Understanding Measurement of Upper Limb Activity and Goal Attainment within Neurorehabilitation

Shannon Pike

BHSc Occupational Therapy

A thesis submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

Department of Community and Clinical Allied Health
School of Allied Health, Human Services and Sport
College of Science, Health and Engineering

La Trobe University
Victoria, Australia

May 2021

Table of Contents

List of Tables

List of Figures

List of Abbreviations

10MWT	Ten Metre Walk Test
ABI	Acquired Brain Impairment
ARAT	Action Research Arm Test
ArmA	Arm Activity Measure
AqoL	Assessment of Quality of Life
BI	Barthel Index
BI (C&W)	Barthel Index (Colin & Wade)
CMSA	Chedoke-McMaster Stroke Assessment
COSMIN	Consensus-based Standards for the selection of health Measurement Instruments
CP	Cerebral Palsy
DAS	Disability Assessment Scale
EQ-5D	EuroQol – 5 dimension
FAT	Frenchay Arm Test
mFAT	modified Frenchay Arm Test
FIM	Functional Independence Measure
GAS	Goal Attainment Scale
GAS – 10pt	Goal Attainment Scale – 10 point
GAS – Light	Goal Attainment Scale – Light
Global Ax	Global Assessment Scale
ICF	International Classification of Functioning, Disability and Health
KleinBell ADL	Klein-Bell Activities of Daily Living scale
LASIS	Leeds Adult Spasticity Impact Scale
SF-36	Medical Outcome Study 36-Item Short-Form Health Survey
MAL	Motor Activity Log
MAL-5	Motor Activity Log – 5
MAL-28	Motor Activity Log – 28
mo	Month
MI	Motricity Index
MS	Multiple Sclerosis
N	Number
NHPT	Nine Hole Peg Test
OHS	Oxford Handicap Scale
PDS/CBS	Patient Disability Scale / Carer Burden Scale

PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-P	Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols
RMA	Rivermead Motor Assessment
RMA-UL	Rivermead Motor Assessment - Upper Limb
SA-SIP	Stroke-Adapted Version of the Sickness Impact Profile
SD	Standard deviation
SIS	Stroke Impact Scale
TBI	Traumatic Brain Injury
UL MAS	Upper Limb Motor Assessment Scale
WHO	World Health Organisation
Yr	Year

Abstract

Background: Measurement of upper limb performance outcomes following ABI is integral to effective neurorehabilitation. Yet there is scant evidence about upper limb use from the ABI survivor's perspective and population-specific psychometric evidence on existing outcome measures. This thesis addresses these gaps.

Aim: To describe the experience of upper limb motor impairment following ABI, to characterise the impact on every-day life, and to understand measurement of upper limb activity and goal attainment within neurorehabilitation. Thus informing the objective of this thesis: to determine the possibility to accomplish precise and individualised measurement of upper limb activity and goal attainment in neurorehabilitation.

Method: Five studies were conducted. Qualitative enquiry with stroke survivors revealed their experience of living with upper limb impairment from spasticity (Study-1) and how the impact can be characterised using the ICF (Study-2). A systematic review using COSMIN synthesised published psychometric evidence for outcome measures of everyday upper limb use (Study-3). A repeated-measures study compared performance of two goal attainment scaling methods (Goal Attainment Scaling and GAS-Light) (Study-4), and cross-sectional two group survey explored therapist perceptions of clinical utility and patient acceptance of goal attainment scaling methods (Study-5).

Results: The experience of having upper limb impairment required continual adaptation and adjustment using processes contextualised in the body, time, a life situation and in relation to services and assistive technology. The ICF Comprehensive Core Set for Stroke with eight recommended code additions captured impacts of this upper limb impairment. Of 30 measurement tools appraised, four were recommended; the top one being the ArmA. GAS-Light was as valid, reliable and sensitive as GAS. GAS-Light had stronger clinical utility properties and was acceptable to patients.

Conclusion: This thesis revealed lived-experience perspectives and ICF-linked impacts that can inform patient-centred approaches to goal setting in neurorehabilitation. The

thesis contributes new evidence about existing measures to inform measurement of upper limb impairment impacts on activity and participation for ABI survivors in neurorehabilitation. Together these findings support evidence-based approaches to upper limb neurorehabilitation.

Statement of authorship

Except where reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis accepted for the award of any other degree or diploma. No other person's work has been used without due acknowledgement in the main text of the thesis. This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

Shannon Pike is the sole author of Chapter 1 (Introduction and Background), Chapter 8 (Clinical Utility of Goal Attainment Scaling Methods) and Chapter 7 (Discussion). The remaining chapters are multi-authored publications on which Shannon Pike was the lead author and completed the majority of work across all studies. The conception, design and management of studies; data collection and analysis; writing and subsequent revisions of publications; as well as response to journal peer-review was led by the PhD candidate, under the supervision of her supervisors. Co-authors provided advice and/or assistance with study planning, design and revision of the final draft of publications. Publication permissions are provided in Appendix A.

All research procedures reported in this thesis were approved by relevant Ethics Committees prior to the commencement of each study. Copies of ethics approvals are provided in Appendix B. Copies of study registrations are provided in Appendix C.

This work was supported by an Australian Government Research Training Program Scholarship, La Trobe University School of Allied Health Postgraduate Support Grant and NSW Health Postgraduate scholarship.

Signature: _____

Shannon Pike

Date: 14th May 2021

Dissemination of findings

Peer-reviewed publications

Pike, S., Lannin, N. A., Cusick, A., Wales, K., Turner-Stokes, L., & Ashford, S. (2015). A systematic review protocol to evaluate the psychometric properties of measures of function within adult neurorehabilitation. *Systematic Reviews, 4*(1), 86. doi: 10.1186/s13643-015-0076-5

Pike, S., Lannin, N. A., Wales, K., & Cusick, A. (2018). A systematic review of the psychometric properties of the Action Research Arm Test in neurorehabilitation. *Australian Occupational Therapy Journal, 65*(5), 449-471. doi:10.1111/1440-1630.12527

Pike, S., Cusick, A., Wales, K., Cameron, L., Turner-Stokes, L., Ashford, S., & Lannin, N. A. (2021). Psychometric properties of measures of upper limb activity performance in adults with and without spasticity undergoing neurorehabilitation – A systematic review. *PLOSONE,* 16(2): e0246288. doi.org/10.1371/journal.pone.0246288

Pike, S., Lannin, N. A., Cameron, L., Palit, M., Cusick, A. (2021). Chronic stroke survivors with upper limb spasticity: linking experience to the ICF. *Disability and Rehabilitation.* doi: 10.1080/09638288.2021.1894490.

Pike, S., Cusick, A., Turner-Stokes, L., Buckley, D., Li Teng Han, M., Lannin, N. A. (under review). Comparison of Goal Attainment Scaling (GAS) and GAS-Light performance in neurorehabilitation.

Pike, S., Lannin, N. A., Cameron, L., Palit, M., Schneider, E., & Cusick, A. (under review). Perspectives of stroke survivors with chronic upper limb spasticity: a qualitative study revealing contexts and processes of adjustment and adaptation.

Published abstracts

Pike, S., Lannin, N. A., Cusick, A., Wales, K., Ashford, S., & Turner-Stokes, L. (2015). A systematic review of measurement properties of functional assessments including their validity as measures of every-day activity performance and life participation. *Australian Occupational Therapy Journal, 62 (Suppl. 1): 122.* doi:10.1111/1440-1630.12212_2

Pike, S., Lannin, N. A., Cusick, A., & Wales, K. (2017). Measuring everyday activity performance and life participation: A systematic review of the nine-hole peg test. *Australian Occupational Therapy Journal, 64 (Suppl. 2): 89.* doi: 10.1111/1440-1630.12406

Pike, S., Lannin, N. A., Cusick, A., & Wales, K. (2017). A systematic review of the psychometric properties of the upper-limb motor assessment scale and its ability to measure everyday activity performance. *Australian Occupational Therapy Journal, 64 (Suppl. 2): 15.* doi: 10.1111/1440-1630.12400

Pike, S., Lannin, N. A., Cameron, L., Wales, K., Ashford, S., Turner-Stokes, L., & Cusick, A. (2019). Using psychometric information about upper limb measurement tools to select best-quality assessments for use with adults in neurorehabilitation. *Australian Occupational Therapy Journal, 66 (Suppl. 1): 96.* doi: 10.1111/1440-1630.12586

Oral conference presentations

Pike, S., Lannin, N. A., Cusick, A., & Wales, K., Ashford, S., & Turner-Stokes, L. (2015). A systematic review of measurement properties of functional assessments including their validity as measures of every-day activity performance and life participation. *Oral presentation at the Occupational Therapy National Conference, 1-3rd July 2015, Melbourne.*

Pike, S., Lannin, N., Cusick, A., & Wales, K. (2017). Measuring everyday activity performance and life participation: A systematic review of the nine-hole peg test. *Oral presentation at the Occupational Therapy National Conference, 19-21st July 2017, Perth.*

Pike, S., Lannin, N. A., Cusick, A., & Wales, K. (2017). A systematic review of the psychometric properties of the upper-limb motor assessment scale and its ability to measure everyday activity performance. *Oral presentation at the Occupational Therapy National Conference, 19-21st July 2017, Perth.*

Pike, S., Lannin, N. A., Cameron, L., Wales, K., Ashford, S., Turner-Stokes, L., & Cusick, A. (2019). Using psychometric information about upper limb measurement tools to select best-quality assessments for use with adults in neurorehabilitation. *Oral presentation at the Occupational Therapy National Conference, 10 – 12th July 2019, Sydney.*

Posters

Pike, S., Lannin, N. A., Wales, K. & Cusick, A. (2018). A systematic review of measurement properties of the Action Research Arm Test (ARAT) for assessment of functional upper limb performance with people who have neurological spasticity. *Poster presentation at the 11th Pan Pacific Conference in Rehabilitation Hong Kong 17th – 18th November 2018.*

Acknowledgements

Thank you to the many people who contributed to my successful completion of this program of research. I wish to express my sincere appreciation and gratitude to my primary supervisor Professor Natasha Lannin for your guidance throughout this thesis. I am grateful for the time you have generously invested in my development as a clinician-researcher. I continue to be inspired by the brilliance of your intellect, your research and clinical expertise, and your passion for excellence to improve patient outcomes.

I thank my external supervisor, Professor Anne Cusick, for your mentorship within my research and more broadly as a clinician researcher. Thank you for sharing your expertise in qualitative research methods. I will always admire your ability to write so eloquently, and am grateful for the continued opportunities you provide me to develop and refine my writing. You have provided me with a lifelong gift, thank you.

I wish to sincerely thank study participants. Thank you for your time, knowledge and sharing your experience of neurorehabilitation.

I thank my colleagues, in particular the Wagga Wagga rehabilitation team, Dr Kylie Wales, and my friends, for your continued interest, support and encouragement throughout my PhD journey.

Most importantly, I thank my family. I thank my parents for their emotional and practical support, their intuition and impeccable timing to offer help before I needed to ask; I can never thank you enough for being present. I thank my parents-in-law for your continued support and for regularly checking in on my progress. And finally, I acknowledge and thank Matt for his continuous support. Your support never wavered, and even when I doubted myself, your patience and reassurance kept me going. Thank you for reminding me of what I am capable of. To Mason and Torah, you have only known me to be “doing my research”, thank you for being my most exuberant supporters. I think what I am most proud of is that the commitment that you have seen me demonstrate has contributed to the determined traits I see in both of you.

Chapter 1:

Introduction

1.0. Introduction

This thesis seeks to understand the impact of upper limb motor impairment arising from acquired brain impairment (ABI) so as to enable individualised measurement of upper limb activity and participation outcomes within neurorehabilitation. A large number of tools purport to measure activity and participation outcomes in neurorehabilitation clinical populations – be that through observation of performance, self-report or by measuring the intervention or neurorehabilitation goals met. Not all these tools have evidence of psychometric properties relevant to neurorehabilitation clinical populations and the quality of tool applicability and rigour is therefore mixed. Clear recommendations that use psychometric evidence to inform clinician measurement and goal setting practice are lacking. The variety of approaches and tools used in practice is a likely consequence; there is no gold standard and a limited evidence base for comparison of tool psychometric properties.

This program of research will seek to characterise the experience of everyday life with upper limb motor impairment, identify and synthesise existing evidence about the measurement of upper limb activity, and to test the concurrent validity of goal setting tools to provide information that can be used in tool selection by neurorehabilitation clinicians. There will be a particular focus on tools which have been used with people with upper limb spasticity. In addition to having information about psychometric properties of outcome measurement tools relevant to upper limb activity, it is important to have information about what activities are meaningful to the person living with an upper limb motor impairment. Qualitative methodology will be used to illuminate the experience of people with the upper limb impairment of spasticity. Together the qualitative and quantitative findings will inform a greater understanding of measurement of meaningful activity and participation outcomes for people with upper limb motor impairment following ABI and in particular spasticity.

To place this program of research in context, Chapter 1 will review three key topic areas pertinent to the aim of the PhD. First, to provide a context for the study, ABI in Australia will be summarised. Specifically, the upper limb motor impairments experienced

following ABI and the impact of these impairments on activity and participation will be discussed (section 1.1 – 1.4). Secondly, contemporary neurorehabilitation in Australia will be reviewed and how it addresses motor impairments, as well as the importance of using a goal centred approach will be discussed (section 1.5). The importance of a goal-centred approach in neurorehabilitation and how this is thought to improve activity and participation after ABI will be key to this understanding. Measurement in clinical practice, specifically psychometric properties of tools and how knowledge of these properties can inform the selection of outcome measurement tools will be discussed (section 1.5). Finally, the importance of the ABI survivors voice and perspective of the experience of living with an upper limb motor impairment to inform measurement will be discussed (section 1.6). Approaches to the measurement of activity and participation goals and clinical outcomes will be discussed and links between the common motor impairments and limitations experienced after ABI, the importance of therapy aimed at meaningful goal attainment, and capturing change during neurorehabilitation will be drawn together. In this way, Chapter 1 will demonstrate gaps in the evidence, inconsistencies in practice, and the need for the studies presented in the remainder of the thesis.

1.1. Acquired Brain Impairment (ABI)

ABI is one of the most common causes of significant and life-long disability worldwide, leading to socioeconomic disadvantage and reduced quality of life (Kavanagh et al., 2015). In Australia alone, approximately 1 in 45 people acquire an ABI every year (Australian Institute of Health and Welfare, 2007). Considered an umbrella term, ABI describes any damage to the brain that occurs during or following birth (Australian Institute of Health and Welfare, 2007). Common causes include stroke, head trauma, neurodegenerative conditions, hypoxia and cerebral palsy (Australian Institute of Health and Welfare, 2007). The individual studies included in this program of research include participants with an ABI. Recent data from the Australasian Rehabilitation Outcomes Centre shows that stroke is the most common of these causes (Australasian Rehabilitation Outcomes Centre, 2020a, 2020b). Each individual study presented across thesis chapters will define the particular ABI diagnoses of study participants; primarily

including stroke and trauma, with a smaller number having a diagnosis of multiple sclerosis and cerebral palsy.

Stroke occurs when the blood supply carrying oxygen and nutrients is interrupted, either by the blockage of an artery (ischaemic) or bleeding within the brain (haemorrhagic) resulting in brain cell death (Stroke Foundation, 2020). The effect of stroke, and the resulting impairments, depend on the extent of cell death and infarct location within the brain. Advances in acute stroke management have increased survival rates with a 30% decline in stroke deaths between 1981 and 2018 in Australia, the site of this program of studies (Australian Institute of Health and Welfare, 2020), although there remains an estimated 475,000 stroke survivors living in Australian communities (Cadilhac et al., 2019; Deloitte Access Economics, 2017). Despite medical advances and clinical guidelines to inform acute management and neurorehabilitation practices (Hebert et al., 2016; Intercollegiate Stroke Working Party, 2008; Lees, 2009; National Clinical Guideline Centre (NICE), 2013; Ottawa Panel, 2006; Scottish Intercollegiate Guidelines Network (SIGN), 2010; Stroke Foundation, 2020; The European Stroke Organisation Executive Committee, 2008), stroke remains a leading cause of global disability (Donkor, 2018).

ABI also arises from traumatic aetiologies and similar to stroke, result in highly variable presentations dependent on the severity of the initial injury. The incidence of TBI internationally varies widely by age and between countries, with reported incidence rates limited due to many cases not being reported or recognised by healthcare professionals (Nguyen et al., 2016). In New South Wales (Australia), 99/100,000 persons per year sustain a TBI, with rates distinctly higher for those aged between 15-19 years and greater than 75 years (Pozzato, Tate, Rosenkoetter, & Cameron, 2019). Falls, motor vehicle accidents and assault are primary causes (Pozzato et al., 2019). Whilst overall hospitalisation rates have not increased in the last 20 years (Pozzato et al., 2019; Tate, Lane-Brown, Myles, & Cameron, 2020), TBI remains a significant cause of disability in Australia.

Irrespective of cause, the effects of ABI are devastating. ABI results in a wide range of complex and diverse neurological deficits disrupting the survivor's everyday life. The

level of disruption to independence, function and engagement in everyday life experienced after ABI will depend on several factors. This includes the severity of damage to the brain, acute spontaneous cell regeneration and recovery of oedema and inflammation, and the degree of neurological reorganisation or neural plasticity consolidated via repetitive task specific movement (Li, 2017). Contextual factors also influence the clinical presentation and how the brain impairment is experienced (Australian Institute of Health and Welfare, 2007; Ponsford, 1995; Turner-Stokes, Pick, Nair, Disler, & Wade, 2015). Limitations performing everyday activities within the home such as self-care and domestic tasks, as well as restrictions within the community such as shopping, leisure activities and employment are commonly experienced (Ponsford et al., 2014; Tate et al., 2020; Tse et al., 2019; Ytterberg, Dyback, Bergstrom, Guidetti, & Eriksson, 2017). A common sequalae of ABI is motor impairment in one or more upper limbs. Being able to proficiently use the arm and hand (i.e. upper limb) after an ABI is integral to the performance of everyday tasks and activities.

Several impairments can arise following an ABI, which either individually or in combination leads to difficulty producing coordinated and controlled arm and hand movements. These impairments, such as the inability to move or to coordinate movement, to detect, discriminate or recognise somatosensation, or to sense the position or limb movement, all impact on the use of the upper limb in everyday activity (Carey, 1995, 2012; Doyle, Bennett, Fasoli, & McKenna, 2010; Meyer, Karttunen, Thijs, Feys, & Verheyden, 2014). Impairments in cognition and visuospatial perception also impact on the use of the upper limb in everyday activity (Bosma, Nijboer, Caljouw, & Achterberg, 2020; Walker, Sunderland, Sharma, & Walker, 2004), making planning, initiating or controlling movements challenging. Whilst clinical presentations post-ABI vary; upper limb motor impairments are the most common to arise after stroke, with more than half of all stroke survivors reporting a failure to regain use of their affected upper limb (Kwah, Harvey, Diong, & Herbert, 2013). Furthermore, decreased use of the affected upper limb despite a level of motor control return, known as learned non-use, is common, particularly when motor impairment and spasticity are experienced in combination (Hirsch et al 2021). Thus, addressing upper limb motor impairments and

learned non-use is a common goal in stroke neurorehabilitation programs using interventions that specifically target increasing activity and participation.

1.2. Upper Motor Neurone Impairments

The damage that occurs to the cortex after ABI from stroke or trauma leads to the upper motor neurone syndrome (Barnes, 2008; Bhimani & Anderson, 2014). Negative and positive features characterise the upper motor neurone syndrome (Pandyan, Hermens, Conway, & Johnson, 2018), each of which impact on upper limb use. Negative features include a reduction in motor activity such as muscle weakness, loss of dexterity, reduced coordination and fatigue (Ada, O'Dwyer, & O'Neill, 2006; Barnes, 2008) while positive features include increased tendon reflexes, a positive Babinksi sign, clonus, spasticity, dyssynergic patterns of co-contraction during movement, abnormal postures, flexor and extensor spasms (Barnes, 2008; Sommerfeld, Eek, Svensson, Holmqvist, & Von Arbin, 2004).

1.2.1. Upper Limb Spasticity

A common positive feature of an upper motor neurone impairment is spasticity. Spasticity, in this thesis, is recognised as one of several impairments that can be experienced following ABI. Post-ABI spasticity is often focal due to being localised to a small number of muscles (Williams et al., 2020). The reported prevalence of upper limb spasticity as a result of ABI is highly variable; following a stroke, for example, it is reported to occur in 19% (Sommerfeld et al., 2004) through to 46% (Opheim, Danielsson, Alt Murphy, Persson, & Sunnerhagen, 2014; Urban et al., 2010) and as high as 87% (Malhotra et al., 2008) of survivors.

Historical and contemporary challenges establishing a universally accepted, scientifically valid and clinically useful definition of spasticity (Ibuki & Bernhardt, 2007; Johnson & Pandyan, 2008; Malhotra, Pandyan, Day, Jones, & Hermens, 2009; Pandyan et al., 2005) have contributed to inconsistencies in the identification of the presence of spasticity. Lance (1980) chaired a panel who developed the most widely used and accepted definition, where they defined spasticity as “a motor disorder characterised by a velocity-dependent increase in tonic stretch reflex (muscle tone) with exaggerated

tendon jerks, resulting from hyperexcitability of the stretch reflex as one component of the upper motor neurone syndrome" (Lance, 1980). Although impaired voluntary movement and abnormal posture may be associated with spasticity, they do not define it (Ibuki & Bernhardt, 2007). Thus, this definition highlights that spasticity is not the only component of the upper motor neurone syndrome and other features are likely to be at play when movement is impaired. Despite the clarification Lance's definition provided, the spasticity term has been used to label all stiffness and impaired movement (Bhimani & Anderson, 2014; Ibuki & Bernhardt, 2007), which complicates clinical discussions about movement disorders and treatment planning.

Valid and reliable measurement of spasticity is dependent on consistently identifying its presence and severity. Selected outcome measurement tools are often not congruent with defined clinical features (Fheodoroff et al., 2016). As a result, tools may not measure the construct of spasticity as intended. Challenges defining and measuring upper limb spasticity add to the complexity of understanding the impact of an ABI on activity and participation. This is further discussed in Sections 1.3 and 1.4.

1.3. Understanding the Impact of Upper Limb Motor Impairment on Everyday Life: The ICF

The International Classification of Functioning, Disability and Health (ICF) provides a common language and framework to describe and understand disability and health states (World Health Organisation, 2002). The ICF shifts the focus of disability beyond the health condition to better understand the impact on the person at a body or body part level (impairments in body function and structure), on the whole person (activity limitations), and the whole person in a life situation (participation restrictions) (World Health Organisation, 2002). The biopsychosocial model provides the theoretical foundation for the ICF (World Health Organisation, 2002, 2013). Thus, the ICF recognises the complex and dynamic interaction between the person's health condition, the impact of that condition on body function and structure, activity and participation and their unique environmental and personal factors on how they experience disability (World Health Organisation, 2002). Environmental and personal factors are contextualised to the person and can act as barriers or facilitators to activity and participation (World

Health Organisation, 2002). Environmental factors include physical, social and attitudinal contexts (World Health Organisation, 2002). All elements in the ICF are classified in a coded taxonomy with the exception of personal factors since these are particular to the individual and may include age, sex, race, lifestyle, habits, education and profession (World Health Organisation, 2002).

The relationship proposed in the ICF between the three levels of body structure and function (impairment), activity (limitations) and participation (restrictions) is not linear (Playford, 2020) and the delineation between activity and participation is intricate. Action executed in a social environment may be considered participation, and participation always involves the execution of an action or task (World Health Organisation, 2013). Thereby whilst the two terms are defined as explicitly different and are considered separate, they are highly related and linked and consequently are presented in a single list within the ICF (World Health Organisation, 2013). This program of study adopts this integrated approach to activities and participation, as recommended by the World Health Organisation (2013).

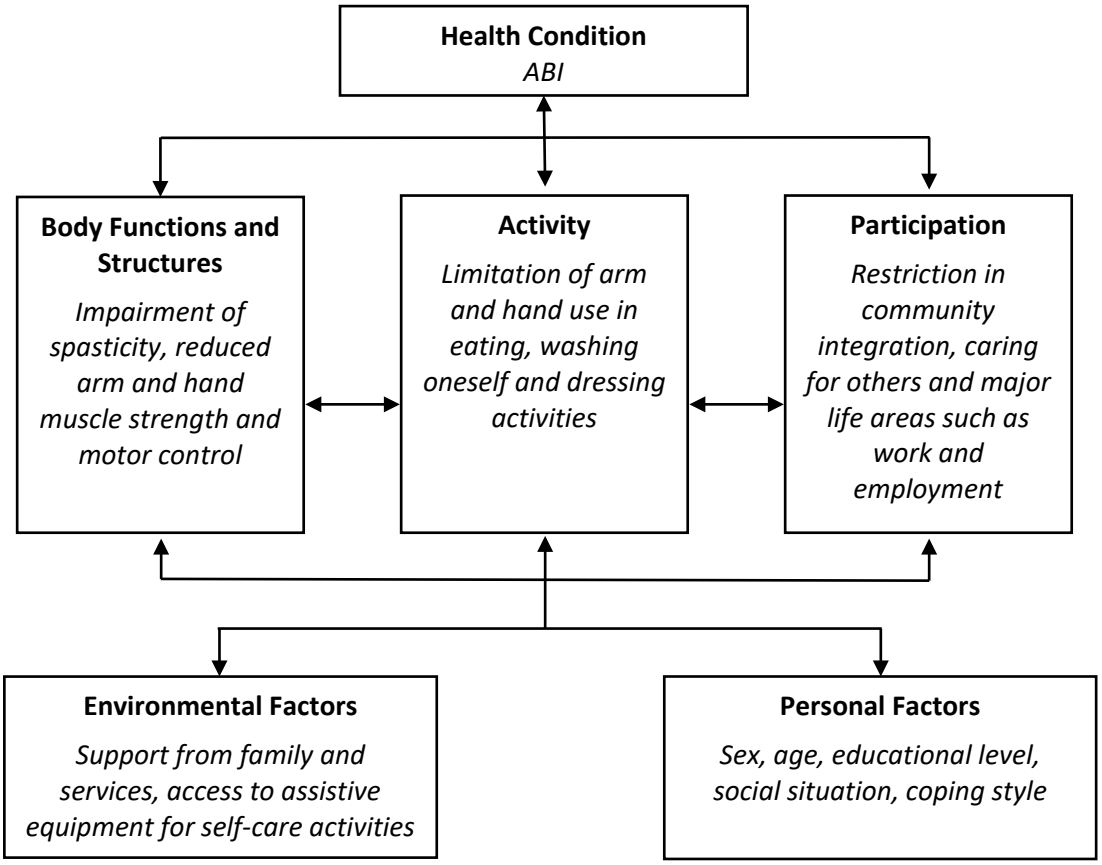


Figure 1.1. *The ICF as applied to upper limb motor impairment post-ABI*

In this thesis, the ICF framework as depicted in Figure 1.1 is used to consider and understand the interplay between upper limb motor function on the capacity and performance of activity or ability to participate in a life situation, whilst considering the influence of environmental and personal factors. Thus impairments, limitations and restrictions can be understood. For example, one individual post-stroke living alone and required to manage all domestic tasks is likely to experience the impact of upper limb spasticity differently to someone with the same diagnosis and impairment living with a supportive carer.

1.4. Relationships Between Upper Limb Motor Function, Activity and Participation

Use of the upper limb in everyday activities is complex. Simple through to multidimensional tasks are performed such as reaching, stabilising, grasping and

manipulating objects (Lamers, Kelchtermans, Baert, & Feys, 2014) to allow adults to perform meaningful activities. The activity performed dictates the level of upper limb motor performance required, and whether one or both upper limbs are used. For example, bilateral upper limb use is more common when performing activities (Kilbreath & Heard, 2005) and interacting within the environment (Yozbatıran, Baskurt, Baskurt, Ozakbas, & Idiman, 2006). Understanding how individual motor impairments impact on the ability to perform an activity and ultimately a person's capacity to participate is needed in neurorehabilitation. Evidence for this relationship will now be presented.

1.4.1. Upper Limb Spasticity, Impairment and Activity

Investigations have demonstrated that upper limb spasticity contributes to limited passive range of motion of the elbow, wrist and fingers, and increased pain (Andringa et al., 2019; Doan et al., 2012). Upper limb spasticity may also contribute to the development of contractures within the first four months post-stroke (Ada et al., 2006). Thus exploring ways to reduce upper limb spasticity as a means to improve upper limb motor control and activity has received significant clinical attention. To date, pharmacological interventions such as botulinum toxin A and adjunct therapies such as casting or motor training have been used to improve functional hand and arm use (Intiso et al., 2013; Kinnear, Lannin, Cusick, Harvey, & Rawicki, 2014; Shaw et al., 2011; Sommerfeld et al., 2004). While studies suggest that the ability to care for the affected upper limb, such as maintaining hand hygiene and ability to complete dressing tasks, improves when spasticity is reduced (Andringa et al., 2019; Doan et al., 2012; Shaw et al., 2011; Turner-Stokes, Fheodoroff, Jacinto, & Maisonobe, 2013), clinical trial evidence shows that reducing spasticity does not improve arm and hand use (Andringa et al., 2019). Evidence regarding the relationship between upper limb spasticity and participation restrictions for adults is scarce (Andringa et al., 2019; Williams et al., 2020) despite spasticity reducing health related quality of life (Doan et al., 2012) and increasing burden (Andringa et al., 2019). This evidence gap highlights the need to investigate the effect of upper limb spasticity on participation outcomes for adults post-ABI.

1.4.2. Upper Limb Motor Impairment

Loss of motor function following ABI has the most disabling effect on the efficient use of the upper limb in everyday activity, not spasticity (Burridge, Turk, Notley, Pickering, & Simpson, 2009; Canning, Ada, Adams, & O'Dwyer, 2004; Harris & Eng, 2007; Harris & Eng, 2010). Muscle weakness in particular, is the primary motor impairment limiting the performance of purposeful movements and everyday activity post-stroke (Ada et al., 2006; Bohannon, Warren, & Cogman, 1991; Boissy, Bourbonnais, Carlotti, Gravel, & Arsenault, 1999; Burridge et al., 2009; Harris & Eng, 2007). Impaired motor control, active range of movement and reduced dexterity post-stroke (Burridge et al., 2009) and for people with multiple sclerosis (Cattaneo, Lamers, Bertoni, Feys, & Jonsdottir, 2017) limits the performance of activities such as dressing and cooking where control of the upper limb and manipulation of objects is essential. In this way, upper limb motor impairment impacts on the survivor's ability to engage in activities outside of their home such as gardening, sports including golf and bowling, crafts and games and the ability to participate in roles such as caring for grandchildren or volunteering in the community (White, Mackenzie, Magin, & Pollack, 2008). Current evidence specific to upper limb motor impairment demonstrates that motor impairment does impact activity and participation (Cawood, Visagie, & Mji, 2016). However, activity restrictions are more strongly related to participation restriction than measures of impairment (Faria-Fortini, Michaelsen, Cassiano, & Teixeira-Salmela, 2011; Gadidi, Katz-Leurer, Carmeli, & Bornstein, 2011; Harris & Eng, 2007). Knowledge of the relationships between upper limb motor impairment, activity and participation enable neurorehabilitation clinicians to better consider the focus of measurement and to select the most appropriate outcome measurement tools to meet the needs of the survivor throughout their recovery and adaptation to living with residual impairments. Furthermore, informing effective and efficient patient-centred neurorehabilitation interventions (Harris & Eng, 2007) post-ABI.

1.4.3. ICF Core Sets to Assist in Targeted Measurement and Intervention

Consideration of the entire ICF framework within clinical practice is not feasible due to being comprised of over 1400 categories (Grill & Stucki, 2011). The development of tailored core sets containing a select number of categories deemed relevant and specific

to individual diagnoses (including stroke) and settings (including neurorehabilitation), aims to facilitate use in practice. The core sets define common standards for what should be measured and may assist in standardised outcome measurement tool selection (Grill et al., 2011).

Table 1.1 compares and contrasts categories of comprehensive core sets relevant to this thesis. The core set for children and youth with cerebral palsy is presented as the adult core set is currently under development. Differences in the lived experience from an adult perspective is likely to result in revisions to included categories (Limsakul et al., 2020).

Of interest are the broad scope and the similarities between the overall number of categories included across the diagnostic core sets. Notable differences include the inclusion of respiratory function impairments for multiple sclerosis and cerebral palsy, thermoregulatory function impairments for multiple sclerosis only, and an increased number of activities and participation categories within the traumatic brain injury core set. Comparison of the setting-based core sets reveals the post-acute core set provides greater opportunity to capture and understand the breadth of impact of ABI in comparison to the rehabilitation core set.

Table 1.1. *Comparison of ICF Core Sets*

	ICF Chapters and second-level categories	Rehabilitation	Post-acute	Stroke	TBI	MS	CP
Body Functions							
b110	Consciousness functions						
b114	Orientation functions						
b117	Intellectual functions						
b126	Temperament and personality functions						
b130	Energy and drive functions						
b1300	Energy level						
b1301	Motivation						
b1308	Energy and drive functions, other specified (fatigue)						
b134	Sleep Functions						
b140	Attention Functions						
b144	Memory functions						
b147	Psychomotor functions						
b152	Emotional functions						
b156	Perceptual functions						
b160	Thought functions						
b163	Basic cognitive functions						
b164	Higher-level cognitive functions						
b167	Mental functions of language						
b172	Calculation functions						
b176	Mental functions of sequencing complex movements						
b180	Experience of self and time functions						
b210	Seeing functions						
b215	Functions of structures adjoining the eye						
b2152	Functions of external muscles of the eye						
b230	Hearing functions						
b235	Vestibular functions						
b240	Sensations associated with hearing and vestibular functions						
b255	Smell function						
b260	Proprioceptive function						
b265	Touch functions						
b270	Sensory functions related to temperature and other stimuli						
b280	Sensation of pain						
b310	Voice functions						
b320	Articulation functions						
b330	Fluency and rhythm of speech functions						

	ICF Chapters and second-level categories	Rehabilitation	Post-acute	Stroke	TBI	MS	CP
b340	Alternative vocalisation functions						
b410	Heart functions						
b415	Blood vessel functions						
b420	Blood pressure functions						
b430	Haematological system functions						
b435	Immunological system functions						
b440	Respiration functions						
b445	Respiratory muscle functions						
b450	Additional respiratory functions						
b4501	Transportation of airway mucus						
b455	Exercise tolerance functions						
b510	Ingestions functions						
b5104	Salivation						
b5105	Swallowing						
b515	Digestive functions						
b525	Defecation functions						
b530	Weight maintenance functions						
b535	Sensations associated with the digestive system						
b540	General metabolic functions						
b545	Water, mineral and electrolyte balance functions						
b550	Thermoregulatory functions						
b5500	Body temperature						
b5508	Thermoregulatory functions, other specified (sensitivity to heat)						
b5508	Thermoregulatory functions, other specified (sensitivity to cold)						
b555	Endocrine and gland functions						
b620	Urination functions						
b630	Sensations associated with urinary functions						
b640	Sexual functions						
b710	Mobility of joint functions						
b715	Stability of joint functions						
b730	Muscle power functions						
b735	Muscle tone functions						
b740	Muscle endurance functions						
b750	Motor reflex functions						
b755	Involuntary movement reaction functions						
b760	Control of voluntary movement functions						
b765	Involuntary movement functions						
b7650	Involuntary contraction of muscles						

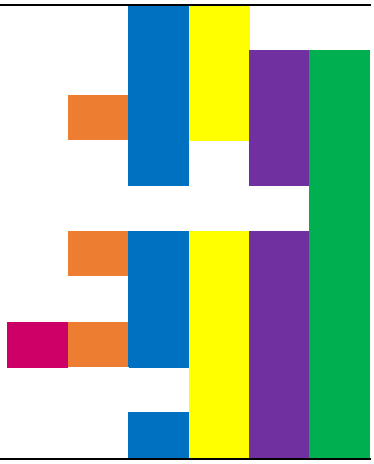
	ICF Chapters and second-level categories	Rehabilitation	Post-acute	Stroke	TBI	MS	CP
b7651	Tremor						
b770	Gait pattern functions						
b780	Sensations related to muscles and movement functions						
b810	Protective functions of the skin						
Body Structures							
s110	Structure of brain						
s120	Spinal cord and related structures						
s130	Structure of meninges						
s320	Structure of mouth						
s410	Structure of cardiovascular system						
s430	Structure of respiratory system						
s530	Structure of stomach						
s610	Structure of urinary system						
s710	Structure of head and neck regions						
s720	structure of shoulder region						
s730	Structure of upper extremity						
s750	Structure of lower extremity						
s760	Structure of trunk						
s7700	Bones						
s7703	Extra-articular ligaments, fasciae, extra muscular aponeuroses, retinacula, septs, bursae, unspecified						
s810	Structure of areas of skin						
Activities and Participation							
d110	Watching						
d115	Listening						
d120	Other purposeful sensing						
d130	Copying						
d131	Learning through actions with objects						
d133	Acquiring language						
d135	Rehearsing						
d137	Acquiring concepts						
d140	Learning to read						
d145	Learning to write						
d155	Acquiring skills						
d160	Focusing attention						
d163	Thinking						

	ICF Chapters and second-level categories	Rehabilitation	Post-acute	Stroke	TBI	MS	CP
d166	Reading						
d170	Writing						
d172	Calculating						
d175	Solving problems						
d177	Making decisions						
d210	Undertaking a single task						
d220	Undertaking multiple tasks						
d230	Carrying out daily routine						
d240	Handling stress and other psychological demands						
d250	Managing one's own behaviour						
d310	Communicating with – receiving – spoken messages						
d315	Communicating with – receiving – nonverbal messages						
d325	Communicating with – receiving – written messages						
d330	Speaking						
d331	Pre-talking						
d335	Producing nonverbal messages						
d345	Writing messages						
d350	Conversation						
d360	Using communication devices and techniques						
d410	Changing basic body position						
d415	Maintaining a body position						
d420	Transferring oneself						
d430	Lifting and carrying objects						
d435	Moving objects with lower extremities						
d440	Fine hand use						
d445	Hand and arm use						
d450	Walking						
d455	Moving around						
d460	Moving around in different locations						
d465	Moving around using equipment						
d470	Using transportation						
d475	Driving						
d510	Washing oneself						
d520	Caring for body parts						
d530	Toileting						
d540	Dressing						
d550	Eating						
d560	Drinking						

	ICF Chapters and second-level categories	Rehabilitation	Post-acute	Stroke	TBI	MS	CP
d570	Looking after one's health						
d620	Acquisition of goods and services						
d630	Preparing meals						
d640	Doing housework						
d650	Caring for household objects						
d660	Assisting others						
d710	Basic interpersonal interactions						
d720	Complex interpersonal interactions						
d730	Relating with strangers						
d740	Formal relationships						
d750	Informal social relationships						
d760	Family relationships						
d770	Intimate relationships						
d815	Preschool education						
d820	School education						
d825	Vocational training						
d830	Higher education						
d840	Apprenticeship (work preparation)						
d845	Acquiring, keeping and terminating a job						
d850	Remunerative employment						
d855	Non-remunerative employment						
d860	Basic economic transactions						
d865	Complex economic transactions						
d870	Economic self-sufficiency						
d880	Engagement in play						
d910	Community life						
d920	Recreation and leisure						
d930	Religion and spirituality						
	Environmental Factors						
e110	Products or substances for personal consumption						
e1100	Food						
e1101	Drugs						
e1108	Non-medical drugs and alcohol						
e115	Products and technology for personal use in daily living						
e120	Products and technology for personal indoor and outdoor mobility and transportation						
e125	Products and technology for communication						

	ICF Chapters and second-level categories	Rehabilitation	Post-acute	Stroke	TBI	MS	CP
e130	Products and technology for education						
e135	Products and technology for employment						
e140	Products and technology for culture, recreation and sport						
e150	Design, construction and building products and technology of buildings for public use						
e155	Design, construction and building products and technology of buildings for private use						
e160	Products and technology of land development						
e165	Assets						
e210	Physical geography						
e225	Climate						
e2250	Temperature						
e2251	Humidity						
e2253	Precipitation						
e250	Sound						
e310	Immediate family						
e315	Extended family						
e320	Friends						
e325	Acquaintances, peers, colleagues, neighbours and community members						
e330	People in positions of authority						
e340	Personal care providers and personal assistants						
e355	Health professionals						
e360	Other professionals						
e410	Individual attitudes of immediate family members						
e415	Individual attitudes of extended family members						
e420	Individual attitude of friends						
e425	Individual attitudes of acquaintances, peers, colleagues, neighbours and community members						
e430	Individual attitudes of people in position of authority						
e440	Individual attitudes of personal care providers and personal assistants						
e450	Individual attitudes of health professionals						
e455	Individual attitude of health-related professionals						
e460	Societal attitudes						
e465	Social norms, practices and ideologies						
e515	Architecture and construction services, systems and policies						
e525	Housing services, systems and policies						

	ICF Chapters and second-level categories	Rehabilitation	Post-acute	Stroke	TBI	MS	CP
e535	Communication services, systems and policies						
e540	Transportation services, systems and policies						
e550	Legal services, systems and policies						
e555	Associations and organisational services, systems and policies						
e560	Media services, systems and policies						
e570	Social security services, systems and policies						
e575	General social support services, systems and policies						
e580	Health services, systems and policies						
e585	Education and training services, systems and policies						
e590	Labour and employment services, systems and policies						



Rehab: ICF Rehabilitation Set, Post-acute: Comprehensive ICF Core Set for Neurological Conditions for Post-Acute Care, Stroke: Comprehensive ICF Core Set for Stroke, TBI: Comprehensive ICF Core Set for Traumatic Brain Injury, MS: Comprehensive ICF Core Set for Multiple Sclerosis, CP: Comprehensive Core Set for Children and Youth with Cerebral Palsy from birth to 18 years of age.

Published data evaluating the comprehensiveness and content validity of individual core sets supports their applicability (Algurén, Lundgren-Nilsson, & Sunnerhagen, 2010; Geyh et al., 2004; Glässel, Coenen, Kollerits, & Cieza, 2012; Glässel, Coenen, Kollerits, & Cieza, 2014; Glässel, Kirchberger, Kollerits, Amann, & Cieza, 2011; Glässel et al., 2010; Grill et al., 2011; Lemberg, Kirchberger, Stucki, & Cieza, 2010; Paanalahti, Alt Murphy, Lundgren-Nilsson, & Sunnerhagen, 2014; Paanalahti, Lundgren-Nilsson, Arndt, & Sunnerhagen, 2013). What remains unknown, however, is whether the Comprehensive Core Set for stroke is applicable to and of sufficient breadth to capture and understand the impact of post-stroke upper limb spasticity. This knowledge gap will be addressed within Study 2 within this thesis, and is presented in Chapter 2b.

1.5. Contemporary Australian Neurorehabilitation to Increase Activity Post-ABI

Neurorehabilitation aims to reduce disability and optimise the quality of life through reducing impairments, supporting the return to meaningful activities and participation in the community following an ABI (Barnes, 2003). Wade (2020) identified important features that characterise effective rehabilitation, drawing evidence from ABI

populations to inform his conclusion. Contextually, he determined that rehabilitation may be effective for any adult with a disability at any stage of their illness and could be delivered in any hospital or community setting. Including a process of problem-solving which is framed by the biopsychosocial model of illness, rehabilitation requires an expert multidisciplinary team to collaboratively set goals that reflect patient preferences and to deliver interventions in a person-centred way (Wade, 2020). A key component of effective rehabilitation was said to be monitoring change and evaluation of the rehabilitation program (Wade, 2020), highlighting the importance of meaningful, accurate measurement of outcomes within rehabilitation. A culmination of these key attributes is reflected within national and international guidelines (Hebert et al., 2016; Intercollegiate Stroke Working Party, 2008; National Clinical Guideline Centre (NICE), 2013; Ottawa Panel, 2006; Stroke Foundation, 2020; The European Stroke Organisation Executive Committee, 2008). This thesis will focus on one of the attributes of effective neurorehabilitation proposed by Wade (2020), that of consistent measurement of outcomes to demonstrate change, and to be able to set goals and to identify when goals are achieved.

1.5.1. Goal-directed Neurorehabilitation

Patient-centred goal setting is a recognised integral component of neurorehabilitation, directing an individualised approach to target the unique impact of ABI on the survivor (Hebert et al., 2016; Intercollegiate Stroke Working Party, 2008; Stroke Foundation, 2020). Levack and Siegert (2014, p. 11) define rehabilitation goals as "…a desired future state to be achieved by a person with a disability as a result of rehabilitation activities. Rehabilitation goals are actively selected, intentionally created, have purpose and are shared (where possible) by the people participating in the activities and intervention designed to address the consequences of acquired disability". As defined earlier in this chapter, upper limb motor impairments are experienced in a unique way by each person after ABI so identifying individual, specific outcomes, that when achieved, are perceived to reduce their disability, is important. Patient-centred goal setting allows neurorehabilitation to be tailored to each person and individualised outcomes to be evaluated (Donnelly & Carswell, 2002; Turner-Stokes, 2009).

The evidence base underpinning goal setting and goal-directed neurorehabilitation is limited by research of low methodological quality and heterogeneity (Levack et al., 2015; Rosewilliam, Roskell, & Pandyan, 2011; Sugavanam, Mead, Bulley, Donaghy, & van Wijck, 2013). Despite the research limitations, goal setting in practice appears to enhance communication between the patient and their neurorehabilitation team, positively influence perceived participation in neurorehabilitation, adherence to neurorehabilitation programmes, immediate performance in motor and cognitive tasks, psychosocial outcomes including health-related quality of life, and goal attainment (Levack et al., 2015; Rosewilliam et al., 2011; Sugavanam et al., 2013). Stroke survivors, for example, view patient-centred goal setting as a rewarding process that maintains hope and forward momentum and can empower or disempower pending the interactions and clinician approach (Lloyd, Bannigan, Sugavanam, & Freeman, 2018).

Adhering to best practice goal setting is not a reality of current neurorehabilitation practice (Lloyd et al., 2018; Plant, Tyson, Kirk, & Parsons, 2016; Rose, Rosewilliam, & Soundy, 2017; Rosewilliam et al., 2011). Goal setting often does not occur, and when implemented, the process that is implemented fails to comply with a patient-centred approach (Lloyd et al., 2018; Rosewilliam et al., 2011). Neurorehabilitation professionals report they facilitate strong patient-centred practices, yet patients express a passive experience with a perceived lack of control over their goals (Lloyd et al., 2018; Rosewilliam et al., 2011). Evidence of professional led goal setting practice is found within the types of goals generated (Rosewilliam et al., 2011). Devised goals are often impairment-based rather than the activity-based goals which are preferred by patients, and the goals fail to incorporate their social, activity and participation needs (Rosewilliam et al., 2011). The experience of setting goals from the perspective of the patient-participants and clinician-participants in the study comparing goal attainment methods will be explored and is presented in Chapter 7.

1.5.2. Clinical Evaluation of the Experience of Upper Limb Impairment

Clinical practice guidelines (Hebert et al., 2016; Stroke Foundation, 2020; Wheeler & Acord-Vira, 2016; Winstein et al., 2016) recommend outcome measurement within neurorehabilitation. Outcome measurement tools provide the method to measure the

motor impairment, upper limb performance and use in an activity, but also the impact of neurorehabilitation (by comparing pre- and post-neurorehabilitation scores). Data obtained in measurement is thus key to deliver best practice: to determine neurorehabilitation need, determine goal achievement, determine treatment effectiveness and efficacy, and to identify when treatment protocols must change (Burridge et al., 2019; Duncan Millar, van Wijck, Pollock, & Ali, 2019; Elovic, Simone, & Zafonte, 2004). Standardised outcome measurement is also crucial in quality assurance and clinical accountability to healthcare systems and individual services (Ashford, Slade, Malaprade, & Turner-Stokes, 2008; Hurn, Kneebone, & Cropley, 2006).

Clinicians have a large number of outcome measurement tools at their disposal (Santisteban et al., 2016). When selecting a suitable tool, clinicians should consider the purpose of the tool and the focus or scope of data the tool aims to obtain. Emerging technologies such as kinematic assessments and accelerometers may also provide future developments in this space (Kwakkel et al., 2019). Clinicians should also consider whether the tool is objective (performance-based) or subjective (self-report, proxy informants). Clinical utility characteristics such as feasibility, availability, both physically and economically, are key considerations (Ashford et al., 2008; Greenhalgh, Long, Brettle, & Grant, 1998). Most importantly, clinicians should select psychometrically sound tools that are valid, reliable, responsive to clinically meaningful change (Ashford et al., 2008; Elovic et al., 2004) and ideally, assess across the relevant domains of the ICF (Burridge et al., 2019).

1.5.3. Appraisal of Outcome Measurement Tool Psychometric Properties

Comprehensive evaluation determines the psychometric properties of outcome measurement tools (Greenhalgh et al., 1998). Studies evaluating the psychometric properties of outcome measurement tools should be of high methodological quality so as to provide trustworthy results (Mokkink et al., 2010). Standards and methodology to appraise the psychometric properties of outcome measurement tools have evolved following the development of the first minimum set of standards in 1966 (Rosenkoetter & Tate, 2018; Scientific Advisory Committee of the Medical Outcomes, 2002). This has included further refinement of existing tools and the development of new critical

appraisal tools by research groups, including the Scientific Advisory Committee of the Medical Outcomes Trust (SACMOT) (SACMOT, 2002), the Evaluating the Measurement of Patient-Reported Outcomes (EMPRO) tool (Valderas et al., 2008) and Consensus-based Standards for the selection of health status Measurement INstruments (COSMIN) Initiative (Mokkink et al., 2010; Prinsen et al., 2018).

The COSMIN initiative (https://www.cosmin.nl/) involves an international multidisciplinary team of researchers with expertise in the development and evaluation of outcome measurement tools. The initiative aims to improve the selection of outcome measurement tools within clinical practice and research through developing methodology and practical tools to guide the selection of the most suitable outcome measurement tools. Specifically, COSMIN practical tools provide a comprehensive guideline to search, select and appraise the methodology and results of individual studies evaluating the psychometric properties of the measure against defined criteria (Mokkink et al., 2010; Prinsen et al., 2018). COSMIN evaluates the following psychometric properties: internal consistency, reliability (test-retest, inter-rater, intra-rater), measurement error, content validity (including face validity), structural validity, hypothesis testing, cross-cultural validity, criterion validity, responsiveness, interpretability and generalisability. COSMIN also provides a guideline to synthesise and interpret the methodological quality of included studies and reported results to guide the selection of the most suitable measurement tool for the given situation (Mokkink et al., 2010; Prinsen et al., 2018). COSMIN was adopted in Study 3 to identify the current evidence for the psychometric properties of upper limb performance measurement tools when used with adults with and without spasticity. Chapter 3 presents the study protocol, and Chapter 4 and 5 present the results of the review.

1.5.4. Psychometric Evidence and Measurement Practices

The evidence base regarding the psychometric properties of measures of upper limb activity and participation is growing; however many outcome measurement tools have significant gaps remaining within their evidence of psychometrics (Alt Murphy, Resteghini, Feys, & Lamers, 2015; Ashford et al., 2008; Ashford & Turner-Stokes, 2013). Limitations are reflected within clinical guidelines, where direction concerning specific

tool selection and timing of measurement is still unclear (Burridge et al., 2019). Researchers continue to seek to clarify and gain consensus on the most important areas to measure, the timing of when measurement should occur and how to select the most appropriate outcome measurement tools (Burridge et al., 2019; Duncan Millar et al., 2019; Kwakkel et al., 2017; Kwakkel et al., 2019). The selection and use of tools to measure upper limb activity limitations and participation restrictions post-stroke (Alt Murphy et al., 2015; Ashford et al., 2008; Connell & Tyson, 2012; Duncan & Murray, 2012; Duncan Millar et al., 2019; Santisteban et al., 2016; Turner-Stokes & Turner-Stokes, 1997) and upper limb spasticity management specifically (Ashford & Turner-Stokes, 2013; Cusick, Lannin, & Kinnear, 2015; Hinderer & Gupta, 1996; Kinnear et al., 2014), have been investigated. Investigations found low, inconsistent and highly variable tool use across clinicians, researchers, neurorehabilitation services and countries (Burridge et al., 2019; Burton, Tyson, & McGovern, 2013; Duncan Millar et al., 2019; Santisteban et al., 2016). Furthermore, tools were often administered in versions that had been modified following original development and psychometric testing or were incomplete, using a selection of the tool only (Burridge et al., 2019; Burton et al., 2013; Duncan Millar et al., 2019; Santisteban et al., 2016). Reviews have demonstrated large numbers of outcome measurement tools are included in clinical trials with 31 to 144 identified in trials post-stroke (Alt Murphy et al., 2015; Dekker et al., 2019; Duncan Millar et al., 2019; Santisteban et al., 2016) and up to 31 in mixed neurorehabilitation (Connell & Tyson, 2012). This work has led to recommended outcome measurement tools within clinical trials for the ICF domains of body functions and structure, and activity limitations, however an agreed-upon outcome measurement tool to measure participation performance after stroke remains (Kwakkel et al., 2017). Furthermore, the consensus-based recommendations acknowledge that suggested tools may not evaluate specific outcomes or neurorehabilitation research questions requiring the inclusion of additional tools (Kwakkel et al., 2017). There is therefore a recognised need for further exploration and development to fill this knowledge gap.

Current practices reveal the need to increase psychometric evidence, reduce the number of outcome measurement tools used to enable pooling of results and to evaluate outcomes of importance to survivors that often are poorly captured (Duncan

Millar et al., 2019). Standardised outcome measurement tools have been criticised as lacking sensitivity to capture small or specific change that is relevant and of great importance to survivors (Grenville & Lyne, 1995; Hurn et al., 2006). Furthermore, the allocation of a number to represent a level of performance or degree of change in one's performance on a tool or scale in a clinical environment has been highlighted to lack any meaning for survivors (Elovic et al., 2004). The need for individualised approaches to meaningful measurement is well recognised (Duncan Millar et al., 2019; Grenville & Lyne, 1995; Hurn et al., 2006; Williams et al., 2020) with the inclusion of goal setting to capture individualised outcomes alongside standardised outcome measurement tools proposed as one possible solution.

1.5.5. Tools to Measure Goal Attainment

This thesis earlier defined goal setting as an integral yet challenging component of neurorehabilitation. Plant et al. (2016) identified clinician related, ABI survivor related, service and organisational level barriers and facilitators to goal setting and offered practice recommendations. One specific recommendation is to formally document goals and to use an explicit method to set and evaluate goal achievement (Plant et al., 2016; Rosewilliam et al., 2011). Whilst no single approach is recommended, the Goal Attainment Scale (GAS) (Kiresuk & Sherman, 1968) is one tool with the capability to set, document and evaluate the extent of achievement of individual goals complimenting other standardised outcome measurement tools (Turner-Stokes, 2009). Kiresuk and Sherman (1968) introduced GAS within a mental health clinical setting. Since its inception, GAS has been widely used across other areas of healthcare including focal spasticity management (Ashford & Turner-Stokes, 2006; Sheean, Lannin, Turner-Stokes, Rawicki, & Snow, 2010; Turner-Stokes et al., 2010; Williams et al., 2020) and adult neurorehabilitation (Hurn et al., 2006; Khan, Pallant, & Turner-Stokes, 2008). GAS has been well-validated and recommended as an adjunct to standardised measures within clinical practice and research (Bovend'Eerdt, Dawes, Izadi, & Wade, 2011; Brock et al., 2009; Joyce, Rockwood, & Mate-Kole, 1994; Khan et al., 2008; Lannin, 2003). Whilst there are many advantages to using GAS, clinicians have also highlighted challenges when using the tool within the clinical setting, most notably the tool's utility (Turner-Stokes, 2009). Clinicians report the tool to be time-consuming, the use of zero and

negative numbers to be discouraging to patients, and limits in recording partial goal achievement despite acknowledged benefits to individualised outcome evaluation and service evaluation (Turner-Stokes, 2009). In response to these clinical challenges, Turner-Stokes (2009) proposed the GAS-Light method, a briefer version that proposed increased clinical utility. There is currently a paucity of evidence regarding GAS-Light that can be used by clinicians, researchers and services in deciding whether to use this tool or the original GAS. This evidence gap will be addressed in Chapter 6 within a repeated measure design comparing GAS-Light performance to GAS performance with adults undergoing neurorehabilitation for upper limb motor impairments. The clinical utility of GAS and GAS-Light is explored in Chapter 7 with clinicians via survey and via measurement of acceptance of GAS-Light by ABI survivors.

1.6. The Lived Experience of Upper Limb Spasticity Following ABI

There has been a global rise in research regarding perspectives of patients / carers / consumers of healthcare and measuring what matters to them (Calvert, Kyte, Price, Valderas, & Hjollund, 2019). The use of patient-rated measures is one method providing an opportunity to engage consumers in the measurement of healthcare outcomes. Patient-rated measures have been shown to add value to more global scales which may lack sensitivity in detecting impairment, limitations and restrictions. Stewart and Cramer (2013) suggest the inclusion of patient-reported outcome measures to obtain greater understanding of motor impairments, particularly when mild deficits after stroke, for example, are often not detected in global standardised measures despite having a significant and long term effect on daily life. Whilst patient-rated measures can provide insightful data, they do not elicit the depth and breadth of impact required to understand the lived experience. In the area of upper limb motor impairment and specifically post-stroke spasticity, there is limited qualitative evidence from the survivor's perspective to understand this lived impact (Esquenazi, 2011; Kerstens et al., 2020). Chapter 2a aims to address this gap by exploring the experience of living with spasticity using a qualitative method to understand the impact on activity and participation.

Despite the paucity of evidence about the experience from the perspective of the survivor, there is an emerging body of relevant qualitative research from which insights can be drawn regarding the impact of spasticity from any cause (Barnes, Kocer, Murie Fernandez, Balcaitiene, & Fheodoroff, 2017; Patel et al., 2020), the experience of post-stroke upper limb impairment (Barker & Brauer, 2005; Purton, Sim, & Hunter, 2020; Waddell, Tabak, Strube, Haire-Joshu, & Lang, 2019) and perceptions of stroke survivors in general regarding recovery and/or neurorehabilitation (Arntzen, Borg, & Hamran, 2015; Becker, 1993; Clarke, 2009; Faircloth, Boylstein, Rittman, Young, & Gubrium, 2004; Kirkevold, 2002; Kitt, Wang, Harvey-Fitzgerald, Kayes, & Saywell, 2016; Lloyd et al., 2018; Luker, Lynch, Bernhardsson, Bennett, & Bernhardt, 2015; Lund, Mangset, Wyller, & Sveen, 2015; Manning, MacFarlane, Hickey, & Franklin, 2019; Pallesen, 2014; Peoples, Satink, & Steultjens, 2011; Salter, Hellings, Foley, & Teasell, 2008; Sarre et al., 2014; Thomas, Allison, & Latour, 2018; Tornbom, Lundalv, & Sunnerhagen, 2019; Torregosa, Sada, & Perez, 2018; Walder & Molineux, 2017).

1.7. A Summary of Knowledge Gaps

The literature presented in sections 1.1 – 1.6 identified three gaps in the evidence base that form the rationale for the study series.

1.7.1. What is the Impact of Upper Limb Spasticity on the ABI Survivor?

Adults who experience ABI, particularly post-stroke, experience upper limb impairments which lead to activity limitations and participation restrictions. Many people do not regain efficient use of their upper limb with lifelong limitations and restrictions endured (Kwah et al., 2013). Unfortunately, there is limited understanding of the impact of upper limb spasticity on their lived experience following ABI, nor on how this influences the ability to engage in everyday activities from the perspective of the survivor. With this significant gap in the literature, it is no surprise that a disparity between outcomes deemed important by ABI survivors engaging in neurorehabilitation with those identified by clinicians has been identified (Duncan Millar et al., 2019). Key attributes of neurorehabilitation, previously defined in this chapter, identify the essential components of measuring the outcomes of individualised neurorehabilitation programs to support ABI survivors to return to living in their own communities. Yet, without fully

understanding the lived experience of impairment, limitations and restrictions in everyday life, valid, reliable and responsive approaches to outcome measurement cannot occur. There is a need to investigate that lived experience.

1.7.2. How Should Goals be Set, Documented and Evaluated?

A further gap raised in this review of the literature is on documenting and evaluating goal achievement. There is suggested benefit to adopt patient-centred goal setting within neurorehabilitation. GAS has been identified as one method to set, document and evaluate goal achievement. However, routine use is hampered by poor clinical utility. Clinical utility of goal attainment scaling methods remain unknown and will be explored in this thesis.

1.7.3. Evidence Gap in Psychometric Properties of Tools

This chapter has also shown that upper limb motor impairments have varying impacts on performance of everyday activities. Research shows that reducing impairments such as spasticity does not automatically improve the ability to perform an upper limb activity, but that a broader view of evaluating impairments and their impact on activity and participation is needed. Therefore, outcome measurement post-ABI must assess across the breadth of the ICF, inclusive of impairment, activity limitations and participation restrictions. Clinicians should be aware of the strengths and limitations of individual outcome measurement tools and ensure tool selection and use are congruent with a patient-centred approach (Donnelly & Carswell, 2002). Currently, clinicians do not have an evidence base to inform their selection of outcome measurement tools, particularly when evaluating activity limitations and participation restrictions for adults with upper limb spasticity post-ABI. This knowledge gap will be explored in this thesis.

1.8. Research Questions

This program of studies will answer the overarching question: Is it possible to accomplish precise and individualised measurement of upper limb activity and goal attainment in neurorehabilitation? Across five studies, the following eight research questions will be addressed:

Question 1: What is the lived experience of upper limb spasticity post-stroke?

Question 2: Can the lived experience of stroke survivors with upper limb spasticity be captured through linking to the ICF?

Question 3: Does the ICF Comprehensive and/or Brief Core Sets for stroke capture and characterise the unique experience of living with upper limb spasticity?

Questions 4: What are the psychometric properties and reported utility of outcome measurement tools used by allied health clinicians working in neurorehabilitation to address upper limb activity and participation goals?

Question 5: What are the psychometric properties and reported utility of the Action Research Arm Test?

Question 6: What are the psychometric properties of GAS-Light when compared to the more well-established GAS?

i. Is GAS-Light valid in an Australian neurorehabilitation setting when determining upper limb activity and/or performance goals?

ii. Is GAS-Light sensitive enough to detect real-life changes and the attainment of goals in adult patients undergoing neurorehabilitation for upper limb motor impairment?

Question 7: What is the clinical utility of GAS-Light when compared to GAS when used by clinicians working with ABI survivor's undergoing neurorehabilitation for upper limb motor impairment?

Question 8: How do ABI survivors experience goal setting and do they accept GAS-Light within neurorehabilitation?

The studies aim to provide evidence that will fill the current gap regarding the experience of living with upper limb spasticity and the impact on everyday life for adults following ABI. Furthermore, evidence will be provided to support the measurement of upper limb activity and participation outcomes via standardised tools, goal setting and goal attainment in clinical practice.

1.9. Outline of the Thesis

Each study is presented in a separate chapter with a final chapter to discuss findings, consider implications and make recommendations. The first study (Chapter 2a) adopts a qualitative approach to understand the perspectives of living with upper limb spasticity after stroke. Thematic analysis was used to identify meaningful concepts shared by stroke survivors. The emerging themes provide the opportunity to gain insight into the areas of activity and participation that may be impacted and to understand the lived experience for a particular group of people post-stroke to guide outcome measurement approaches (Research Question 1).

The second study (Chapter 2b) links interviews completed with stroke survivors to the ICF using established linking rules (Cieza, Fayed, Bickenbach, & Prodinger, 2019; Cieza et al., 2005). The results of this study provide insights into the experience of living with upper limb spasticity through the lens of the ICF to consider the impairments, activity limitations, participation restrictions, environmental barriers and facilitators and personal factors influencing experience. This study will provide insights into whether the Brief and Comprehensive Core Sets for stroke are appropriate for capturing and characterising the lived experience of spasticity (Research Question 2 and 3).

The third study is a systematic review and application of the COSMIN (Mokkink et al., 2010; Prinsen et al., 2018) principles. This study was designed to locate and evaluate the psychometric properties of outcome measurement tools systematically identified by Ashford & Turner-Stokes (2013) to measure active and/or passive function in the context of everyday real-life activities. Such a systematic review has not been completed to date and thus this study aims to contribute to filling voids in the literature pertaining to specific outcome measures and their psychometric properties. This study's protocol is

presented in Chapter 3, and the results presented in Chapter 4 (Research Question 4). Chapter 5 extracts results and presents the psychometric properties and clinical utility of the Action Research Arm Test, a commonly used and recommended measure of activity in clinical guidelines (Research Question 5).

The results of the study comparing GAS and GAS-Light when used with adults with upper limb motor impairment post-ABI are presented in Chapter 6 (Research Question 6). Clinicians' perceptions of the clinical utility of the two methods are explored via clinician surveys and presented in Chapter 7 (Research Question 7). Chapter 7 also presents the experience of patients participating in goal setting and use of both goal attainment scaling methods using patient surveys (Research Question 8). Survey results provide patient perspectives for clinicians regarding outcome measure appropriateness. Both Chapter 6 and 7 provide data that can guide tool selection and goal setting practice within upper limb neurorehabilitation.

Chapter 8 integrates results from all studies and provides insights into the experience of living with upper limb spasticity. This qualitative information coupled with evidence of the psychometric properties of measurement tools and goal attainment methods used to evaluate activity and participation goals in neurorehabilitation, will enhance individualised and precise measurement. Clinical implications and recommendations for research will also be provided.

1.10. Chapter Synopsis

This introductory chapter has introduced the upper limb motor impairments experienced post-ABI and discussed the interplay between motor impairments, activity limitations and participation restrictions. The current challenges when measuring upper limb activity and participation outcomes, and gaps in the psychometric properties of tools used to measure those outcomes have been outlined. A gap in knowledge regarding the experience of living with upper limb spasticity has also been highlighted. Thus, the following chapters present the study series devised to fill the identified knowledge gaps, and inform clinicians in their choice of outcome measurement tools and goal setting practices.

Chapter 2:

The Experience of Living with Upper Limb Spasticity

2.0. Preface to Chapter 2

Great attention has been placed on both the clinical measurement and management of upper limb spasticity following ABI. Comparatively, there has been very minimal enquiry into the experience of living with post-stroke upper limb spasticity from the perspective of the stroke survivor (Esquenazi, 2011; Kerstens et al., 2020). Likewise, the ICF and Core Sets, as introduced in Chapter 1, have been recognised as valuable frameworks to consider and understand the impairments in body functions, activity limitations and participation restrictions that may be experienced following a range of neurological diagnoses, including stroke (Geyh et al., 2004). Of interest, neither framework have been applied to understand the experience of living with post-stroke upper limb spasticity. Hence the applicability of the Core Sets, in particular, remains unknown — the studies presented in Chapter 2 aimed to address these identified knowledge gaps.

Specifically, the aims of Chapter 2 were:

- To explore the perspectives of stroke survivors who had recently completed a structured multidisciplinary neurorehabilitation program for people with chronic upper limb spasticity about their experiences of living with upper limb spasticity. This aim is explored in Chapter 2a.
- To identify the impact of upper limb spasticity on stroke survivors by linking their experience revealed through interviews to the ICF (World Health Organisation, 2001). In doing so, characterising their unique experience and exploring the extent of coverage provided by the Brief and Comprehensive Core Sets. This aim is explored in Chapter 2b.

Qualitative data elicited through enquiry completed within a clinical trial (Lannin, Ada, English, Ratcliffe, & Crotty, 2018; Lannin et al., 2020; Lannin et al., 2020) registered at www.ANZCTR.org.au (ANZCTR12615000616572) informed Chapter 2.

The qualitative data was interrogated via two distinctly different data analysis methods to address study aims and answer research questions and are therefore presented in the following two subchapters, 2a and 2b.

Chapter 2a:

Living with Upper Limb Spasticity: Perspectives of Stroke Survivors

The work covered in this chapter has been submitted for publication as:

Pike, S., Lannin, N.A., Cameron, L., Palit, M., Schneider, E., & Cusick, A. (under review). Perspectives of stroke survivors with chronic upper limb spasticity: a qualitative study revealing contexts and processes of adjustment and adaptation.

See Appendix B for ethics approval, Appendix C for trial registration and Appendix D for supplementary material.

2a.0. Chapter Overview

There has been great clinical attention given to the measurement of upper limb spasticity following an ABI. This attention, however, has primarily focused on quantitative approaches and intervention studies to determine the effects of pharmacological and/or neurorehabilitation interventions. To date, there has been very minimal enquiry into the experience of living with upper limb spasticity. This chapter presents a study that was designed to address this gap in neurorehabilitation research and knowledge.

2a.1. Introduction

Post-stroke upper limb spasticity occurs in up to 46% of stroke survivors at 12 months (Opheim et al., 2014). Those who experience spasticity after stroke experience difficulty in using their limb in everyday activities, limb positioning and comfort, and maintaining optimal condition of the limb (Francisco & McGuire, 2012). While physical impacts of spasticity are well documented, there is limited evidence regarding the experience of stroke survivors living with this condition (Esquenazi, 2011; Kerstens et al., 2020).

Kerstens et al. (2020) enquired about the experienced consequences of chronic post-stroke spasticity with a focus on physical impairments and activity limitations, the experienced effects of botulinum toxin treatment and whether current spasticity management addresses survivor needs. Interviews revealed spasticity-related impairments and activity limitations within categories of stiffness, posture, pain and other sensations, loss of motor control, fatigue and shame, fluctuations in spasticity related to botulinum toxin and the need for professional support and feedback (Kerstens et al., 2020). A further study using qualitative methods with stroke survivors who have spasticity (Levy et al., 2021) explored their views regarding, and factors impacting adherence to an intense neurorehabilitation program involving exercise, their focus was not on their experience of living with/having spasticity. They found adherence could be explained by enablers in motivation (automatic, reflective) and opportunity (social), whilst barriers to adherence were capability (physical) and motivation (reflective). Although there is minimal specific evidence, there is an emerging body of relevant qualitative research regarding: (a) impacts of spasticity from any cause; (b) the

experience of post-stroke upper limb impairment; and (c) perceptions of stroke survivors in general regarding recovery and/or neurorehabilitation.

Clarke (2009) argued that qualitative evidence was needed to "understand the often paradoxical observation that some people are emotionally devastated by stroke-related impairments, whereas others manage to retain their sense of wellbeing in the face of declining function" (p.239). Post-stroke experience from a survivor perspective has been characterised as: an 'unfolding illness' (Kirkevold, 2002); biographical or life course disruption (Becker, 1993; Faircloth et al., 2004); sudden transformation and slow adaptation (Salter et al., 2008); a period of adjustment (Sarre et al., 2014); a time of continuous coping with ongoing issues in the body and self, using strategies including resignation and personal growth (Pallesen, 2014); a process of reintegration back into the community and living a meaningful life (Walder & Molineux, 2017); a threat (Lund et al., 2015); a state of reorientation in life (Tornbom et al., 2019); and a 'struggle' (Arntzen et al., 2015). These themes, generated through research inquiry, are not dissimilar to constructions of meaning derived from survivor 'blogs' by Thomas et al. (2018). They found these blogs revealed social interaction, finding a life purpose, emotional wellbeing, improving function and increasing independence were more important than impairment reduction. Torregosa et al. (2018) also revealed "finding meaning in life" was important, with post-stroke adjustment requiring "time, life goal reconfiguration, willpower, humour, and network support" (p.361).

To sustain neurorehabilitation, stroke survivors engage in maintaining hope and keeping forward momentum (Lloyd et al., 2018). They assume power, with neurorehabilitation enabling empowerment (Peoples et al., 2011). When considering elements of neurorehabilitation that may be important, Luker et al. (2015) identified nine issues of importance to survivors: "(1) physical activity is valued; (2) bored and alone; (3) patient-centred therapy; (4) recreation is also neurorehabilitation; (5) dependency and lack of control; (6) fostering autonomy; (7) power of communication and information; (8) motivation needs nurturing; and (9) fatigue can overwhelm" (p.1694). Attention to these issues helps neurorehabilitation to be informed by survivor experience and preferences. Identification of person-focussed approaches was also identified by Kitt et al. (2016) in

their study to explore moderate to severe stroke survivor perspectives on community-based exercise groups. While Luker et al. (2015) and Kitt et al. (2016) examined stroke survivor perspectives of aspects of neurorehabilitation, Mannering et al. (2019) focussed on the experience of stroke survivors with aphasia using a systematic review. They found stroke survivors with aphasia had a desire to contribute positively to society, motivation to take charge of their condition and wanted to navigate health systems to obtain relevant and collaborative long-term services from aphasia-aware professionals (Manning et al., 2019). Social networks and having social companionship were important enablers.

In the paucity of evidence about the experience of spasticity from a stroke survivor viewpoint, perspectives from people with spasticity from any cause are relevant. Global internet surveys of people living with spasticity aimed to characterise spasticity symptoms and understand their burden and impact on the ability to work, perform daily activities and quality of life (Patel et al., 2020) and explore perceptions of health-related quality of life (Barnes et al., 2017). The authors were interested in the gap between expected and experienced health (Barnes et al., 2017). Self-reports of participants found spasticity had broad adverse impacts on daily life, the ability to work, quality of life, independence and mood (Barnes et al., 2017; Patel et al., 2020). These adverse impacts contrasted with their high expectations of treatment to enable return to work, take care of themselves and be free of muscle spasm (Barnes et al., 2017). While most had received physiotherapy treatment, many reported receiving inadequate information from physicians and long delays between spasticity onset and Botulinum neurotoxin treatments (Barnes et al., 2017). The authors proposed that while health professionals may have a clear understanding of the life impact of spasticity, people who live with it need more information to calibrate their expectations (Barnes et al., 2017).

The few qualitative studies relating to post-stroke upper limb experience do not specifically enquire about or present information on spasticity. Indeed in the studies located, the term ‘spasticity’ is not mentioned in the text or in participant descriptions. But since spasticity may be a cause of upper limb impairment or disability, findings of these studies provide relevant background:-

- Barker and Brauer (2005) explored what carers and stroke survivors (0.4 to 13 years post-stroke) with upper limb impairments think recovery is, what influences it, and what can maximise it. While none of their findings were specific to spasticity (and spasticity was not reported in quotes or derived concepts), their findings raised similar themes to those discussed in spasticity-specific studies - physical loss and disruption, the tension between losing and keeping hope, and the need to strive for change.

- Perspectives of stroke survivors with upper limb hemiplegia in the first six months of recovery were explored by Waddell et al. (2019). Again, while they did not mention spasticity in the article, the key findings mapped well to the spasticity-specific studies in that belief, motivation and confidence in upper limb recovery remained consistently high in the first six months, independent of clinical factors.

- Purton et al. (2020) investigated perspectives of people with upper limb dysfunction after stroke. They found that upper limb dysfunction brought: (a) “an altered way of life” in personal care, meaningful and valued activities and meaningful life roles and relationships; and (b) a “disrupted self” with feeling devalued, disrupted self-image and changes in identity.

These three studies, specific to the experience of stroke survivors with upper limb impairment or dysfunction, have similarities with literature presented earlier regarding life with spasticity from any cause and stroke survivor perspectives in general. Perspectives of post-stroke survivors regarding their experience of upper limb spasticity is limited. To advance the implementation of evidence-informed person-centred neurorehabilitation and community care, “insider” perspectives of stroke survivors who have upper limb spasticity is needed. This study aims to address this gap.

2a.2. Study Aim

This research explored perspectives of stroke survivors who had recently completed a structured multidisciplinary neurorehabilitation program for people with chronic upper limb spasticity, about their experiences of living with upper limb spasticity.

2a.3. Methods

The study used a qualitative approach to understand what the experience of having upper limb spasticity was like from the perspective of adults whose spasticity was caused by a stroke. Experience was interrogated using content analysis, independent of any theoretical frame, to reveal themes describing lived experience from the point of view of participants. The framework to report the study method and results in this paper was informed by the *JBI Critical Appraisal Checklist for Qualitative Research* (Joanna Briggs Institute, 2017). The study was nested in a clinical trial which was approved by the relevant Human Research Ethics Committees and is registered at www.ANZCTR.org.au (ANZCTR12615000616572).

2a.3.1. The Research Conceptual Framework

A prospective conceptual framework was not adopted to interrogate data. The aim was to understand participant experience from their perspective; thus a phenomenological approach was adopted.

2a.3.2. The Researchers

The research team were occupational therapists and a rehabilitation physician, all with extensive neurorehabilitation experience. They practice and are registered in Australia. One member of the research team provided the motor training program, another member collected the interview data. The lead investigator (thesis author) engaged directly with the data through the management and analysis processes with assistance and supervision from her thesis co-supervisor. All research team members reviewed the analysed data providing feedback to contribute to the final results. All research team members aided in the interpretation of results, which were led by the thesis author.

2a.3.3. Sampling

The sample was drawn from a population of patients who had been discharged from a neurorehabilitation program (Lannin et al., 2018; Lannin et al., 2020). These people were identified to have the attributes needed to answer the research question. They each had chronic stroke, upper limb spasticity and an upper limb motor training neurorehabilitation experience. To be included in the trial program from which the sample was drawn, participants had to meet the following inclusion criteria: greater than three months post-stroke; scheduled to receive a botulinum toxin-A injection to muscle(s) that crossed the wrist; not currently receiving upper limb neurorehabilitation; and having cognition with normal range (<5 adjusted errors on the Short Portable Mental Status Questionnaire (Pfeiffer, 1975)). All participants received the same motor training program prior to their interview (Lannin et al., 2018). This involved face-to-face contact with treating physicians, occupational therapists and physical therapists.

All participants in the trial who had consented to interviews were considered eligible. Using purposive sampling, 12 participants were selected to ensure the following attributes were included in the interview participant pool: sex (male/female), productivity role (employed/ retired/ unemployed/ home duties) and time post-stroke (more recent/more chronic). These attributes were selected so insights could be gained from people in a variety of life situations. The number of people invited to interview was limited by the amount of project funding available to pay for data collection and transcription.

2a.3.4. Study Context

Participants lived in the community and had completed a 12-week motor training program that was designed for per-protocol implementation (Lannin et al., 2018; Lannin et al., 2020). All participants received a botulinum toxin-A injection as part of the program prior to commencing motor training. Physicians selected muscles for injection and dose based on the distribution of spasticity and patient goals generated through a structured goal setting process. This motor training program was implemented in an outpatient clinic setting and included participants' self-practice in their home environment. Occupational therapists and physical therapists provided patient

education, serial casting (two weeks), movement training aiming to decrease weakness (electrical stimulation and progressive resistance exercises) and exercises to improve active movement (10 weeks) (Lannin et al., 2020).

2a.3.5. Data Collection

Following completion in their motor training program, qualitative interviews were conducted by one of the research team members, a registered occupational therapist with neurorehabilitation expertise in the clinic (n=5), by phone (n=3), at the person's home (n=1) and at the local community library (n=1). The interviewer had not been involved in the delivery of the neurorehabilitation program. One interview was held with each participant and this lasted from 26 to 69 minutes (mean 33 minutes, SD 14.66; median 33.5 minutes). Participants determined the length of the interview. The interviewer could probe answers to questions to clarify points made by participants.

2a.3.6. Materials

A semi-structured interview guide (see Appendix D) was used to invite participants to share their experience of 'what it is like having spasticity'. Other questions asked about: when they perceived their spasticity had started; what they noticed about spasticity as it developed; what they had done previously to manage it; impacts on their roles, responsibilities and ability to do things. Participant age, gender, living situation, spasticity severity, pre-stroke hand dominance, affected upper limb and time since neurorehabilitation program completion was noted by the interviewer.

2a.3.7. Data Management

Interview audio recordings were transcribed by an independent person specifically employed for this task. Interview data was de-identified prior to use by the research team and pseudonyms allocated. Demographic and clinical data for each participant was matched by the independent person to the pseudonym and entered into an Excel™ spreadsheet. Deidentified transcripts were uploaded into NVivo software version 12 (QSR International, 2018) using participant pseudonyms by the thesis author. Beginning

concepts were coded on transcripts and as meaning threads emerged these were set out in a text table.

2a.3.8. Data Analysis

Participant characteristics were summarised. Thematic analysis of transcripts sought to synthesise experience through identification of meaning units for individuals, and using open coding approaches, developing theme categories that described experience for the cohort. Given the brevity of some of the transcripts and the post-program context of interviews, it was accepted that construction of a rich and contextualised description of experience would not be possible. Analysis thus aimed to identify themes characteristic of participant experience rather than trying to construe relationships between concept categories and meaning units.

To do this, first, interview data was uploaded into NVivo software version 12 (QSR International, 2018) and data was tagged using common words from the data into topic categories. In discussion, the thesis author and her co-supervisor developed meaning units and through iterative engagement with data sought to capture concept categories that described participant experience of living with post-stroke spasticity. The transcripts were interrogated using content analysis of words and phrases to identify mutually exclusive and internally consistent categories to build a tentative meaning framework. The thesis author and her co-supervisor iteratively refined the meaning framework, deepening theoretical sensitivity by revisiting qualitative evidence previously presented in the literature to identify experience and perceptions in our data that had been previously identified and that experience which was new to these participants. Preliminary findings were presented to the research team for their perspectives on the authenticity of the framework in capturing their understanding of participant experience. Participants were not consulted. Findings were then considered in light of existing qualitative evidence of stroke survivor experience with similarities and differences presented in the Discussion.

2a.4. Results

Of 12 participants invited, 10 agreed to participate (9 males: age 20 -77 years, median 49 years, one female: 50 years). One participant, 'Lee', had his spouse 'Val' also participate in the interview with him. Table 2a.1 presents participant characteristics and pseudonyms. Of particular note is that the pre-stroke dominant upper limb was affected in four of the 10 participants and all had spasticity, ranging from slight (1) to severe (3). All participants had very low levels of arm and hand use in their affected arm (indicated by Box and Blocks Test scores), suggesting that these participants were significantly limited in their upper limb function, not just due to spasticity but also motor control. All participants had either mildly reduced or intact cognition, and two experienced some level of expressive aphasia.

2a.4.1. Interviewer Data

Transcripts revealed little variation in interviewer wording of questions or prompts. Words spoken by the interviewer which elicited affirmative or negative responses by people with expressive aphasia were included in data reports.

2a.4.2. Thematic Analysis Key Finding

Participant data revealed lived experience was framed by two organising concepts – lived experience in *context* and *continuous* adaptive processes. The experience of living with upper limb spasticity starts with a stroke but it never ends. It is continuous adaptation to new contexts - their body, their life situation and health services - using processes of expecting, learning, practising, evaluating, listing and committing. These contexts and processes are now elaborated.

Experience is contextualised in the body, in time, in a life situation and in relation to health and home care services. Stroke survivors with upper limb spasticity *contextualised* their experience in the body (body parts – 'my' and 'the'; upper limb movement; the feeling of spasticity in muscles; upper limb sensation including pain; upper limb strength); in time (time since stroke; time since spasticity was noticed; time taken); in a specific life situation (roles in social networks, home environment and arrangements); and in services and technology access and usage (health services;

community and home care services; and assistive technology). Each of these is now elaborated with data examples.

2a.4.3. Body

Participants spoke about parts of their body including shoulder (Che), arm (Tom), forearms (Joh), hand (Bob), palm (Kim), fingers (Kit) and thumb (Zac). Participants referred to body parts using first person pronoun “my left arm” (Zac), “my shoulder, my hands and fingers” (Joh) indicating connection with self, as well as signal-words to indicate body parts were separate to self, for example “the arm” (Lee); “the palm” (Kim); “the left hand” (Kit); “put it in another area rather than just push it” (Bob); “I have to grab it and get it under control so it’s normal” (Tom). One participant personified the body part (e.g. “Trevor”, was the name given by Kim to her arm).

Participants described having upper limb movement; “It only moves when I yawn ... because I’ve got a reflex action” (Che); “my arm starts to tremor” (Tom); “I think I’ve got movement in my upper arm, yeah left arm, but nothing in my hand” (Zac), or lack of movement “I knew I had no movement… try and get me to just move a little” (Bob); “I couldn’t move it” (Zac); “to help me move my arm” (Dan).

Only three of the participants used the word “spasticity” in the interviews (Che, Joh and Tom; e.g., “I’ve still got minor spasticity” (Joh). But most described either what it felt like when noticeable, “tightness … it got cramped” (Che); “tighter” (Kim); “automatically squeezing the thing” (Kit); “tremor” (Tom); “tight” (Zac) or when it was less pronounced, “it got better in bed it used to relax” (Bob); “not always so tight” (Kim); “but now it’s relaxed and stuff” (Dan).

Three participants described upper limb sensation experiences. This involved: sensation returning over time, “when I first woke up from the stroke, my arm, right arm was not numb, but then it started getting more sensation going into it” (Che); increased sensitivity, “especially it's worse if I put - like if I ran some cold water goes on it - it feels like it's burning. Or if something hot falls on it, that's when it really kills” (Che); and

Table 2a.1. *Demographic and clinical characteristics of participants [pseudonyms]*

Participant	Kim	Tom	Ted	Kit	Che	Bob	Joh	Dan	Lee	Zac
Demographic Characteristics										
Sex	F	M	M	M	M	M	M	M	M	M
Age *(yrs)*	50	52	38	62	41	77	38	20	67	49
Highest education level	U	H	U	H	U	H	U	H	H	H
Living situation	Alone	Alone	With others	Alone	With others	With others	With others	With others	With others	With others
Years since stroke on day of interview	8.5	9.2	4.4	11.1	3.8	4.9	6.1	2.5	6.0	2.1
Pre-stroke hand dominance	Right	Right	Right	Right	Right	Right	Right	Right	Right	Right
Clinical Characteristics										
Side of hemiplegia	Left	Left	Right	Left	Left	Right	Right	Right	Left	Left
Prior BTX-A Intervention	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Uses walking aid or assistance	No	No	Yes	No	No	Yes	Yes	No	No	No
Neglect*	0	1	0	0	0	0	1	0	1	0
Box and Blocks Score - Baseline	0	0	1	1	0	0	1	0	0	0
Box and Blocks Score - 3 mo	0	0	3	8	0	2	18	0	0	0
Baseline Spasticity	3	3	3	1	2	2	2	2	1	3
Sensation**	1	1	1	1	1	0	0	1	2	2
Cognitive measure***	0	0	0	3	1	3	0	2	1	0

F: Female, M: Male, H: High School, U: Undergraduate, BTX-A: Botulinum Toxin – A, *0: nil neglect, 1: slight neglect, 2: severe neglect, mo: months, **0: normal sensation, 1: impaired sensation, 2: no sensation, ***number of errors on Short Portable Mental Status Questionnaire

limited sensation, “Just cold … if something cold touches it I feel it … [and] I can feel pain” (Zac).

Three participants mentioned upper limb strength. Bob described himself as strong “I’m still quite stronger than I was”, Joh described a change in upper limb strength since the stroke “It was just one of the dormant muscles but with no strength” and Zac quantified his strength “Apparently I've got a seven-kilogram grip on my left hand, which I had only a four-kilogram grip when I started”.

2a.4.4. Time

All participants, except Kit, identified time since stroke, with six recalling the time in years, two identifying approximate dates and one recalling time in months. Kit, Tom and Lee did not identify the time when they first noticed spasticity, but for the others this was described in relation to: the stroke itself, “it started that moment [of the stroke] I can remember the first week. My hand was paralysed but also shaking” (Joh); “It’s always been there since the stroke” (Zac); a date “I think February of 2016” (Dan); and in relation to the stroke-care continuum “Oh probably when I was in [name of rehabilitation hospital] post-stroke ... I think I was in [name of acute hospital] for about three weeks, firstly in intensive care after the first operation… Then I was sent to rehab there at [name of rehabilitation hospital]” (Ted). Other mentions of time include time taken to: travel to and from therapy “Because I don’t have a car, I couldn’t drive so he had to bring me in, so it was a big ask from him you know?” (Zac); time to practice therapy tasks “putting aside time every day to make sure you did the exercise … like it probably took a good hour a day to bowl it over” (Tom); “Lee spent hours and hours of training” (Val); the frequency of therapy sessions over time “once a month. Used to be weekly, then it got pushed out to fortnightly now the physio just feels that once a month is suffice to see me” (Tom); and time in relation to living situation “we dated a month before the stroke” (Joh); “my father passed away two years ago” (Kim); “were just about to leave [our old house]” (Bob).

2a.4.5. Life Situation - Roles

Participants identified roles in social networks of family, friends and community. Each is now explored with data examples. In family, participants identified partner-roles of being a *single individual,* "as a single person" (Kit); *fiancé,* "my girlfriend soon to be my wife" (Joh); *spouse,* "[my wife] and I we do what we want" (Bob); *divorcee,* "I separated with my wife" (Zac) and *widow,* "my partner [died] oh about 10 weeks ago" (Kim). Other family roles were also identified including *mother,* "my two children" (Kim); *father,* "the kids took over" (Bob); "my youngest son pops in on a regular basis so I catch up with him" (Tom), "my three kids are … 12, 10 and 8" (Zac); *adult-child in contact with parents*, "I talk to my parents once a week" (Joh); *adult-child living with and dependent on parents*, "[my mum does] Just general stuff like showering and stuff like that" (Dan); "My parents help me to get dressed every morning " (Zac); and *sibling,* "I talk to my sisters twice a week" (Joh); "my brothers if I see them I stop and say hello" (Kim).

Friends were mentioned by three participants. Kim referred to a "best friend" and "friends" visiting her. Kit had a "mate" he went fishing with and Tom said "[I have] not a lot of friends. I've got a mate who drops around every now and again".

Community productivity and leisure roles were identified by six participants. Past productivity roles were builder (Bob), school student (Dan) and baggage handler (Zac). One participant had employment – this was Joh who worked as a disability support person – and Bob contributed to the household by going shopping. Bob was the only person to identify a past leisure activity which was cooking for the family on Sundays. Current leisure roles included being a fly-fisherman with other mates (Kit) and getting manicures (Kim).

2a.4.6. Life Situation – Home Environment and Arrangements

Four participants describe significant changes to where and who they lived with after stroke compared to where and who they lived with before the stroke. This involved family breakdown, moving to parents to receive daily assistance and changing homes because maintenance was no longer possible. These are now explored with data examples.

- Zac discussed family breakdown and interstate relocation following his stroke; "…separated with my wife and my three kids are still [in the warmer state]" where he had been living. The move interstate was a major change because he had been there "…for the last 20 years" and after the stroke and family change, he had to live with his parents who help him "get dressed every morning" and with "meals, cooking, showering".

- Dan also experienced relocation: "I was living alone, yeah. I was finishing high school" but he returned home to live with his mum who helped him with "general stuff like showering" after his stroke.

- Bob described changes to his living situation post-stroke, "I thought our place … I was very good when I was able to do much, but she [wife] can't do too much, and I can't. So I suggested that I buy the other half [of my daughter's property to live there]".

- Kim relocated post-stroke following the death of her partner, "…the house I've moved into … he died so he decided … I could live there because it's a much bigger house".

Changes to living situation also occurred in ways participants did not perceive were stroke related. Joh's relationship with his girlfriend progressed from pre-stroke separate dwellings, "We dated a month before the stroke" when he was living with his sister, to post-stroke cohabiting "So she has been living with me for about four years" and their plans to marry.

Other participants shared information about living alone and support they received from family, "My daughter, when she's home she'll cook for me or sometimes she gets me out of the shower to get me onto the bed and things like that" (Kim). Tom described living alone and a lack of support, "my youngest son pops in on a regular basis so I catch up with him. I've got no real family for support" (Tom).

2a.4.7. Services and Technology Access and Usage

Nine participants identified accessing or trying to access health services, for example "I'm still seeing the spasticity clinic" (Che), "I tried to get physio" (Tom), "outpatient therapy" (Ted). A number of specific health professions/ professionals were used by five participants – "occupational therapist" (Che, Kit, Joh), "physio" (Tom), "the neurologist" (Joh), "I asked lots of doctors and neurologists too" (Zac). Two commented that health professionals provided "guidance" (Zac) and an "opportunity for me to call someone and ask was there too, so that was good" (Zac).

Five participants described accessing home care services and the types of assistance they received with home-based activities. The services are now identified with examples: "we've both got a help package [laundry]" (Bob), to access the community; "I have a home help sort of person and I can use them up to nine hours a week and I decide what we're going to do" (Kim); "I have one or I had two carers, but now it's just the one…they usually just go out with me to eat and stuff like that" (Dan); "I have a support worker that comes and picks me up and takes me wherever I want to go… yeah pretty much getting out and about. If I want to go to the bank or shopping" (Zac), or support with a combination of home and community based-activities; "I have a PCA [personal care assistant] come once a week and takes me shopping and paying bills and that. Then she'll come back on the day after on the Friday and cleans for me" (Tom).

Participants described assistive technology usage to move around "I've been wearing an AFO [Ankle-Foot Orthosis]" (Zac); "I've got a little cart I sometimes use [to go to shops], I zoom off around there sometimes" (Bob); "sitting in a wheelchair" (Joh). Other participants discussed assistive technology used as therapy aides, "Electric stimulation" (Zac); "yeah an e-stim – I've got one of those" (Che); "electrode machine" (Lee). Kim described the way she used household items to assist with the management of her spasticity, "Well of a night I have a paint roll holder or whatever, and I sleep with it of a night so that it's out… It sort of, for me, to get my palm out" and Zac described having "…certain clamps and procedures that I do things with".

Living with upper limb post-stroke spasticity involves continuous adaptation through the processes of expecting, learning, practising, evaluating, listing and committing

Living with post-stroke spasticity involved *processes* implemented in contexts previously described; expecting (what recovery might and might not be possible; what neurorehabilitation might achieve); learning (about stroke; about neurorehabilitation; and about intervention information and skills); practising (exercises; stretches; doing tasks a certain way); evaluating (functional changes; change in impairments; effect of interventions; impacts of spasticity; change then, now, in the future); listing (what I cannot do since the stroke; what I can do now; ADL and IADL); and committing (to the new me; to my enduring personal qualities; to being useful; to helping other stroke survivors; to making time/putting time in; to accessing and using resources, and inability to commit).

2a.4.8. Expecting

Participants described expectations along a continuum of what recovery might be possible, "I was expecting to get movement back in my hand, arm" (Zac) to expectations for no further recovery "I didn't expect much because I know that now that it is a slow process" (Kit). Three participants described uncertainty about expectations of recovery or possible recovery "[starting trial] I was just hoping something would happen. I [had] tried to get physio as often as I can because I thought well something might work sooner or later" (Bob); "…when I first got asked I was pretty keen to try anything to get any sort of improvement in my movement" (Tom); "we were looking for the light at the end of the tunnel to hopefully achieve something" (Lee and Val).

Participants described their expectations of what neurorehabilitation might achieve "All I want from out of it was to hold a cup, an empty cup, and put it to my lips and put it down – that was all I wanted nothing more than that" (Kim); "…if the fingers would just grasp something so he could use the hand even 10 per cent or 20 per cent" (Lee and Val); "My goal was improving my hand function" (Ted); "…using the left hand to do

things for – in my case was just using it to open up the door" (Kit); "I was hoping that I could be able to carry something in the hand" (Tom).

2a.4.9. Learning

Some participants demonstrated that they had learnt and understood medical and neurorehabilitation terms about stroke because they used it in describing their own experience. For example, "because I've got a *reflex action*" (Che); "I had a disease in my heart and it caused a big stroke so I'm still suffering *expressive aphasia* and I think *minor spasticity*" (Joh). In addition participant pre-stroke knowledge about stroke was used to explain experience, for example "I thought oh, this couldn't be a stroke - because *I know about strokes* from – in teaching" (Kim). Participants acknowledged their own learning about neurorehabilitation, "I started to *understand* much better what I was supposed to do. I found then the exercise, well seemed to be really good, real better… I *was doing it properly*" (Bob); "I *know now* that it is a slow process" (Kit). Participants also described the importance of learning as part of neurorehabilitation intervention "because if somebody is with me I got the motivation plus the technicality of it. That means if I do something wrong, I've got [interviewer says 'feedback'] yes, yes, yes" (Kit); "…they taught me how to do it properly" (Zac); "I reckon I know the theory with Botox" (Ted) and "I know all about Botox" (Tom).

2a.4.10. Practising

Participants discussed practicing exercises including stretches before, within and after completion of the trial. These are now explored with examples. Two participants identified they did their own physical exercises specifically to work on their upper limb before the trial: "Pretty much going to the gym five days a week. I'm not an amateur gym goer but I have a specific program with strength, power lifting, gymnastics" (Joh); "I did some stretches prior to this" [trial] (Zac). During the trial, three participants reported many hours were spent practising, "But anyway, he just - he spent hours and hours of training and - or we spent hours of training. It was not a problem" (Lee and Val). After the trial, trial exercise practice continued, "I think we continued on with a lot of the practice of the exercises for months afterwards, but then we eventually stopped". Whilst others wanted to continue to practice but rather independently sought new exercises

and ways to do it, "So, I went to the gym and found, by Google, exercises" (Joh). Bob thought that it was his persistent practise to use his upper limb that brought improvement, "…at the end [of the trial] I started to see things happening" so "that's why I still move" and "…keep trying something else… instead of just when I do the exercises". The perception of practice being linked to improvement was also perceived by Kit and Tom: "Let me say I find out the way to switch it off in my - that means because that is automatically when I put something between two of them, what I was squeezing no matter what I want it or not. After simulation and all that stretching exercises it helped me [bait his fishing hook] (Kit); "I really push myself on my walking side of things, you know to walk as far as I can and that makes me feel better (Tom).

2a.4.11. Evaluating

Participants evaluated aspects of their lived experience. This included change in functional ability to perform tasks, severity and presence of impairments, impacts of neurorehabilitation interventions, and how they felt about the spasticity. These are now explored with data examples.

Functional change in the ability to perform a task was evaluated by participants in household tasks, "It takes me five minutes versus a one minute normal person, but I take my time [changing bed linen]" (Joh); leisure, "I was going fishing. I knew I couldn't do that what I used to do – that means fly fishing. You need both hands to do it…I couldn't control my left" (Kit); and self-care, "By the time you did a complete shower and get dressed again, and that sort of thing, by the time it's [laughs] – yeah, you're tired" and "…it's too difficult. I'll do it [cook] again when I can" (Bob).

Participants evaluated change in their observed and perceived impairments, including: changes to their upper limb spasticity, "I think its [the spasticity] actually the same, mild or a little bit actually decreased" (Joh); "its [the spasticity] just basically been the same…yeah its [the spasticity] plateaued out" (Zac); "my arm was up here most of the time, but now its relaxed" (Dan); changes to their upper limb strength, "But in my arm it was just so weak before and it is now a little bit of strength" (Joh); "My arm still probably, pretty much hangs in its own place" (Tom); and changes to their upper limb

sensation “I noticed when I first woke up from the stroke, like my arm, right arm was, not numb, but then it started getting more sensation going into it” (Che).

Participants evaluated the effect of interventions. There was generalised functional improvement, “doing that exercises I find out myself I was going for that [door] handle using the left hand automatically” (Kit); “I think that everything helps. I think this is a step in my progress” (Zac); and specific improvement from targeted effort, “I’m improving from always stretching” (Che). Interventions could also be identified as effective in maintaining an outcome, “I think it’s maintaining, and so I do every day a stretch, not to increase the motion, but just to combat the spasticity because if not it will get like a claw and that will be a burden” (Joh). Some interventions were identified as not being effective, “…it [botulinum toxin injection] really didn’t do anything” (Kim). Two participants did evaluate interventions but were not able to draw strong conclusions about it, “I think steady improvement. I think, can’t tell, or maybe can tell. Very hard to tell with me” (Ted); “it’s a bit hard to tell…like it is what it is. It’s hard for me to pick any improvement in anything now” (Tom).

Two participants expressed feelings about the spasticity itself – rather than feelings about having spasticity. Joh said “spasticity for me is a mild distraction…. It’s like an insult” and Tom described the spasticity as “a real bloody nuisance sometimes…”.

Evaluation happened at time points and over time. There was change then, “in the hospital I couldn’t move it” (Zac); “…when I went [shopping] to put my right hand out … I couldn’t, I could only put my left hand out” (Kim); “it felt like I didn’t do anything anymore” (Bob). There was change now, “I couldn’t do anything, now I can do something and that’s good. I keep trying something else” (Bob)**;** “But lately when I started slowly, slowly I find out it [upper limb] works… That means I could control” (Kit); and thoughts about change or no change in the future, “Generally I sort of got to a point where I accepted the fact now I am the way I am and that I’m not going to get any better” (Tom); “He had done his fantastic best. It was just great. So we conceded defeat, but it wasn’t in a bad way. It’s ‘well we’ve done this and this is what we’ve achieved and

maybe one day down the track a miracle might happen', and we left it at that [laughs]" (Lee and Val).

2a.4.12. Listing

When responding to the prompt enquiry about how spasticity affected their everyday lives, most participants listed activities of daily living tasks they could not do since the stroke. For example, "She [personal care assistant] does pretty much the things I can't do like change the sheets on my bed and makes it, vacuums for me and mop floors and clean the bathroom" (Tom); "…trying to get my hand open and get something into my hand is difficult" (Tom); "my parents help me to get dressed every morning" (Zac); "I couldn't drive, I couldn't work, I couldn't get around" (Zac). They also listed what they can now do, "I can draw"(Bob); "using my clothes is much better than it was"(Bob); "I can walk, I can drive I can do almost every normal activity on my own" (Joh); "I do all my own washing I do all my own cooking and wash my dishes and everything" (Tom); "I go shopping and I walk up and down every aisle of the supermarket…I can walk a whole city block" (Tom).

2a.4.13. Committing

Participants described an intellectual and emotional process of committing to a continued but different intrapersonal life. Tom committed himself to the new me, "I accept the fact now that I am the way I am and that I'm not going to get any better. There's going to be no miracle cure and the best thing I can do is improve my stamina and my fitness, which I do now" (Tom). Joh committed himself to the person he was and is with to enduring personal qualities, "I am super independent, I'm stubborn on that" … "I have a self-independence of spirit" … "I'm the same man. My personality didn't change, my life has definitely changed". Bob did this too, sharing how he lost feeling useful as he did before the stroke, "the kids took over things that I used to do for them or for me automatically…", but he committed to getting improvements in function so he could be as useful as possible, for example helping with shopping, "I can go to the shops… I zoom off around there sometimes and buy some things" and dressing himself without help, "that's why I'm pleased with the - see, that's the first double thing that I've been able to do" [manage trouser zip]). For Zac and Joh they described a commitment to

helping other stroke survivors, “I feel like I was doing something to help. Like, you know, for future people that had a stroke, and if it does help them, that's fantastic” (Zac).

Intrapersonal commitment required commitment of time and effort. Some participants did this while others recognised they did not. Participants identified they needed to commit to making time for neurorehabilitation or have key support people in their lives make that commitment too. For example, participants needed to make time for neurorehabilitation interventions, “I was pretty keen on the idea and I probably threw myself at it to give it the best chance possible” (Tom); “putting aside time every day to make sure you did the exercise” (Tom); “he spent hours and hours of training” (Lee and Val), or daily activities, “I'm changing the sheets, or the doona covers. It takes me five minutes versus a one-minute normal person, but I take my time” (Joh). Support people also needed to commit time, “it was a big ask for my dad to come in here every day” (Zac).

Kit described the inability to commit without support, despite his interest, “I needed something like this, like attending some physiotherapy sessions in here. That’s practically 100 per cent of it because I know myself I wouldn’t do it at home”, “I am that person when I don’t have a whip under my hat, that means I do everything except what I [am] supposed to do” (Kit). Kim also described an inability to commit but for her it was to pre stroke roles, “Then when I had my stroke, it was like I couldn’t be bothered thinking about it [preparing family Christmas lunch], what to do and anything”.

2a.5. Discussion

This study explored the perspectives of stroke survivors who had recently completed a structured multidisciplinary neurorehabilitation program for people with chronic upper limb spasticity. The aim was to reveal their perspectives on the experience of living with upper limb spasticity. We analysed and synthesised experience using content analysis to create meaning threads. Key findings from this study were that stroke survivors with spasticity organised experience using two overarching constructs of ‘contexts’ and ‘processes’. These are similar to findings of previous stroke-survivor qualitative research. In relation to contexts, our findings are similar to other studies that reveal the “situated”

nature of spasticity which occurs in the body, affecting the everyday life and lifestyle of a person in a unique physical and social environment. In relation to processes, our participants, like others in previous research, found the experience of being a stroke survivor is one without end. Spasticity is an inherent part of surviving where continuous adaptation is required using a range of strategies to adapt to the "journey of discovery I never wanted to take" (Thomas et al., 2018). Our study specified and characterised those processes from the perspective of stroke survivors with upper limb spasticity.

2a.5.1. Contexts

Findings from this study support previous research identifying the significant influence of the social and physical environments on the experience of living with a stroke. Stroke survivors with upper limb spasticity clearly articulated the meaningful activities they could no longer participate in or complete independently and revealed the practical, emotional and psychological support needed from immediate family, community-based support services and health professionals. This finding reflects previous research from stroke survivors with spasticity (Kerstens et al., 2020) without spasticity (Purton et al., 2020) and adults with spasticity from any cause (Barnes et al., 2017). Whilst this study identified a need for support, carer stress and burden were not discussed nor raised as a concern by participants in comparison to that revealed in previous research including adults with spasticity (Ganapathy et al., 2015). Language used by stroke survivors to refer to spasticity in this study also reflected previous findings from research with adults with spasticity from both post-stroke and non-stroke samples (Barnes et al., 2017; Bhimani, McAlpine, & Henly, 2012; Kerstens et al., 2020). Stroke survivors commonly describe the feeling of spasticity within the muscles rather than label the impairment 'spasticity' and as revealed in this study by one participant (Joh), survivors often don't know the term spasticity nor understand what it means. This insight reinforces previous findings and recommendations for healthcare professionals to have knowledge of and use the word choice preferred by stroke survivors to describe spasticity to enhance understanding and communication (Bhimani et al., 2012).

Participants provided further insights into the embodied experience not yet illuminated in previous research. Palleson (2014) found survivors five years post-stroke experienced

a vulnerable, unreliable and inconvenient body and Purton et al. (2020) revealed 'the disrupted self' with subordinate themes of disrupted self-image and changes in identity for stroke survivors attributed to upper limb dysfunction. However, those findings do not encapsulate the sense of disconnect to the affected limb and the adoption of a separate identity for the affected upper limb that emerged from this study. This unique finding within the embodied experience of upper limb spasticity requires deeper exploration in future research.

2a.5.2. Processes

Our participants articulated processes previously identified within themes of life course disruption (Faircloth et al., 2004), the struggle (Arntzen et al., 2015) and coping with loss (Purton et al., 2020), adapting to change (Shipley, Luker, Thijs, & Bernhardt, 2018) and building identity (Lou, Carstensen, Jørgensen, & Nielsen, 2017; Pallesen, 2014). Furthermore, stroke survivors in this study shared experience that holds similarities with previous research identifying the importance to 'keep the door open' (Barker & Brauer, 2005), to keep forward momentum and maintain hope (Lloyd et al., 2018) for current and future recovery despite stroke chronicity or the current level of dependence on others to carry out daily activities. Experience elicited from our study supports fostering autonomy through neurorehabilitation to regain control over one's life (Luker et al., 2015) and the importance of interactions with accessible and trusted health professionals to learn and direct one's own recovery (Manning et al., 2019). Autonomy and a sense of control for our participants was fostered through participants own learning and evaluation. Learning; about stroke, about the neurorehabilitation process, how to adapt or carry out activities differently and the active self-evaluation of spasticity onset, changes in spasticity severity over time, functional abilities and changes or perceived lack of change to their abilities and effect of therapy on their spasticity and upper limb use. Participants in our study also revealed a desire to contribute positively to society (Manning et al., 2019). Participants were willing to commit to a time and energy consuming trial not only for desired personal benefit, but to assist health professionals to learn from them to build stroke knowledge to, in turn, help other stroke survivors.

The exploration data using content analysis revealed context, process and mind-set meaning threads unique to the participants which provides new insight into the experience of having upper limb spasticity after stroke. At the same time, the findings reflect concepts previously identified in qualitative research about stroke survivor experience in general, the experience of stroke survivors with upper limb dysfunction, and the experience of people who have spasticity from any cause. Our findings suggests that while stroke survivors with spasticity who have completed a multidisciplinary upper limb neurorehabilitation program are like others, on a never-ending journey, they have a suite of processes to draw on as part of continuous adaptation to static and changing contexts within and outside their bodies. This may make the difference between people feeling like passengers in a journey they did not want to go on or drivers in the harsh landscape of what life threw at them.

This study provides a beginning insight into the experience of stroke survivors with upper limb spasticity. Findings from this study focus attention on the importance and desire voiced by survivors with spasticity to learn and adapt consolidates current recommendations to ensure stroke survivors are provided opportunity and resources to support active learning and management of their condition and neurorehabilitation. This finding is balanced with ensuring access to enabling support from family, community services and health professionals. This study emphasises the necessity for healthcare professionals to ensure their selection and use of language to label and define spasticity matches the preferences of survivors. This is an identified current gap in clinical practice and research, with a potential risk of suboptimal communication or missed opportunity to understand the impact of upper limb spasticity on the survivor. Healthcare professional are also presented with a further challenge of preserving hope for this cohort of stroke survivors. Despite a strong theme of hope, participants were also reserved in their expectations for recovery which must be considered in light of current challenges to increase upper limb use for chronic stroke survivors with spasticity (Lannin et al., 2020). More research should seek to understand the experience of this important, but neglected, cohort of stroke survivors. This would not only to help inform providers of person-centred care but, since spasticity is a chronic condition with significant

impairment and disability risks, it may help empower people with spasticity to keep using strategies found to be helpful by others who have journeyed before them.

When considering our findings, we acknowledge that these participants had particular attributes which provide uniquely informed insights about upper limb spasticity. They had completed a 12-week program focussing on post-stroke upper limb neurorehabilitation for people with spasticity. Their perspectives may have been framed by their program experience which used strategies including goal setting, pharmacological and physical interventions, training and practice, and relationships with health professionals. Our participants may thus have more understanding about the location, nature and trajectory of upper limb spasticity with and without administered interventions and self-directed management. That said, it is noteworthy only three participants used the word “spasticity” in their interviews. All study participants had very low levels of arm and hand use, therefore their experience and ability to participate in everyday life activities was likely also influenced by their reduced motor control and not just specific to spasticity.

There were limitations in this study. First, the study context was a possible limitation. It was conducted after the completion of a neurorehabilitation program where having upper limb spasticity was an inclusion criterion and an explicit focus of the intervention. Participants had interventions that enhanced their understanding of their spasticity. In one sense this could be construed as a limitation because they are likely to have had superior exposure to spasticity terminology, neurorehabilitation strategies directed towards spasticity management, and interactions with professionals about upper limb spasticity when compared with most post-stroke survivors. At the same time this ‘limitation’ could be a strength of the study because the enquiry focus was a topic area they were familiar and comfortable with as participant ‘experts’. Secondly, sampling was a limitation. It was conducted prospectively using diversity of participant demographic attributes as criteria. These attributes were selected on face value rather than informed by a theoretical framework. Data sampling within the participant group was not possible because interviews were held on only one occasion, findings should therefore be seen as indicative.

2a.6. Conclusions

This study presents new information that illuminates the experience of stroke survivors with upper limb spasticity after recent completion of a targeted neurorehabilitation program. Findings indicate that experience is contextualised in time, their body and in their life situation of environments and roles. Further findings suggest the experience of life with upper limb spasticity following stroke is a dynamic one with multiple processes in play and a mindset of expectation fuelled by hope. These findings are complementary to prior qualitative evidence of stroke survivor experience and the experience of people with spasticity from any cause. Future research using longitudinal, in-depth, repeated interviewing and triangulation with other data sources is needed to see whether themes seen in this and prior qualitative research resonate in other samples and settings. Future research using a variety of approaches, could explore whether patterns emerging across qualitative studies have potential for interrogation as propositions that may be generalisable to other samples and settings.

2a.7. Chapter Synopsis

Chapter 2a presented findings from a qualitative enquiry, illuminating what adults living with post-stroke upper limb spasticity experience, specifically the areas of everyday life they perceive to be most impacted due to this impairment. As discussed in both Chapter 2a and Chapter 1, understanding the effect of upper limb spasticity directly from the survivor's perspective was previously limited. Whilst this study contributes to this currently neglected area, it revealed a further knowledge gap: 'Does the complete ICF (World Health Organisation, 2001) and individual Core Sets for stroke (Geyh et al., 2004) capture this lived experience?' Chapter 2b presents a study that was conducted to answer this research question.

Chapter 2b:

Linking the Lived Experience to the ICF

The work covered in this chapter has been published as:

Pike, S., Lannin, N. A., Cameron, L., Palit, M., & Cusick, A. (2021). Chronic stroke survivors with upper limb spasticity: linking experience to the ICF. *Disability and Rehabilitation,* doi: 10.1080/09638288.2021.1894490

See Appendix B for ethics approval, Appendix C for trial registration and Appendix E for supplementary material.

2b.0. Chapter Overview

Consistent with neurorehabilitation literature, this thesis has applied the ICF as a valuable framework within which to understand impairments in body functions, activity limitations and participation restrictions that may be experienced by people with upper limb motor impairments following ABI. Chapter 1 introduced the ICF and Core Sets relevant to clinical populations with ABI, one of which is stroke. A number of studies have evaluated the applicability and validity of the ICF Core Sets for stroke. However, the applicability of the Core Sets for stroke to survivors with upper limb spasticity is yet to be considered. This chapter presents a study that contributes to this evidence-based gap.

2b.1. Background

Stroke is a leading cause of long-term disability and economic burden worldwide. Up to 70% of survivors experience upper limb sensorimotor impairments as a result of their stroke (Lawrence et al., 2001) with many continuing to live with persistent long-term impairments that impact the use of affected limbs in everyday activities (Nakayama, Jørgensen, Raaschou, & Olsen, 1994). Spasticity, as presented in Chapter 1, is one post-stroke impairment with reported prevalence as high as 87% (Malhotra et al., 2008). Spasticity often adversely impacts the ability to use and care for the affected upper limb, can lead to contracture development and pain, and contribute to increased carer burden (Ganapathy et al., 2015; Sunnerhagen, Olver, & Francisco, 2013).

Despite the high incidence of upper limb spasticity following stroke and the extent and seriousness of potential impacts on function and quality of life, there has been little qualitative research on the way upper limb spasticity affects the lives of survivors (Barnes et al., 2017). The ICF (World Health Organisation, 2001) has been used as a conceptual framework to understand potential impacts of health conditions, including stroke. As discussed in Chapter 1, the application of the ICF to this area may shed important light on the ongoing experience of stroke survivors who experience spasticity.

In stroke neurorehabilitation the ICF Core Set for stroke (Geyh et al., 2004) has been applied a number of times in different clinical and community samples to validate the taxonomy itself. Studies with stroke survivors living in Finland (Paanalahti et al., 2014) and Sweden (Paanalahti et al., 2013) largely confirmed the content validity of the Comprehensive version of the Core Set for stroke. Further, studies with stroke survivors living in Germany (Glässel et al., 2012), Brazil (Riberto, Lopes, Chiappetta, Lourenção, & Battistella, 2013), China (Wang et al., 2014) and Sweden (Algurén et al., 2010) largely confirmed the content validity of the extended version of the Core Set for stroke, but recommendations were made to reduce some of the categories in some studies. Other validity studies have supported the content of the extended Core Set for stroke from the perspectives of: male versus female stroke survivors (Glässel et al., 2014) and health professionals, including physicians (Lemberg et al., 2010), occupational therapists (Glässel et al., 2010) and physical therapists (Glässel et al., 2011). The impact of stroke on younger survivors living in Sweden has been examined through application of the extended Core Set for stroke (Snögren & Sunnerhagen, 2009), and the impact of stroke on older Australian women (Tavener et al., 2015) via application of the Brief Core Set for stroke. None of these validity or impact studies considered clinical attributes of stroke survivors themselves in relation to core set validity or applicability.

While the development of the ICF Core Set for stroke has provided a framework specific to describe impairments in body structures and function that arise after stroke (Geyh et al., 2004), it's applicability to stroke survivors with upper limb spasticity is yet to be explored. All of the study reports previously cited (Algurén et al., 2010; Glässel et al., 2012; Glässel et al., 2014; Glässel et al., 2011; Glässel et al., 2010; Lemberg et al., 2010; Paanalahti et al., 2014; Paanalahti et al., 2013; Riberto et al., 2013; Snögren & Sunnerhagen, 2009; Tavener et al., 2015; Wang et al., 2014) had gaps in the reporting of participant characteristics from the perspective of the person living with spasticity; thus no understanding of the prevalence or contribution of post-stroke upper limb spasticity to the patterns identified in the ICF Core Sets was possible.

Although applicability of the ICF and ICF Core Sets for stroke survivors with spasticity specifically has not been conducted, the ICF has been used in studies with people

undergoing upper limb focal spasticity neurorehabilitation (Eftekhar, Mochizuki, Dutta, Richardson, & Brooks, 2016; Nott, Barden, & Baguley, 2014; Turner-Stokes et al., 2010). These studies linked participant neurorehabilitation goals to the ICF framework, not specifically to the stroke Core Set, to identify areas of importance and to evaluate the effect of neurorehabilitation on goal achievement across ICF domains and categories. Goal types were linked to 12 (Eftekhar et al., 2016), 16 (Turner-Stokes et al., 2010) and 18 (Nott et al., 2014) second-level categories included in the Comprehensive Core Set for stroke, concentrated in the Activities and Participation domain (Communication, Mobility, Self-Care, Domestic Life, Community/Social Chapters) followed by the Body Functions domain (Sensory and Pain, Neuromusculoskeletal Chapters). Whilst these studies provide insight into goal types using ICF categories, they do not necessarily capture the impacts and the experience of upper limb spasticity from a personal rather than goal directed perspective.

2b.2. Study Aim

To date the experience of stroke survivors with upper limb spasticity has not yet been linked to ICF categories to reveal impacts from their perspective. The aim of this study, therefore, was to identify the impact of upper limb spasticity on stroke survivors by linking their experience revealed through interviews to the ICF (World Health Organisation, 2001). In doing so their unique experience will be characterised and the extent of coverage provided by the Brief and Comprehensive Core Sets will be explored.

2b.3. Research Questions

1. Can the lived experience of stroke survivors with upper limb spasticity be captured through linking to the ICF?

2. Does the Comprehensive and Brief ICF Core Sets for stroke capture and characterise the unique experience of living with upper limb spasticity?

2b.4. Methods

6.4.1. Study Design

To understand the impacts of upper limb spasticity following stroke, qualitative methodologies of purposive sample selection, semi-structured interviews and structured content-analysis were employed. In general terms our study design replicated that used by Paanalahti et al. (2013), but our purpose was to reveal participant experience through the ICF rather than validate ICF framework concepts. The study was nested in a clinical trial exploring the impact of a protocol based upper limb neurorehabilitation intervention (Lannin et al., 2018; Lannin et al., 2020) which was approved by university and hospital Human Research Ethics Committees (see Appendix B), and was registered at www.ANZCTR.org.au (ANZCTR12615000616572) (see Appendix C).

2b.4.2. Participants

The sample was drawn from a population of patients who had been discharged from a structured multidisciplinary neurorehabilitation program (Lannin et al., 2018; Lannin et al., 2020). These people were identified to have the attributes needed to answer the research question. They each had chronic stroke, upper limb spasticity and they had all been enrolled in and completed the same neurorehabilitation program which included interventions specifically targeting upper limb impairments and limitations. To be in the neurorehabilitation program, participants had to meet the following inclusion criteria: greater than three months post-stroke; scheduled to receive a botulinum toxin-A injection to muscle(s) that crossed the wrist (in accordance with the Pharmaceutical Benefits Scheme); not currently receiving other upper limb neurorehabilitation; and having good functional cognition, which was deemed to be <5 errors on the Short Portable Mental Status Questionnaire (Pfeiffer, 1975).

All participants who completed the trial at the Melbourne site, spoke conversational English, and did not experience severe aphasia were considered eligible. Using purposive sampling, 12 of 33 eligible participants were selected to include the following attributes in the interview participant pool: gender (male/female), productivity role (employed/ retired/ unemployed/ home duties), and longevity of survival post-stroke (more recent,

more chronic). All participants agreeing to participate in the interviews provided written, informed consent to interviews prior to commencing.

6.4.3. Study Context

All participants were community dwelling and had completed a 12-week neurorehabilitation program (Lannin et al., 2018; Lannin et al., 2020). All participants received a botulinum toxin-A injection as part of the program prior to commencing multidisciplinary therapy. Physicians selected muscles for injection and dose based on the distribution of spasticity and patient goals generated through a structured goal setting process. The goal setting process did not use the ICF or ICF terminology. The neurorehabilitation program was implemented in an outpatient clinic setting and participants' home environment. Occupational therapists and physiotherapists provided patient education, serial casting (two weeks), movement training aiming to decrease weakness (electrical stimulation and progressive resistance exercises) and exercises to improve active movement (10 weeks).

6.4.4. Data Collection

Interviews were conducted by a member of the research team, a registered occupational therapist with neurorehabilitation expertise (not the thesis author) in the clinic (n=5), by phone (n=3), at the local community library (n=1) and at the person's home (n=1). The interviewer had not been involved in delivery of the neurorehabilitation program. One interview was held with each participant and this lasted from 26 to 69 minutes (mean 33 minutes, SD 14.66; median 33.5 minutes). Participants determined the length of interview. The interviewer could probe answers to questions to clarify points made by participants. Data were audio recorded, transcribed and de-identified to participant pseudonyms prior to viewing by the whole research team.

6.4.5. Materials

A semi structured interview guide (see Appendix D) was used to invite participants to share their experience of 'what it is like having spasticity'. Follow up probe questions included when they perceived their spasticity had started; what they noticed about

spasticity as it developed; what they had done previously to manage it; impacts on their roles, responsibilities and ability to do things. The interviewer did not use ICF terms, was unaware at the time that data would later be linked to the ICF, and interview questions were developed prior to ICF linkage.

Participant clinical and demographic data was extracted from the neurorehabilitation program trial data base: age; gender; hand-dominance; date of stroke; education level; living situation; history of prior botulinum toxin-A injections; side affected with spasticity; and results of the following standardised tests administered at baseline: Box and Block Test (Platz, Pinkowski, van Wijck, & Johnson, 2005) (administered at baseline and 3 months), Functional Independence Measure motor subscale (Keith, Granger, Hamilton, & Sherwin, 1987), Line Bisection Test (Schenkenberg, Bradford, & Ajax, 1980), and the Short Portable Mental Status Questionnaire (Pfeiffer, 1975). The level of spasticity at baseline was measured using the Tardieu Scale and reported as a score of 0 to 4, where 0 is no spasticity (Patrick & Ada, 2006). The time since neurorehabilitation program completion was noted by the interviewer.

6.4.6. Data Management

Interview audio recordings were transcribed by an independent person specifically employed for this task. Interview data was deidentified prior to use by the research team. Demographic and clinical data for each participant was matched by the independent person to the pseudonym and entered into an Excel™ spreadsheet.

6.4.7. Data Analysis

Demographic and clinical data were presented for each participant and aggregated to characterise the sample. To understand the impact of spasticity on functioning disability and health, participant experience data was mapped to the ICF, initially using the refined ICF Comprehensive Core Set for stroke (Geyh et al., 2004). Data that could not be mapped to categories within the Core Set were linked to other categories in the ICF as required. The Brief Core Set for stroke was mapped by extracting relevant items from the Comprehensive Set so that gaps and overlaps between the two could be identified. Different levels of the ICF are presented (with codes) in this paper using different font-

styles to help readers navigate the classification hierarchy as follows: **Domains, *Chapters***, *second-level* and *third-level categories*.

The ICF ten linking rules (Cieza et al., 2019; Cieza et al., 2005) were used with minor adaptation as follows:-

1. A sentence or phrase in the data was identified as a 'meaningful concept' unit.

2. The meaningful concept unit was allocated to the ICF **Domain**, then ***Chapter*** using a *second* and if possible *third-level category*. The allocation was made that best included the unit concept and excluded other potential meanings.

3. Text concepts were read on 'face value' without imputing meaning. For example, "I do all my own washing, I do all my own cooking and wash my dishes and everything" was the meaningful concept text. The domain selected for this text was **d Activity and Participation**; the chapter within the domain for the text was ***d6 Domestic Life***; and the second-level category for the text was *Doing housework (d640)*.

To record the ICF linking, an Excel™ spreadsheet was constructed by investigators (thesis author and co-supervisor) that had separate tabs for each domain. The **Activity and Participation** domains were considered a single fully overlapping list of categories, as recommended by WHO (World Health Organisation, 2013). Within each tab sheet, rows were prospectively populated with relevant ICF ***Chapters***, *second* and *third-level* categories. Columns identified the participant and the meaningful concept data extracted. Each meaningful concept was recorded separately even if named by the same participant. The final column was the sum of the number of meaningful concepts within each row.

Data linking was led by the thesis author. Linked data was iteratively presented to her co-supervisor for review and following discussion a consensus was reached which allocated most meaningful concept units to one of the ICF Comprehensive Core Set for

stroke categories, to 'Personal Factors' or 'unable to be linked'. The consensus map and unallocated units were presented to the rest of the research team for feedback and for discussion and allocation of remaining units. Following this a consensus final map and visual summary of ICF ***Chapter*** and *second* or *third-level* category coverage was prepared. Meaningful concept units within the Brief Core Set as compared to the Comprehensive Core Set were identified.

2b.5. Results

Of 12 participants invited, 10 agreed to participate (9 males: age 20 -77 years, median 49 years, one female: 50). One participant, Lee, had his spouse Val also participate in the interview with him. Table 2b.1 presents participant characteristics and pseudonyms. Participants were between 2.1 and 11.1 years post-stroke with the majority living with others. The pre-stroke dominant hand was affected in 4 of the 10 participants. Of particular note is that all had spasticity, ranging from slight to severe, and very low levels of upper limb activity in their affected arm (greater than six standard deviations below the normative means per age bracket (Mathiowetz, Volland, Kashman, & Weber, 1985) for the Box and Block test (Platz, Pinkowski, van Wijck, & Johnson, 2005)), most had a somatosensory impairment and only three had a sensory neglect. All participants had either mildly reduced or intact cognition.

2b.5.1. Participant Data

Participant interview data linking results are presented as narrative text (below) and within Table 2b.2 using a similar approach as Paanalahti et al. (2013). Meaningful concepts were linked to **Body Function** (41%), 31% to **Activities and Participation**, 26% to **Environmental Factors** and 2% to **Body Structures**. Nine meaningful concepts were linked to Personal Factors and 10 were not eligible for linking. Of the 10 participants 6 had no meaningful concepts mapped to **Body Structure** (Ted, Kit, Bob, Dan, Lee, Zac) and only one participant made no mention of concepts that could be mapped to **Activity and Participation** (Che). All participants had concepts that could be mapped to **Body Function** and **Environment**. Domains will be presented in ICF order. Concepts that could not be linked are then presented.

Table 2b.1. *Demographic and clinical characteristics of participants [pseudonyms]*

Participant	Kim	Tom	Ted	Kit	Che	Bob	Joh	Dan	Lee	Zac
Demographic Characteristics										
Sex	F	M	M	M	M	M	M	M	M	M
Age *(yrs)*	50	52	38	62	41	77	38	20	67	49
Highest education level	U	H	U	H	U	H	U	H	H	H
Living situation	Alone	Alone	With others	Alone	With others	With others	With others	With others	With others	With others
Years since stroke on day of interview	8.5	9.2	4.4	11.1	3.8	4.9	6.1	2.5	6.0	2.1
Pre-stroke hand dominance	Right	Right	Right	Right	Right	Right	Right	Right	Right	Right
Clinical Characteristics										
Side of hemiplegia	Left	Left	Right	Left	Left	Right	Right	Right	Left	Left
Prior BTX-A Intervention	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Uses walking aid or assistance	No	No	Yes	No	No	Yes	Yes	No	No	No
Neglect*	0	1	0	0	0	0	1	0	1	0
Box and Blocks Score Baseline	0	0	1	1	0	0	1	0	0	0
Box and Blocks Score 3 months	0	0	3	8	0	2	18	0	0	0
Baseline Spasticity	3	3	3	1	2	2	2	2	1	3
Sensation**	1	1	1	1	1	0	0	1	2	2
Cognitive measure***	0	0	0	3	1	3	0	2	1	0

F: Female M: Male H: High School, U: Undergraduate, BTX-A: Botulinum Toxin – A, *0: nil neglect, 1: slight neglect, 2: severe neglect, **0: normal sensation, 1: impaired sensation, 2: no sensation, ***number of errors on Short Portable Mental Status Questionnaire.

Table 2b.2. *ICF Comprehensive Core Set for stroke second-level categories*

ICF Code	ICF Category Title	Meaningful concepts mapped	
		Included category (n=) M/F	Additional category (n=) M/F
	Body Functions n = 178		
Chapter 1: Mental Functions			
b110	**Consciousness functions**		
b114	**Orientation functions**		
b117	**Intellectual functions**		
b126	**Temperament and personality functions**	**10/0**	
b130	**Energy and drive functions**	**35/3**	
b134	**Sleep functions**		
b140	**Attention functions**	**1/1**	
b144	**Memory functions**	**2/0**	
b152	**Emotional functions**	**11/1**	
b156	**Perceptual functions**		
b164	**Higher-level cognitive functions**	**13/0**	
b167	**Mental functions of language**	**2/0**	
b172	**Calculation functions**		
b176	**Mental functions of sequencing complex movements**		
b180	**Experience of self and time functions**	**3/2**	
Chapter 2: Sensory functions and pain			
b210	**Seeing functions**		
b215	**Functions of structures adjoining the eye**		
b260	**Proprioceptive function**	**1/0**	
b265	**Touch function**	**4/0**	
b270	**Sensory functions related to temperature and other stimuli**	**2/0**	
b280	**Sensation of pain**	**3/0**	
Chapter 3: Voice and speech functions			
b310	**Voice functions**		
b320	**Articulation functions**		
b330	**Fluency and rhythm of speech functions**		
Chapter 4: Functions of the cardiovascular, haematological, immunological, and respiratory systems			
b410	**Heart functions**		
b415	**Blood vessel functions**		
b420	**Blood pressure functions**		
b455	**Exercise tolerance functions**	**7/0**	
Chapter 5: Functions of the digestive, metabolic, and endocrine systems			
b510	**Ingestion functions**		
b525	**Defecation functions**		
Chapter 6: Genitourinary and reproductive functions			
b620	**Urination functions**		

ICF Code	ICF Category Title	Meaningful concepts mapped	
		Included category (n=) M/F	Additional category (n=) M/F
b640	**Sexual functions**		
Chapter 7: Neuromusculoskeletal and movement-related functions			
b710	**Mobility of joint functions**	**14/2**	
b715	**Stability of joint functions**	**1/0**	
b730	**Muscle power functions**	**12/0**	
b735	**Muscle tone functions**	**14/4**	
b740	**Muscle endurance functions**		
b750	**Motor reflex functions**	**2/0**	
b755	**Involuntary movement reaction functions**	**1/0**	
b760	**Control of voluntary movement functions**	**22/2**	
b765	*Involuntary movement functions*		*3/0*
b770	**Gait pattern functions**		
	Body Structures n = 9		
Chapter 1: Structure of the nervous system			
s110	**Structure of brain**		
Chapter 4: Structures of the cardiovascular, immunological, and respiratory systems			
s410	**Structure of cardiovascular system**	**3/0**	
Chapter 7: Structures related to movement			
s720	**Structure of shoulder region**	**1/0**	
s730	**Structure of upper extremity**	**3/0**	
s750	**Structure of lower extremity**	**1/1**	
	Activities and Participation n = 137		
Chapter 1: Learning and applying knowledge			
d115	**Listening**		
d138	*Acquiring information*		*1/0*
d155	**Acquiring skills**		
d160	**Focusing attention**		
d166	**Reading**		
d163	*Thinking*		*6/1*
d170	**Writing**		
d172	**Calculating**		
d175	**Solving problems**	**1/0**	
Chapter 2: General tasks and demands			
d210	**Undertaking a single task**	**1/0**	
d220	**Undertaking multiple tasks**		
d230	**Carrying out daily routine**	**4/0**	
d240	**Handling stress and other psychological demands**	**1/0**	
Chapter 3: Communication			
d310	**Communicating with - receiving - spoken messages**		

ICF Code	ICF Category Title	Meaningful concepts mapped	
		Included category (n=) M/F	Additional category (n=) M/F
d315	**Communicating with - receiving - nonverbal messages**		
d325	**Communicating with - receiving - written messages**		
d330	**Speaking**	**4/0**	
d335	**Producing nonverbal messages**		
d345	**Writing messages**		
d350	**Conversation**	**2/0**	
d360	**Using communication devices and techniques**		
Chapter 4: Mobility			
d410	**Changing basic body position**	**1/0**	
d415	**Maintaining a body position**		
d420	**Transferring oneself**		
d430	**Lifting and carrying objects**	**2/1**	
d440	**Fine hand use**	**11/1**	
d445	**Hand and arm use**	**21/0**	
d450	**Walking**	**9/0**	
d455	**Moving around**		
d460	**Moving around in different locations**	**2/0**	
d465	**Moving around using equipment**		
d470	**Using transportation**	**1/1**	
d475	**Driving**	**5/0**	
Chapter 5: Self-Care			
d510	**Washing oneself**	**6/1**	
d520	**Caring for body parts**	**0/1**	
d530	**Toileting**		
d540	**Dressing**	**8/0**	
d550	**Eating**	**2/1**	
d570	**Looking after one's health**		
Chapter 6: Domestic life			
d610	*Acquiring a place to live*		*1/1*
d620	**Acquisition of goods and services**	**4/0**	
d630	**Preparing meals**	**5/1**	
d640	**Doing housework**	**3/0**	
d650	*Caring for household objects*		*1/1*
Chapter 7: Interpersonal interactions and relationships			
d710	**Basic interpersonal interactions**		
d720	*Complex interpersonal interactions*		*3/0*
d750	**Informal social relationships**		
d760	**Family relationships**	**1/2**	
d770	**Intimate relationships**	**1/0**	
Chapter 8: Major life areas			
d820	*School education*		*1/0*

ICF Code	ICF Category Title	Meaningful concepts mapped	
		Included category (n=) M/F	Additional category (n=) M/F
d845	**Acquiring, keeping and terminating a job**	**1/0**	
d850	**Remunerative employment**	**1/0**	
d855	**Non-remunerative employment**		
d860	**Basic economic transactions**		
d870	**Economic self-sufficiency**	**2/0**	
Chapter 9: Community, social, and civic life			
d910	**Community life**		
d920	**Recreation and leisure**	**10/3**	
	Environmental Factors n = 115		
Chapter 1: Products and technology			
e110	**Products or substances for personal consumption**	**1/0**	
e115	**Products and technology for personal use in daily living**	**19/3**	
e120	**Products and technology for personal indoor and outdoor mobility and transportation**	**3/0**	
e125	**Products and technology for communication**		
e135	**Products and technology for employment**		
e150	**Design, construction and building products and technology of buildings for public use**		
e155	**Design, construction and building products and technology of buildings for private use**		
e165	**Assets**		
Chapter 2: Natural environment and human-made changes to environment			
e210	**Physical geography**	**3/1**	
e225	*Climate*		*2/0*
Chapter 3: Support and relationships			
e310	**Immediate family**	**18/4**	
e315	**Extended family**		
e320	**Friends**	**2/2**	
e325	**Acquaintances, peers, colleagues, neighbours and community members**	**1/0**	
e340	**Personal care providers and personal assistant**	**8/1**	
e355	**Health professionals**	**19/0**	
e360	**Other professionals**	**1/0**	
Chapter 4: Attitudes			
e410	**Individual attitudes of immediate family members**		
e420	**Individual attitude of friends**		
e425	**Individual attitudes of acquaintances, peers, colleagues, neighbours and community members**		

ICF Code	ICF Category Title	Meaningful concepts mapped	
		Included category (n=) M/F	Additional category (n=) M/F
e440	**Individual attitudes of personal care providers and personal assistants**		
e450	**Individual attitudes of health professionals**		
e455	**Individual attitude of health-related professionals**		
e460	**Societal attitudes**		
Chapter 5: Services, systems, and policies			
e515	**Architecture and construction services, systems and policies**		
e525	**Housing services, systems and policies**		
e535	**Communication services, systems and policies**		
e540	**Transportation services, systems and policies**		
e550	**Legal services, systems and policies**		
e555	**Associations and organizational services, systems and policies**		
e570	**Social security services, systems and policies**		
e575	**General social support services, systems and policies**		
e580	**Health services, systems and policies**	**24/3**	
e590	**Labour and employment services, systems and policies**		

M: male, F: female, **Bold** categories are included in the Comprehensive Core Set for stroke, *italicised* are categories required but not included in the Comprehensive Core Set for stroke

2b.5.2. Body Function

The **Body Function** domain captured the majority of meaningful concepts; n=178 meaningful concepts were linked to 20 of the 41 *second-level* categories included in the Comprehensive Core Set for stroke. N=1 additional second-level categories outside of the core set were required to link experience. The majority of data fell within the two Chapters of ***Mental Functions (b1)*** and ***Neuromusculoskeletal and Movement-Related Functions (b7).*** A very small number of concepts linked to ***Sensory Functions and Pain (b2)*** and ***Functions of the Cardiovascular, Haematological, Immunological and Respiratory Systems (b4)***. There were no concepts linked to the remaining four chapters (***Chapters*** 3, 5, 6 and 8). Due to the large number of Body Function Chapters, paragraph subheadings have been used to structure this section.

Body function – Global Mental Functions: There are eight second-level categories in Global Mental Functions, six of these are in the Core Set for stroke. *Energy and Drive Functions (b130)* was most populated with data; with its third-level category *Motivation (b1301)* having by far the most meaningful concepts identified across nearly all participants (n=8). *Motivation* was revealed as both an enabler "because if somebody is with me I got the motivation" (Kit); and a hindrance "it was like I couldn't be bothered thinking about it, what to do and anything" (Kim). The other third-level category of *Energy and Drive* Functions, *Energy Level (b1300)* had no data. In other core set second-level categories, *Temperament and Personality Functions (b126)* had limited data "I think I'm quieter because of the stroke" (Joh).

Body function – Specific Mental Functions: In Specific Mental Functions, there are 14 second-level categories and nine are in the core set. Six attracted participant data. Most meaningful concept data was generated for *Higher Level Cognitive Functions (b164)*, presented positively "my comprehension has moved up a huge amount" (Bob), and negatively "I knew I couldn't do that what I used to do" (Kit) and *Emotional Functions* (b152) with participants revealing spasticity to be, "a real bloody nuisance sometimes" (Tom), and a "frustrating" and "tough" (Zac) experience. *Experience of Self and Time (b180)* revealed "so by this time, that's why I have Trevor [name given to the impaired arm] (Kim). Other *Specific Mental Functions* which had data included: *Attention* (b140) "concentrate" (Kim), *Memory (b144)* "I don't remember the first couple of days" (Joh) and *Language (b167)* "I had to think about a word because I couldn't say it" (Bob).

Body functions – Sensory Functions and Pain: When considering ***Sensory Functions and Pain (b2)***, there are 18 second-level categories and six are included in the core set. Data represented Additional Sensory Functions and Pain. In the former, participants discussed *Proprioceptive Function (b260)* "Facilitator: What about knowing where your arm is without looking at it? Interviewee: No." (Zac), and varying *Touch Function (b265)* "I've got feeling" (Che), "I can't feel my left arm put it that way" (Zac) and *Sensory Functions Related to Temperature and other Stimuli (b270).* In Pain, the first level of *Sensation of Pain (b280)* captured data.

Body Functions – Cardiovascular and other systems: When considering **Functions of the Cardiovascular, Haematological, Immunological and Respiratory Systems (b4),** four of the sixteen second-level categories are included in the core set. Meaningful data was linked to Additional Functions and Sensations of the Cardiovascular and Respiratory Systems Exercise Tolerance Function (b455) second-level category " I get tired, but not exhausted" (Bob) and third-level category General Physical Endurance (b4550) "the best thing I can do is improve my stamina and my fitness" (Tom).

Body Functions – Neuromuscular and movement-related: When considering ***Neuromusculoskeletal and Movement Related Functions (b7),*** nine of the 17 second-level categories are in the ICF core set. Data was linked to all second-level categories included in the core set, and some to third-level categories, excluding *Muscle Endurance Functions (b740)* and *Gait Pattern Functions (b770)*. Meaningful data linked to Functions of the Joints and Bones primarily populated the *Mobility of joint functions (b710)* second-level category. Participants commented on "range of motion" (Kit, Joh) and stretching the arm "stretch it out and straighten it out" (Tom), "I just keep doing stretching as much as I can" (Zac). Meaningful data mapped to Muscle Function populated two second-level categories:-

1. The first, Muscle Power Function (b730), was revealed in a positive way "I've been trying to get the strength" (Bob), "But in my arm, it was just so weak before and it is now a little bit of strength" (Tom) and negative way "I have no strength in my arm" (Joh). Data was also mapped to the third-level categories, Power of Muscles of One Limb (b7301) "my hand was paralysed" (Joh) and, Power of Muscles of One Side of the Body (b7302) "paralysed down the left side" (Zac).

2. Secondly, meaningful data mapped to the second-level category *Muscle Tone Functions (b735),* named "spasticity" (Joh, Tom), "minor spasticity" (Joh) and described the phenomena "tight", "tighter" (Kim, Che), "tightness" (Dan), "it got better when I was in bed it used to relax" (Bob) and "fingers became very loose" (Kim).

The *Control of Voluntary Movement Functions (b760)* second-level category was also revealed in positive and negative ways. Participants discussed the loss or reduced ability to control movement, "I know I use all my shoulder rather than my arm" (Bob), "my hand didn't work properly" (Kit). Participants also revealed positive changes to control that they experienced "…it starts to work. That means I could control [the fingers]" (Kit), "he can move his arm up a little bit" (Val). *Involuntary Movement Reaction Functions (b755)* was required to link of participant data related to a lack of arm swing reducing "balance" when walking (Tom).

Involuntary Movement Functions (b765), not included in the Comprehensive Core Set, was required to link participant data relating to an involuntary tremor "my arm starts to tremor. I have to grab it and get it under control so it's normal" (Bob).

2b.5.3. Body Structure

The **Body Structure** domain captured a minority of meaningful concepts with a total of 9 identified across 4 of the 5 second-level categories. Nil additional categories outside of the Comprehensive Core Set were required. Meaningful concepts fell within the two Chapters of ***Structures of the Cardiovascular, Immunological and Respiratory Systems (s4)*** and ***Structures Related to Movement (s7)***. There were no concepts linked to the remaining six Chapters (Chapters 1, 2, 3, 5, 6 and 8).

When considering ***Structures of the Cardiovascular, Immunological and Respiratory Systems (s4)*** there are five second-level categories, only one of those, *Structure of Cardiovascular Systems (s410)*, is included in the Comprehensive Core Set for stroke. Data was linked from one participant to that category, more specifically to the third-level categories *Heart (s4100)* "repairing the heart" (Che) and *Arteries (s4101)* "aortic dissection" (Che).

When considering **Structures Related to Movement (s7)** there are 9 second-level categories, with 3 of those, Structure of the Shoulder Region (s720), Structure of Upper Extremity (s730) and Structure of Lower Extremity (s750), included in the Comprehensive Core Set for stroke. A very small number of meaningful concepts were

mapped to those categories. The majority of participant data, using the coding rules, was considered to hold meaningful concepts related to function, thereby mapped to **Body Functions** chapters, or activity, thereby mapped to **Activity and Participation** chapters, rather than anatomical structures.

2b.5.4. Activity and Participation

The **Activity and Participation** domain captured the second largest number of meaningful concepts with a total of 137 identified across all nine chapters. Data was linked to 27 of the 51second-level categories included in the Comprehensive Core Set with an additional 6 second-level categories not included, required for linking. Most meaningful concepts fell within the ***Mobility (d4)*** Chapter, followed by ***Self-Care (d5)*** and ***Domestic Life (d6)*** Chapters. A very small number of concepts linked to ***Learning and Applying Knowledge (d1)***, ***General Tasks and Demands (d2)***, ***Communication (d3)***, ***Major Life Areas (d8)***, ***Interpersonal Interactions and Relationships (d7)***, and ***Community Social and Civic Life (d9)***. Due to the large number of Activity and Participation Chapters, paragraph subheadings have been used to structure this section.

Activity and Participation – Learning and Applying Knowledge: Within ***Learning and Applying Knowledge (d1)***, there are 26 second-level categories with seven of those included in the Comprehensive Core Set. Only one of those seven, *Solving Problems (d175)* was required, “[Facilitator] So, you've got adaptive ways around things. [Interviewee] That's right, yes. I've had to” (Zac). Two second-level categories not included in the core set were required; *Acquiring Information (d138)* to link the experience of learning, and *Thinking (d163)*, "I didn’t expect much because I know that now that it is a slow process" (Bob).

Activity and Participation – General Tasks and Demands: Meaningful concepts were linked to three of the six second-level categories within the ***General Tasks and Demands (d2)*** chapter. *Undertaking a Single Task (d210)*, "It takes me five minutes versus a one minute normal person, but I take my time [changing sheets]" (Joh), C*arrying out Daily Routine (d230)*, “but doing anything during the day, nothing” (Joh), "So, everything I've been doing, I've had to re-adjust. I do things differently these days" (Zac) and *Handling*

Stress and other Psychological Demands (d240), “I used to do it [cooking] when I was working, to stop thinking about work” (Bob).

Activity and Participation – Communication: Meaningful data was linked to two of the 18 second-level categories within the ***Communication (d3)*** chapter, both categories are included in the core set. Data mapped to *Speaking (d330)* revealed “I couldn’t talk at all” (Bob) whilst data mapped to C*onversation (d350)* revealed “I talk to my parents once a week. I talk to my sisters twice a week. My fiancée every day” (Joh).

Activity and Participation – Mobility: When considering ***Mobility (d4)***, there are 21 second-level categories, 12 of those categories are included in the ICF core set, eight of those were populated with data. The majority of meaningful concepts were linked to Carrying, Moving and Handling Objects second-level categories of *Fine Hand Use (d440)*, *Hand and Arm Use (d445)* in positive and negative ways. Data mapped to F*ine Hand Use (d440)* revealed “I can grasp, I can actually grab a hold of something but trying to get my hand open and get something into my hand is difficult” (Tom), data mapped to *Hand and Arm Use (d445)* revealed “using it to open up the door” (Kit) and “I haven’t been able to use my left arm at all” (Zac).

Two participants revealed seven meaningful concepts mapped to W*alking (d450)* revealing “I’m still wobbly” (Joh), "maybe I could walk a bit better if my arm was a bit freer" (Tom) and *Moving Around in Different Locations (d460)*, "walk up and down every aisle of the supermarket" (Tom), "walk a whole city block" (Tom). A smaller number of meaningful concepts were mapped to *Driving (d475)* “I can drive” (Joh), “I couldn’t drive” (Zac).

Activity and Participation – Self-care: The core set for stroke contains six of the nine second-level categories from the ***Self-Care (d5)*** chapter. Participant data was mapped to 4 of those categories, with most meaningful concepts mapped to *Washing Oneself (d510)* and *Dressing (d540)* followed by *Eating (d550)* and C*aring for Body Parts (d520)*. Data mapped to *Washing Oneself (d510)* revealed “Now, a shower, I just enjoy it, it’s easy and not worry about it” (Bob) in contrast to “[assistance with] stuff like showering”

(Dan). Meaningful data mapped to *Dressing (d540)* revealed the importance of being able to dress oneself “my zip, it’s a big thing. Being able to just hold the jumper or a coat and pull it down and tuck them up" (Bob) and the need for assistance “[my parents] help me to get dressed every morning” (Zac).

Activity and Participation – Domestic Life: In ***Domestic Life (d6)*** there are 11 second-level categories. Meaningful concepts were linked to the 3 second-level categories included in the Comprehensive Core Set for stroke. Participant’s revealed data linked evenly across the categories of *Acquisition of Goods and Services (d620)* and its third-level categories of *Shopping (d6200)*, *Preparing Meals (d630)* “I used to cook” (Bob), “I do all of my own cooking” (Zac), and *Doing Housework (d640)* “I do all my own washing” (Tom). Participants also revealed data linked to *Acquiring a Place to Live (d610)* and *Caring for Household Objects (d650)*, which are not included in the core set.

Activity and Participation –Interpersonal Interactions and Relationships: **Interpersonal Interactions and Relationships (d7)** captured data linked to Family Relationships (d760) “But the kids took over things that I used to do for them” (Bob) and Intimate Relationships (d770) “I separated with my wife and my three kids are still in [city name]” (Zac) second-level categories, both included in the core set.

Activity and Participation –Major Life Areas: In **Major Life Areas (d8)** participant data was linked to Remunerative Employment (d850) with one participant revealing “I’m working almost fulltime” (Joh), Acquiring, Keeping and Terminating a Job (d845) “I used to work before that… when I had the stroke I had to stop that, I couldn't work" (Zac) and Economic Self-Sufficiency (d870) to link the financial cost experienced particularly to access services and support. School Education (d820), which is not included in the core set, was required to link data.

Activity and Participation – Community Social and Civic Life: ***Community Social and Civic Life (d9)*** was populated with data from *Recreation and Leisure (d920)*, a second-level category included in the Comprehensive Core Set, from 4 participants. Meaningful

concepts discussed included “draw” (Bob), “kept going playing games” (Bob), “fishing” (Kit) and “pool” (Kim).

2b.5.5. Environmental Factors

The **Environment Factors** domain captured 115 meaningful concepts identified across 4 of the 5 chapters. Data was linked to 11 of the 33 included second-level categories, with an additional second-level category required. Most meaningful concepts fell within the ***Support and Relationships (e3)***, followed by ***Services, Systems and Policies (e5), Products and Technology (e1)*** and finally ***Natural Environment and Human-Made Changes to Environment (e2)***. No data was linked to the ***Attitudes (e4)*** Chapter.

When considering ***Products and Technology (e1),*** eight of the 14 second-level categories are included in the core set. Meaningful concepts were linked to three second-level categories, the first *Products and Technology for Personal Use in Daily Living (e115)* and more specifically the third-level category *Assistive Products and Technology for Personal Use in Daily Living (e1151)* "I have certain clamps and procedures that I do things with" (Zac). Secondly, the *Products and Technology for Personal Indoor and Outdoor Mobility and Transportation (e120)* second-level category and third-level category *Assistive Products and Technology for Personal Indoor and Outdoor Mobility and Transportation (e1201)* was required, “sitting in a wheelchair” (Joh). Thirdly, *Products or Substances for Personal Consumption (e110)* was required to link data related to medication.

When considering ***Natural Environment and Human-Made Changes to Environment (e2)*** there are 13 second-level categories with only one, *Physical Geography (e210)* included in the ICF core set. Meaningful data was linked to this single category revealing “distance made it difficult” (Tom) to access neurorehabilitation services. The effect of *Climate (e225)* was required and was not available within the core set categories, “The hardest thing was the weather. It was cold, and to do that, I had to take everything off. It was cold” (Zac).

When considering ***Support and Relationships (e3)***, there are 13 second-level categories, 7 of these are in the Comprehensive Core Set for stroke. Most meaningful data was

linked to *Immediate Family (e310)* with participants discussing children “I’ve got five kids” (Bob), and support provided “my daughter, when she’s home she’ll cook for me” (Kim), siblings and parents “my parents [help me get dressed every morning]” (Zac). Meaningful concepts were also linked to *Personal Care Providers and Personal Assistants (e340)* highlighting support received “one day a week I have a support worker that comes and picks me up and takes me wherever I want to go” (Zac) and "She does pretty much the things I can’t do like change the sheets on my bed and makes it, vacuums for me and mop floors and clean the bathroom" (Tom). *Acquaintances, Peers, Colleagues, Neighbours and Community Members (e325)* and *Other Professionals (e360)* linked limited data

Whilst participants discussed support from *Health Professional (e355)* as presented above, most meaningful concepts were linked to the ***Services, Systems and Policies (e5)*** Chapter and more specifically to the *Health Services, Systems and Policies (e580)* second-level category. As per the linkage rules (Cieza et al., 2019; Cieza et al., 2005), participants identified different services and facilities proving care or intervention to them “because I’m still seeing the spasticity clinic” (Che).

2b.5.6. ICF Brief Core Set for Stroke Linking

Six second-level categories from the **Body Functions** domain are included in the brief core set. Participant data was linked to only 2 (*Mental Functions of Language b167*, *Muscle Power Functions b730*) of the 6. An additional 20 categories were required to capture experience.

Two second-level categories are included in the **Body Structures** domain in the brief core set, *Structure of Brain (s110)* and *Structure of Upper Extremity (s730)*. Participant data was linked to the latter category only. An additional 3 categories were required to capture experience.

Seven second-level categories from the **Activities and Participation** domain are included in the brief core set. Five of the 7; *Speaking (d330)*, *Walking (d450)*, *Washing Oneself*

(d510), *Dressing (d540)*, *Eating (d550)* were required for linking. An additional 28 categories were required to capture experience.

Three second-level categories from the **Environmental Factors** domain are included in the brief core set. All 3; *Immediate family (e310), Health Professionals (*e355) *and Health Services, Systems and Policies (*e580*)* were required for data linking, with an additional 9 categories required to capture experience.

2b.5.7. ICF Meaningful Concepts by Participant

Comparison of second-level categories across stroke survivors highlighted diverse, unique and similar experiences. Participants varied in the range of meaningful concepts discussed, for example Ted's experience linked to only seven second-level categories whilst Zac's experienced linked to 48 different second-level categories. Che was the only participant with no content linked to Activity and Participation, Kim, the female participant was the only person to have content linked to *Caring for Body Parts (d520)*, and Bob to *Handling Stress and other Psychological Demands (d240)*. All 10 participants shared meaningful concepts linked to *Health Services, Systems and Policies* (e580), and all participants excluding Che, had data linked to the ***Neuromusculoskeletal and Movement-Related Functions*** chapter within the **Body Function** domain. Che and Dan were the only participants who did not identify content linked to the third-level category *Motivation (b1301)*. Participant responses linked to ICF domain are presented in Table 2b.

2b.5.8. Concepts Unable to be Linked

Participants discussed experience related to personal factors. According to the ICF definition of personal factors and linking rules (Cieza et al., 2005), concepts were assigned to personal factors and not linked to second-level categories. Such factors identified included the living situations of participants, relocating to live with family post-stroke, deaths of partners and family members, and relationship status.

A small number of meaningful concepts were unable to be linked to a second-level category. Concepts related to time intervals post-stroke and setting aside time "…setting aside the time to practice" (Lee); change or lack of change in function "[Facilitator] Do you feel like it's got tighter over the two years? [Interviewee] No. [Facilitator] It's just basically been the same…[Interviewee] Yeah, it's plateaued out, yes" (Zac); changes to life roles "What about roles and responsibility within your family? Do you have any particular responsibilities or roles? [Interviewee] Nope" (Dan); and contributing to the stroke community "It was good. It was good to - I feel like I was doing something to help. Like, you know, for future people that had a stroke, and if it does help them, that's fantastic" (Zac).

2b.6. Discussion

This study aimed to understand the impact of upper limb spasticity on stroke survivors by linking their experience revealed through interviews to the ICF Comprehensive Core Set for stroke. The linking of experience to the ICF is an approach used in previous qualitative research with post-stroke survivors (Algurén et al., 2010; Glässel et al., 2012; Glässel et al., 2014; Glässel et al., 2011; Glässel et al., 2010; Paanalahti et al., 2014; Paanalahti et al., 2013; Riberto et al., 2013; Snögren & Sunnerhagen, 2009; Tavener et al., 2015; Wang et al., 2014), but this is the first study to explore the experience of stroke survivors who have upper limb spasticity. Study findings sought to understand what ICF categories were important or noteworthy from the perspective of participants themselves.

Key findings revealed that stroke survivors with upper limb spasticity use words and share topics that are concentrated around global and specific mental functions, functions of the joints and bones, muscles and movements, carrying, moving and handling objects, support and relationships, primarily of immediate family and health professionals, products and technology for personal use in daily living and health services.

The impact of known, prevalent post-stroke cognitive impairments, such as memory, attention and fatigue on survivor experience has been reported (Algurén et al., 2010;

Paanalahti et al., 2013). This perspective was also shared by participants with upper limb spasticity within this study, with most meaningful concepts linked to the ***Mental Functions*** chapter. In contrast to Swedish stroke survivors who identified common meaningful concepts linked to ***Mental Functions***, specifically the third-level category of *Fatigue* (energy) within *Energy and Drive Functions (b130)* (Algurén et al., 2010), participants in our study also focussed on *Energy and Drive Functions (b130),* but within the third-level category *Motivation* (drive). Participants in our study had consented to participate in a protocol based upper limb intervention trial, thus this experience and willingness to participate may have influenced findings. Exploring the perspectives of stroke survivors who have not participated in a trial or recent neurorehabilitation is an area recommended for future research.

The importance of impairments in ***Neuromusculoskeletal and Movement-Related Functions*** and the ability to use the hand and arm in activities highlighted in previous studies (Algurén et al., 2010; Paanalahti et al., 2014; Paanalahti et al., 2013; Riberto et al., 2013; Wang et al., 2014), was echoed by our participants. Participants discussed a lack of strength, the inability to move the upper limb or control voluntary movements and the need to straighten out and stretch the arm that felt 'tight'. Indeed, most meaningful concepts in this Chapter (b7) were linked to *Control of Voluntary Movements (b760)*. This suggests the category b760 could be routinely included when understanding post-stroke experience for survivors with upper limb spasticity.

Despite participants being chronic stroke survivors with very little ability to control and use their affected upper limb, meaningful concepts linked to the **Activity and Participation** domain fell primarily within fine hand use and hand and arm use. Participants discussed upper limb activity along a continuum of nil ability, to noting small changes in ability, to completing basic activity such as opening a door. Maintaining focus on possible upper limb use and the mental drive to improve their ability to use their arm despite time post-stroke and impairment level was highlighted by some of our participants. That said it is clear that activity limitations and participation restrictions could arise from categories across the full range of the ICF including **Environment** not just aspects specific to spasticity.

Participants in this study also conveyed the importance of external support as identified in previous research (Paanalahti et al., 2013). Participants frequently discussed the support from immediate family, assistive devices, services and health professionals as facilitators to enable them to not only live in their own home, but to support engagement in their neurorehabilitation.

This study did not aim to validate the content of the Comprehensive or Brief Core Sets, however considering experience through the lens of the ICF Core Sets, provides beginning insights into the applicability of the frameworks. Findings, similar to previous post-stroke studies, revealed that the Comprehensive Core Set for stroke captured meaningful concepts whilst the Brief Core Set omitted the linking of experience to represent the breadth of impact. Because the study design permitted participants to raise any aspects of their experience and as questions and analysis were not restricted to categories in the Core Set, meaningful concepts were raised that linked to an additional nine second-level categories outside the Comprehensive Core Set.

In **Body Functions**, the need for additional ICF categories to capture participant data is not unique to this study. Of importance to our stroke survivors with upper limb spasticity, and a key finding identified by previous validation studies (Glässel et al., 2012; Glässel et al., 2014; Glässel et al., 2011; Glässel et al., 2010; Lemberg et al., 2010), is the requirement of the second-level category *Involuntary Movement Functions (b765)* to link clonus and tremor associated with post-stroke spasticity.

In **Activity and Participation**, *Caring for Household Objects (d650),* was required as an additional category to the Core Set to link the experience of being unable to continue maintaining their home. Survivors in earlier phases of their post-stroke journey revealed similar experience linking to this additional category to capture the loss of autonomy and becoming more dependent on others to maintain and manage their household (Glässel et al., 2012). Occupational therapists participating in a validation study (Glässel et al., 2010) also proposed inclusion of this category given the frequent impact experienced post-stroke on domestic life. *Basic Interpersonal Relationships (d710)* is included in the Comprehensive Core Set, yet our participants, like other previous studies

(Glässel et al., 2012; Glässel et al., 2010; Paanalahti et al., 2013), revealed experience beyond this category requiring the addition of *Complex Interpersonal Relationships (d720)* to capture the frustrations and interpersonal negotiations undertaken with others who they felt did not understand the impact of spasticity. The second-level category *Thinking (d163),* was required for our participants and for stroke survivors within the first three months post-stroke (Glässel et al., 2012). Three additional categories required were unique to this study. The first, *Acquiring Information (d138)*, was required to link the experience of learning about stroke and neurorehabilitation throughout their participation in the trial. The second, *School Education (d820)* was required and is a category that should be considered for possible inclusion to encompass experience of younger stroke survivors who may still be attending school. And thirdly, *Acquiring a Place to Live (d610)* was required to link experience of participants who discussed needing to find a more suitable home due to challenges maintaining their current residence, a concept closely linked to *Caring for Household Objects (d650).*

In **Environment**, the category *Climate (e225)*, was a required addition, to link our participant experience when they revealed the impact of colder climates in dressing requirements and thus their ability to complete this alone – the cold was an environmental barrier impacting on dressing. Paanalahti et al. (2014) also identified climate was an additional required category, because their stroke survivors reported weather as a barrier that impacted walking.

In light of these findings, consideration was given to the way in which the extended version of the Comprehensive Core Set for stroke captured participant experience. The extended version includes categories from the Core Set for patients with neurological conditions in the acute hospital and in early post-acute rehabilitation facilities to increase applicability beyond chronic post-stroke phases to earlier post-stroke phases (Ewert et al., 2005; Stier-Jarmer et al., 2005). The extended version, when applied to our participants, did not capture the remaining additional meaningful concepts within included categories. The failure to further capture additional meaningful concepts is likely due to the chronicity of our participants, who similar to participants in the study by Paanahlati et al. (2014), were not in the acute phase post-stroke.

Findings from this exploratory study suggests that the Comprehensive Core Set, rather than the Brief or extended versions, should be used as a basis for ICF linkage for people with upper limb spasticity post-stroke but it should not be used in isolation. As a starting point, the core set could be augmented with the **Body Function** third-level category of *Motivation (b1301)* to capture drive independent of fatigue, a potential current limitation. Furthermore, addition of the second-level category *Involuntary Movement Functions (b765)*, the **Activity and Participation** second-level categories *Thinking (d163)*, *Caring for Household Objects (d650)*, *Complex Interpersonal Relationships (d720)* and *School Education (d820)*, and ***Environment*** Chapter category of *Climate (e225)* could enhance representation. The opportunity to link across other ICF categories should be considered if the aim is to capture unique experience of individual participants.

Findings of this study provide insights to neurorehabilitation professionals aiming to implement collaborative and goal directed practice. The breadth of experience, when framed by the ICF, identifies the variety and the scope of health and wellbeing impacts and thus treatment outcomes that may be important or of interest to stroke survivors with upper limb spasticity. The findings may also illuminate the type of concerns patients have, helping to frame the scope of person-centred enquiry in measurement and evaluation phases of neurorehabilitation. The findings may also help suggest the type of information survivors with upper limb spasticity may need to understand their experience in relation to other survivors and evidence-based interventions available, to help formulate treatment expectations or interpret and manage their reactions about what has and is happening to them.

Further research into stroke survivor experience is recommended. Consideration of study design including surveys to capture experience from a broader number of participants in addition to further enquiry with participants who have upper limb spasticity but have not recently participated in neurorehabilitation is recommended to gain representativeness and prevalence of different impacts.

There were limitations in this study. First, the study context was a possible limitation. It was conducted after the completion of a neurorehabilitation program where having

upper limb spasticity was an inclusion criterion and an explicit focus of the intervention. Participants had interventions that enhanced their understanding of their spasticity and how to manage it. In one sense this could be construed as a limitation because they are likely to have had superior exposure to spasticity terminology, neurorehabilitation strategies directed towards spasticity management, and interactions with professionals about upper limb spasticity when compared with most post-stroke survivors. At the same time this 'limitation' could be a strength of the study because the enquiry focus was a topic area they were familiar and comfortable with as participant 'experts'. Secondly, sampling was a limitation. It was conducted prospectively using diversity of participant demographic attributes as criteria. These attributes were selected on face value rather than informed by a theoretical framework. Data sampling within the participant group was not possible because interviews were held on only one occasion. Findings should therefore be seen as indicative.

2b.7. Conclusion

This study presents new information that illuminates the experience of stroke survivors with upper limb spasticity after recent completion of a targeted neurorehabilitation program. Whilst this study did not aim to validate the Core Sets for stroke, findings suggest the Brief Core Set is too limited to link the totality of the experience of stroke survivors with upper limb spasticity. While the Comprehensive Core Set more meaningfully captures experience, it misses important aspects in **Body Functions** *(Involuntary Movement Functions (b765))*, **Activity and Participation** (*Acquiring Information (d138), Thinking (d163), Acquiring a Place to Live (d610), Caring for Household Objects (d650), Complex Interpersonal Relationships (d720), School Education (d820))* and **Environmental Factors** (*Climate (e225))*. These may be useful additional categories to consider if further research demonstrates they are common concerns. Individual experience links to a wide variety of categories outside the Comprehensive Core Set – all participants had one or more categories outside the Core Set. This finding suggests that living with upper limb spasticity involves unique as well as common impacts and experiences. Since the extended Core Set for stroke does not capture the range of categories we found, it may indicate either that the impacts of survivors with upper limb spasticity on functioning disability and health are yet to be considered or that

the circumstances and attributes of our participants are particular and unique. Future research using a variety of approaches, could explore whether additional ICF categories revealed in this study and patterns emerging across qualitative studies have potential for interrogation as propositions that may be generalizable to other samples and settings.

2b.8. Chapter Synopsis

The results of this study identified unique categories needed so as to apply the ICF Comprehensive and Brief Core Sets for stroke to capture and describe the experience of living with upper limb spasticity post-stroke. This Chapter 2b and Chapter 2a have thus explored how stroke survivors experience upper limb spasticity providing unique and important patient-centred insights. The subsequent Chapters will now consider the measurement of the reported areas of impact within the performance of upper limb activity.

Chapter 3:

A Systematic Review Protocol

This study has been published as:

Pike, S., Lannin, N. A., Cusick, A., Wales, K., Turner-Stokes, L., & Ashford, S. (2015). A systematic review protocol to evaluate the psychometric properties of measures of function within adult neuro-rehabilitation. *Systematic reviews*, 4(1), 86. doi:10.1186/s13643-015-0076-5

Trial Registration: Prospective Register of Systematic Reviews (PROSPERO) registration number CRD42014013190.

See Appendix A for publication permission, Appendix C for trial registration, Appendix F for supplementary material, and Appendix I for published manuscript.

3.0. Chapter Overview

Contemporary neurorehabilitation uses the ICF as an organising framework. Chapter 2b reported categories within the Activity and Participation domain of the ICF linked to the experience of living with upper limb spasticity after stroke. As demonstrated in the Introduction (Chapter 1), clinicians require psychometrically validated outcome measurement tools to evaluate this survivor impact in relation to use of the upper limb in everyday activity. This chapter presents the protocol for a systematic review study which will locate, appraise and synthesise available evidence for the psychometric properties of tools used to evaluate upper limb activity and participation outcomes.

3.1. Background to the Study

In recent years there has been a recognition within neurorehabilitation that spasticity management programs must go well beyond the treatment of impairments, in line with contemporary understandings of health emerging from the World Health Organisation's ICF (World Health Organisation, 2001). This framework starts with the assumption that health is not a state independent of individuals in the context of everyday life; thus spasticity, a neuro-muscular condition, cannot be considered independent of the person who has it and their daily life. This makes understanding, measuring and monitoring the impact of neurorehabilitation programs on function in everyday life as important as measuring and monitoring spasticity.

In upper limb neurorehabilitation, function is an important and debated term because impairments can affect function and it is through function that activity and participation goals can be achieved. The term function is used variably within the literature; it alludes to impairments, activity performance and/or participation in life situations in addition to associations with active task performance. Whilst concepts associated with function can vary, operationally, functional use of the spasticity-affected upper limb has been defined by Ashford and Turner- Stokes (2013). That is: active task performance; the affected limb actively completes the task or passive task performance; the task is completed by the affected limb with assistance from the unaffected limb or the task is assisted with or completed by a carer; a key area for spasticity interventions (Ashford & Turner-Stokes,

2013). This three-part operational definition of upper limb function is used in the present study.

Multi-disciplinary person-centred approaches are needed to address neurorehabilitation needs at impairment, activity and participation levels (Demetrios, Khan, Turner-Stokes, Brand, & McSweeney, 2013; Royal College of Physicians, British Society of Rehabilitation Medicine, Chartered Society of Physiotherapy, & Association of Chartered Physiotherapists Interested in Neurology, 2009; Sheean et al., 2010). Neurorehabilitation clinical practice guidelines recommend collaborative goal setting (Playford, Siegert, Levack, & Freeman, 2009; Rosewilliam et al., 2011; Royal College of Physicians et al., 2009; Turner-Stokes, 2009), so that patient preferences and priorities can inform programs. Practice guidelines also recommend the use of standardised outcome measurement tools to measure impairment, activity and participation dimensions of performance relevant to everyday real life (Ashford & Turner-Stokes, 2013; Rosewilliam et al., 2011; Sheean et al., 2010; Turner-Stokes, Fheodoroff, Jacinto, & Maisonobe, 2013). Although most neurorehabilitation clinicians measure treatment outcomes (Ward et al., 2004), evidence suggests that many have limited awareness of the range of outcome measurement tools available (Williams, Olver, Graaff, & Singer, 2012). Those who do use assessment use predominantly impairment-based measures—few use measures that capture activity or participation performance (Bakheit et al., 2010; Sheean et al., 2010). There are measures available. Although an earlier systematic review of functional outcome measures in the hemiparetic upper limb was conducted (Ashford et al., 2008), this study was unable to identify a single valid and reliable outcome measure that captured “real life” function. But a more recent review (Ashford & Turner-Stokes, 2013) identified n = 27 functional assessments used in upper limb neurorehabilitation for people with spasticity (with and without botulinum toxin-A injection). Their inclusion criteria required the outcome measurement tools to explore function in the context of everyday real life. As yet, these tools have not been appraised in relation to the psychometric rigour or clinical utility (Ashford & Turner-Stokes, 2013; Intiso et al., 2013). This study aims to fill that gap.

3.2. Study Aim

This study aims to guide clinicians in their tool selection when evaluating important, individualised outcomes and the efficacy of neurorehabilitation interventions. This study will use the same n = 27 outcome measurement tools from the Ashford and Turner-Stokes (2013) study to investigate the psychometric properties of each and draw conclusions regarding their relative rigour and relevance. A key focus will be the validity of these outcome measurement tools in their ability to capture change in activity performance and life participation. The ICF will be used as the framework to appraise tool content to determine the extent to which items address activity and participation domains in addition to impairment (body structures and function). Determining the content validity of items in relation to these domains is important not only to see how valid the tool is in measuring "health" as it is defined by the ICF but also because these domains reflect common patient goals.

Common neurorehabilitation goals for people with upper limb spasticity include reducing pain, increasing the range of movement, preventing contractures and reducing spasticity to enable movement training, splinting or casting (Sheean et al., 2010; Sheean, 2001; Ward et al., 2004). Other goals relate to increasing a person's ability to perform activities and participate in their life situation (Cusick et al., 2015; Elia, Filippini, Calandrella, & Albanese, 2009; Intiso et al., 2013; Sheean et al., 2010; Turner-Stokes et al., 2010; Turner-Stokes, Fheodoroff, Jacinto, & Maisonobe, 2013). To date, no systematic review has done this.

Outcomes of this review will help clinicians and researchers alike working with people who have upper limb spasticity. Attention has been given to neurorehabilitation for people with upper limb spasticity on function in everyday real life (Ashford & Turner-Stokes, 2013; Intiso et al., 2013; Rosewilliam et al., 2011; Turner-Stokes, Fheodoroff, Jacinto, & Maisonobe, 2013; Turner-Stokes, Fheodoroff, Jacinto, Maisonobe, & Zakine, 2013), but it is a relatively new focus of program evaluation (Ashford & Turner-Stokes, 2013; Rosewilliam et al., 2011; Turner-Stokes, Fheodoroff, Jacinto, & Maisonobe, 2013) and a challenging one (Hinderer & Gupta, 1996; Sheean, 2001; Turner-Stokes et al., 2010). Determining whether or not interventions impact functional outcomes in

everyday life for people with upper limb spasticity has, to date, been complicated by methodological problems, not just in relation to function but also spasticity measurement (Ashford et al., 2008; Platz, Eickhof, Nuyens, & Vuadens, 2005) and the use of weak study designs (Hinderer & Gupta, 1996; Sheean, 2001) as discussed in Chapter 1. There is a need for more research to show that multidisciplinary upper limb spasticity management neurorehabilitation programs impact people's ability to perform activities in everyday real life (Shaw et al., 2011; Simpson et al., 2008; Turner-Stokes, Fheodoroff, Jacinto, & Maisonobe, 2013).

The aims of this systematic review will be:

1. To classify the functional outcome measures reported by Ashford and Turner-Stokes (2013) according to whether activity and/or participation outcomes following upper limb spasticity neurorehabilitation are being assessed; activity performance and participation will be defined according to the ICF model (World Health Organisation, 2001); and

2. To locate all of the existing evidence of the properties of the outcome measures, to evaluate the strength of this evidence and come to a conclusion about the best measure available for the particular purpose of measuring activity and/or participation outcomes following upper limb spasticity neurorehabilitation.

3.3. Method

This systematic review builds on the systematic search conducted by Ashford and Turner-Stokes (2013) by synthesising and appraising the research of the psychometric (measurement) properties of outcome measures reported within the published paper. Their review thoroughly identified outcome measurement tools used to assess activity thus did not require duplication. From the 22 studies located in the published search (Ashford & Turner-Stokes, 2013), n = 33 measurement approaches were identified. On review of those measurement approaches, some were in fact developed for that particular study, for example, three functional tasks (palm hygiene, cutting the

fingernails, placing an arm through the sleeve), and consequently do not have published psychometric properties and were excluded from the current study. The remaining n=27 outcome measures had published research investigating their psychometric (measurement) properties and will therefore form the sample for the present study. The authors acknowledge the creation of a degree of outcome measurement tool selection bias due to this method.

3.4. Data collection

3.4.1. Publication/Study Inclusion Criteria

1. The aim of the study should be to develop or evaluate the measurement properties of a measurement tool identified in the review published by Ashford and Turner-Stokes (Ashford & Turner-Stokes, 2013);

2. The tool should aim to measure activity performance or participation, as defined by the ICF (World Health Organisation, 2001). Activity performance is defined as “the execution of a task or action by an individual” or requires assistance from or be completed by a carer for the individual. Participation is defined as “involvement in a life situation.”

3. The tool is evaluated in adult patients, over 18 years of age, with upper limb spasticity (as defined by the authors of the included studies) or patients before or after botulinum toxin injection engaging in upper limb neurorehabilitation programs (with or without the inclusion of botulinum toxin therapy). A neurorehabilitation program is one that is devised and implemented by a clinician to work towards achievement of identified goals. Participants can be engaging in the neurorehabilitation program whilst a hospital inpatient, transitioning to home or be community-dwelling.

4. All research studies must be original research, and both conducted and published studies in English within peer-reviewed literature will be considered for this review.

3.4.2. Publication/Study Exclusion Criteria

This review is concerned with outcomes of upper limb spasticity neurorehabilitation that identify changes in the performance of an activity or participation as defined by the ICF (World Health Organisation, 2001). Studies that measure activity performance and participation will be included. Studies that measure upper limb spasticity neurorehabilitation outcomes through assessment of upper limb impairments only, including pain, range of movement, contracture and changes in tone, will be excluded. Outcomes that have been modified in any manner or implemented in a language other than English will be excluded.

3.4.3. Search Methods for the Identification of Studies

A search will be conducted in Medical Literature Analysis and Retrieval System Online (MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL) and Excerpta Medica dataBASE (EMBASE). The databases were selected due to indexing a high volume of applicable literature within the areas of life sciences, nursing and allied health and biomedicine. Multiple databases were selected to increase search results due to differing indexing of publications across databases, potentially limiting search results. In MEDLINE, a validated search filter for finding studies on measurement properties will be used (Terwee, Jansma, Riphagen, & de Vet, 2009) (see Appendix F). The translated versions for the Cumulative Index to Nursing and Allied Health Literature (CINAHL) and Excerpta Medica database (EMBASE) were also used. Searches with the names of each included instruments (in the title) in combination with the terms for the study population as described in the search strategy (Appendix F), will be conducted until each instrument has been searched.

3.4.4. Screening

Once all searches have been exhausted, the abstracts will be downloaded into the reference management system EndNote and duplicates deleted. A study deemed as a duplicate will have authors, setting and location, outcome measures implemented, date and duration of study in common. The eligibility criteria will first be applied to the title and abstract, and if deemed relevant, the full manuscript will be retrieved to determine

eligibility of potential studies. The initial screen and selection will be completed independently by the thesis author with the thesis author's supervisor blindly screening a 10 % selection of articles for eligibility. Debate on the inclusion or exclusion of studies will be resolved by an independent third reviewer (thesis author supervisor) and discussion between all three reviewers to reach consensus.

3.4.5. Data Management

Details on studies that were initially selected based on title and abstract, full-text articles that were retrieved and articles included in the review will be documented. Reasons for the exclusion of retrieved full-text articles, particularly in the case of doubtful articles, will also be recorded.

3.4.6. Data Extraction

Data will be extracted from selected studies by the thesis author utilising a standardised data extraction form. This form will record information related to participants, study design, description of botulinum toxin therapy and neurorehabilitation program(s), outcome measures administered and their classification according to the ICF (activity performance and or participation focus), psychometric properties, study inclusion/exclusion criteria if available and a brief summary of the findings. The thesis author's supervisor will crosscheck all COSMIN ratings.

3.4.7. Risk of Bias Assessment

Studies evaluating the measurement properties of an outcome measurement tool require high methodological quality with a low risk bias to guarantee that appropriate conclusions are drawn about the properties of the measure (Terwee et al., 2009). Thus, it is important to evaluate those methodological qualities (De Vet, Terwee, Mokkink, & Knol, 2011). This review will apply the COnsensus-Based Standards for the Selection of Health Status Measurement INstruments (COSMIN) checklist with 4-point scale version (Terwee et al., 2012). This version is recommended by the COSMIN developers for use in systematic reviews of measurement properties. The checklist will be applied to assess the quality of the papers reporting on the psychometric properties of the 27 outcome

measures, evaluating whether each study meets the standards for methodological quality with regard to internal consistency, reliability (test-retest, inter-rater and intra-rater reliability), measurement error, content validity (including face validity), structural validity, hypothesis testing, cross-cultural validity, criterion validity, responsiveness, interpretability and generalisability (Terwee et al., 2012). The 4-point scale will allow a methodological quality rating of either "excellent", "good", "fair" or "poor" to be assigned to the study (Terwee et al., 2012). The COSMIN checklist was developed in an international Delphi study with the focus of evaluating the methodological quality of studies on measurement properties (Terwee et al., 2012). The COSMIN checklist is a modular tool, and the measurement properties evaluated in the study will determine which components or "boxes" need to be completed (Terwee et al., 2012).

3.4.8. Data Analysis

Individual assessment items within the outcome measures will be examined to extract meaningful concepts. Those concepts will then be linked to the ICF framework categories of activity performance and or participation following the linking rules suggested by Cieza et al. (2005), see Appendix F. This linking process will enable the extent to which outcomes are valid measures of activity performance and life participation to be determined. The COSMIN checklist with 4-point scale version (Terwee et al., 2012) as described above will be applied to the selected studies, as per COSMIN guidelines, to appraise the overall methodological quality of studies. From here, Terwee's quality criteria for measurement properties (Terwee et al., 2007) will be applied. Quality criteria for the following nine measurement properties are defined: content validity, internal consistency, criterion validity, construct validity, reproducibility, reliability, responsiveness, floor and ceiling effects and interpretability. This data analysis process will enable conclusions to be drawn regarding the strongest psychometric measure available for the particular purpose of evaluating activity and/or participation outcomes following upper limb neurorehabilitation. Differences in the psychometric properties of outcome measures for patients with and without upper limb spasticity will be discussed.

3.5. Discussion

The systematic review described in this protocol chapter will provide a comprehensive evaluation of the measurement properties of outcome measures assessing activity performance and participation goals for adults with upper limb spasticity undergoing neurorehabilitation. The results of this review will provide health professionals with detailed information to guide clinical decision-making when choosing the most appropriate outcome measurement tool for purpose. Neurorehabilitation clinicians and managers will also be provided with information to permit accurate measurement and monitoring of the relationship between neurorehabilitation and health outcomes in these patients.

3.6. Chapter Synopsis

This chapter presented the protocol for the systematic review study to identify, appraise and synthesise evidence on the psychometric properties of tools used to evaluate activity and participation outcomes. The search scope was restricted to those articles in the Ashford and Turner-Stokes (2013) review. At the time of protocol publication, the search thus excluded tools outside that review. The proposed clinical implications for this study are discussed. The following two Chapters (4, 5) present the results of this study.

Chapter 4:

A Systematic Review

The work covered in this chapter has been published as:

Pike, S., Cusick, A., Wales, K., Cameron, L., Turner-Stokes, L., Ashford, S., & Lannin, N. A. (2021) Psychometric properties of measures of upper limb activity performance in adults with and without spasticity undergoing neurorehabilitation – A systematic review. PLoS ONE 16(2): e0246288. doi.org/10.1371/journal.pone.0246288

See Appendix C for trial registration, Appendix F for supplementary material and Appendix I for published manuscript.

4.0. Chapter Overview

This chapter presents the results of the systematic review, implemented according to the protocol presented in Chapter 3 with methodological amendments to align with the revised COSMIN methodology (Mokkink et al., 2018; Terwee et al., 2018) released after the Chapter 3 protocol was published. In addition a methodological revision was made to include outcome measurement tools identified in the most recent clinical guidelines for spasticity management, which resulted in one additional tool being included. Revisions are captured within the PROSPERO registration information presented in Appendix C. Outcome measurement tools in the main text of the Chapter are named without citations because the citations are part of the results presented.

4.1. Background

The personal experience of an ABI can be profound, impacting on all areas of a person's health and wellbeing. The ICF (World Health Organisation, 2001) provides a framework to consider the impact of an ABI on a person, highlighting both the breadth and complexity of potential issues. While the ICF can classify areas that may be impacted by an ABI, and some rating of impairment and limitation is possible using the ICF core sets (Geyh et al., 2004; Zhang et al., 2018) (presented in Chapter 2b), precise measurement of factors known to be related to activity is essential.

Accurate measurement is key to determining the effect of neurorehabilitation interventions, and therefore measurement tools used in neurorehabilitation should target all levels of functioning, disability and health – this includes activity and participation as much as impairments in body structure and function (Lohmann, Decker, Müller, Strobl, & Grill, 2011). In addition to targeting all levels, measurement should also capture and reflect actual performance of everyday 'real-life' activities outside of the clinical setting (Ashford et al., 2008). Measurement of activity and participation in 'real-life' activities presents many challenges, not least of which is consistency, validity and sensitivity of 'real life' functions.

Several reviews have sought to identify and determine the most suitable outcome measurement tools to measure upper limb impairment and activity for adults with a neurological condition (Alt Murphy et al., 2015; Ashford et al., 2008; Lamers et al., 2014). Scant evidence has been located, and clear gaps have been identified in the presentation of the psychometric quality of the tools in a neurorehabilitation context (Alt Murphy et al., 2015). Furthermore, Alt Murphy et al. (2015), identified that many of the reviews to date failed to critically appraise the methodological quality of the individual studies evaluating the psychometric properties of the tools. Whilst recommendations regarding upper limb evaluation have been made, the tools identified and the evidence regarding the psychometric properties of the tools were not specifically targeted nor extracted from a sample of adults with upper limb spasticity as a result of their neurological condition. Spasticity severity may impact psychometric properties of outcome measurement tools such as responsiveness and sensitivity. Thus influencing which tool should be selected for use based on psychometric evidence. Currently little is known of the psychometric properties of many outcome measurement tools for use with people with upper limb spasticity. Essentially it is unknown if the tools are still valid, reliable and responsive when used with people with upper limb spasticity.

Review work by members of this study's authorship team, Ashford and Turner-Stokes, did identify outcome measurement tools both applicable to the upper limb that assess function in the context of everyday life, and from studies including adults with upper limb spasticity (Ashford & Turner-Stokes, 2013). They demonstrated newer upper limb measurement tools used in neurorehabilitation research which examine activity and participation in the context of everyday real-life activities show promise (Ashford & Turner-Stokes, 2013). There is thus a need for a comprehensive appraisal and synthesis of the psychometric properties of all these tools, to potentially recommend a tool/s for clinical and research use.

4.2. Study Aim

The two aims of this study, therefore, were to firstly critically appraise and summarize the quality of the psychometric properties of previously identified upper limb activity performance measurement tools (Ashford & Turner-Stokes, 2013) when used with

adults with upper limb spasticity using a level of evidence approach and the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines (Prinsen et al., 2018; Terwee et al., 2012; Terwee et al., 2018). Secondly, to determine if the presence of upper limb spasticity impacts which measure should be selected based on psychometric evidence; differences in psychometric properties for the identified measurement tools for adults with a neurological impairment but without upper limb spasticity will be defined.

4.3. Method

A systematic review with COSMIN appraisal was undertaken, with PRISMA guidelines informing reporting.

4.3.1. Identification and Selection of Measurement Tools

The published list of measurement tools by Ashford and Turner-Stokes (Ashford & Turner-Stokes, 2013) was used to identify and select measurement tools for appraisal. As this source systematic review was published in 2013, the most recent clinical guidelines management of spasticity in the upper limb (Royal College of Physicians, British Society of Rehabilitation Medicine, The Chartered Society of Physiotherapy, Association of Chartered Physiotherapists in Neurology, & Royal College of Occupational Therapists, 2018) was also searched so as to identify any potential tools which may have been developed since 2013. One further tool, the Arm Activity Measure (ArmA), was located and subsequently included in the review.

4.3.2. Measurement Tool Inclusion Criteria

To be included, measurement tools had to assess activity or performance as defined by the ICF (World Health Organisation, 2001), and each needed to focus on the upper limb. Activity is defined within the ICF as “the execution of a task or action by an individual” (World Health Organisation, 2001 p10) while participation is defined as “involvement in a life situation” (World Health Organisation, 2001, p.10). In the present study, the official World Health Organisation (WHO) coding of activity and participation was used, that of a single overlapping list of categories (World Health Organisation, 2013); tools that only

evaluate impairment/s (e.g. pain, range of movement, contracture, spasticity) were excluded.

4.3.3. Study Search Strategy

Searches were completed according to the published protocol (presented in Chapter 3), with methodological amendments reported in section 4.3.1. Medical Literature Analysis and Retrieval System Online (MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL) and Excerpta Medical database (EMBASE) were searched from inception to December 2016. Where able, the validated search filter for finding studies on measurement properties was used (Terwee et al., 2009); search terms are presented in Appendix F. COSMIN requires information regarding the development/content validity of the measurement tools to be sought, therefore tool references were identified and obtained when not identified within the search results.

4.3.4. Study Screening

Title and abstracts were downloaded into the reference management system EndNote™. Duplicates were removed and screened for inclusion by one reviewer. To minimise the risk of incorrect inclusion and exclusion of studies; a second reviewer screened a random 25% sample of included studies against inclusion criteria and all excluded papers were reviewed by the senior author. Disagreements were settled through independent review, followed by discussion until a consensus decision was reached. Full text papers were obtained for all included studies and checked to confirm the final inclusion/exclusion decision (Pike et al., 2015).

4.3.5. Study Inclusion and Exclusion Criteria

Studies which included participants both with and without spasticity were included; to be included in the spasticity analysis, evidence of the presence of participant upper limb spasticity was required - not just the mention of 'spasticity' in text. For example, the study by Page, Levine and Hade (2012) reported a Modified Ashworth Scale score of ≥3 as an exclusion criterion; but within the study sample there was no evidence of participants with spasticity ≤3. Thus, this article was deemed to be a study without

upper limb spasticity. In addition, only studies which tested the measurement tool in its *original and complete* form were included. This conservative approach to study selection was taken to ensure maximum possible homogeneity in the evidence base which would be used to underpin tool recommendations for practice use. If a tool was used as a comparator to validate another tool, the study was excluded in accordance with COSMIN methodology. The full protocol is presented in Chapter 3 with methodological amendments reported in section 4.3.1. Inclusion criteria are detailed in Table 4.1.

Table 4.1. *Inclusion criteria*

Design
• Psychometric properties of the identified measurement tools were evaluated
• Original research
• Conducted and published in English within peer reviewed literature
Participants
• Adults (>18 years old)
• ≥ 90% diagnosis of a following neurological condition; Stroke, Multiple Sclerosis, Cerebral Palsy, Traumatic Brain Injury, Anoxia
• With or without upper limb spasticity
• Undergoing neurorehabilitation
Measurement tool
• Measured activity and/or participation
• Nil modifications
• Complete measure administered

4.4. Data analysis

4.4.1. Methodological Quality of Studies

The quality of the included studies was appraised using the COSMIN taxonomy of measurement properties and definitions for health-related patient reported outcomes (Prinsen et al., 2018; Terwee et al., 2012; Terwee et al., 2018) and the COSMIN Risk of Bias checklist (Mokkink et al., 2018) for systematic reviews of patient-reported outcome measures. The methodological quality of each study was individually assessed to evaluate whether it met the standards for measurement tool development, content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for

construct validity and responsiveness. The Risk of Bias checklist rated each measurement property as either “very good”, “adequate”, “doubtful” or “inadequate”. As there is no accepted “gold standard” measure of upper limb activity, criterion validity was not evaluated, and construct validity and responsiveness properties were appraised within the hypothesis testing criteria of COSMIN. Where *a*priori hypotheses were not stated, studies were assigned an appropriate generic hypothesis from the list developed by the COSMIN group (Mokkink et al., 2018). Information regarding the interpretability and generalizability were collected.

4.4.2. Quality of Measurement Properties

The results of individual studies reporting on the psychometric properties were then evaluated using Terwee’s quality criteria for measurement properties (Prinsen et al., 2018), presented in Appendix F. Results were rated as sufficient ‘+’, indeterminant ‘?’or insufficient ‘-’.

4.4.3. Sample Size of Studies

Sample size was only assessed within individual studies evaluating the measurement properties of content validity, structural validity and cross-cultural validity as per COSMIN guidelines. Sample sizes of individual studies evaluating the remaining measurement properties were not assessed via the Risk of Bias Checklist, and sample sizes per those measurement properties were instead pooled at the synthesis stage (Prinsen et al., 2018).

4.4.4. Synthesis of Best Evidence

All identified evidence and results were then pooled and the modified COSMIN GRADE approach used to determine the overall quality of the evidence (Prinsen et al., 2018). The modified COSMIN GRADE approach considers and downgrades the level of evidence and consequently trustworthiness of results depending on the risk of bias (methodological quality), inconsistency of results, imprecision (based on total sample size) and indirectness (evidence from different populations than the population of interest) (Prinsen et al., 2018, p.1151); indirectness was not applicable in this review as

studies conducted in samples other than those specified in the inclusion and exclusion criteria were excluded. The synthesis determines either “high”, “moderate” “low” or “very low” quality levels of ‘sufficient’, ‘insufficient’, ‘inconsistent’ or ‘indeterminant’.

4.5. Results

Of the 33 measurement tools identified in the Ashford and Turner-Stokes review (Ashford & Turner-Stokes, 2013), 29 measurement tools were published tools. We completed searches for these 29 measurement tools, plus the ArmA (i.e. 30 tools in total).

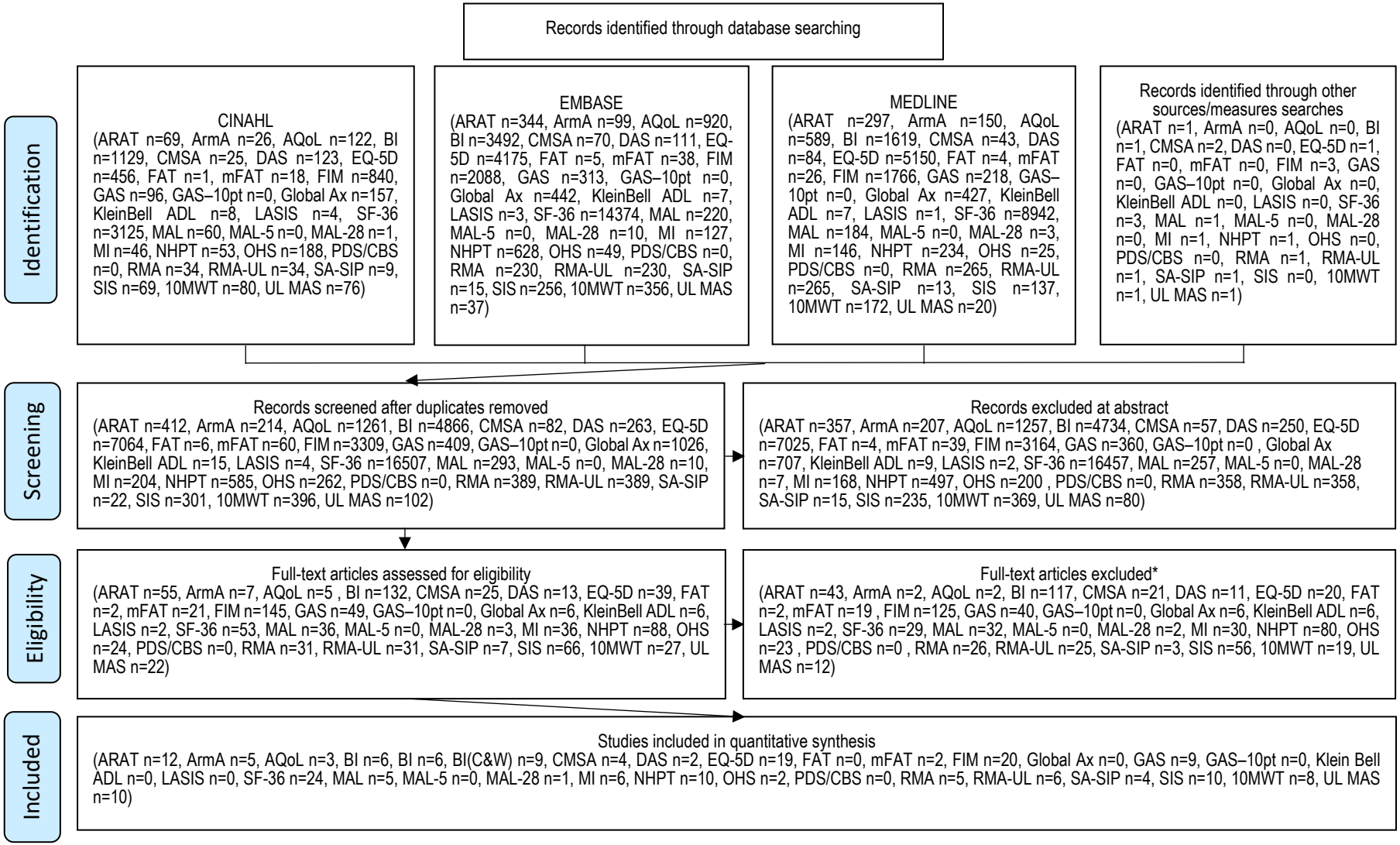
4.5.1. Flow of Studies

The electronic search strategy located 56,288 studies across the individual measurement tools. After screening titles, abstracts and full text, 156 psychometric studies (some evaluating more than one included tool) were included in this systematic review. Our systematic search did not locate any studies evaluating the psychometric properties of the following: Frenchay Arm Test (De Souza, Langton-Hewer, & Miller, 1980), Global Assessment Scale (Smith, Ellis, White, & Moore, 2000), Goal Attainment Scale – 10 point scale (Bhakta, O'Connor, & Cozens, 2008), Klein-Bell Activities of Daily Living Scale (Klein & Bell, 1982), Motor Activity Log-5 (Chang et al., 2009), Leeds Adult Spasticity Impact Scale (Bhakta, Cozens, Chamberlain, & Bamford, 2000) and Patient Disability Scale/Carer Burden Scale (Bhakta et al., 2000). Figure 5.1 presents the flow of papers through the review.

4.5.2. Characteristics of the Studies

The 156 included studies are outlined in Table 4.2. The majority of studies (n = 97, 62%) included post-stroke participants, and of these, most were greater than 6 months post-stroke. The remaining studies included diagnoses of multiple sclerosis (MS), traumatic brain injury (TBI) or mixed neurological participants. Sample characteristics varied across studies and these are detailed in Table 4.2; sample sizes were commonly small (range n=5 to n=148,367; mean=2011 (SD=13,310.6); median=90), with less than n=100 participants in over half of studies (57%) and only n=5 studies including greater than 10 000 participants. The number of studies evaluating each measurement tool varied,

ranging from n = 1 study investigating the Motor Activity Log-28 (MAL-28), to n = 23 for the Medical Outcome Study 36-Item Short-Form Health Survey (SF-36). Participants with upper limb spasticity were specifically identified in n = 15 studies in total (across n = 9 of the included n = 23 measurement tools).



*See Supplementary Table 2 for reasons for full text exclusions. ARAT Action Research Arm Test, ArmA=Arm Activity Measure, AQoL=Assessment of Quality of Life, BI=Barthel Index, CMSA=Chedoke-McMaster Stroke Assessment, DAS=Disability Assessment Scale, EQ-5D=EuroQol–5 dimension, FAT=Frenchay Arm Test, mFAT=modified Frenchay Arm Test, FIM=Functional Independence Measure, GAS=Goal Attainment Scale, GAS–10pt=Goal Attainment Scale–10 point, Global Ax=Global Assessment Scale, KleinBell ADL=Klein-Bell Activities of Daily Living scale, LASIS=Leeds Adult Spasticity Impact Scale, SF-36=Medical Outcome Study 36-Item Short-Form Health Survey, MAL=Motor Activity Log, MAL-5=Motor Activity Log-5, MAL-28=Motor Activity Log-28, MI=Motricity Index, NHPT=Nine Hole Peg Test, OHS=Oxford Handicap Scale, PDS/CBS=Patient Disability Scale/Carer Burden Scale, RMA=Rivermead Motor Assessment, RMA-UL=Rivermead Motor Assessment-Upper Limb, SA-SIP=Stroke-Adapted Version of the Sickness Impact Profile, SIS=Stroke Impact Scale, 10MWT=Ten Metre Walk Test, UL MAS=Upper Limb Motor Assessment Scale.

Figure 4.1. *Prisma flow chart*

Table 4.2. *Characteristics of included studies*

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Adams et al., (1997)	RMA RMA-UL	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 83 Age *(yr),* mean (SD) = Grp 1: 75.39 (6.41), Grp 2: 56.54 (5.73), Grp 3: 56.33 (5.95) Sex, number male (%) = Group 1 (51), Group 2 (62), Group 3 (54) Sample included people with spasticity = not reported	Structural validity
Adams et al., (1997)	RMA RMA-UL	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 51 Age *(yr),* mean (SD) = 74.37 (9.38) Sex, number male (%) = 24 (47) Sample included people with spasticity = not reported	Structural validity
Alderman et al., (2001)	EQ-5D	Diagnosis = Traumatic Brain Injury n = 29, Stroke n = 11 Time since diagnosis (*mo*) = greater than 6 n = 11 Age (*yr*), mean (range) = 39 (19-66) Sex, number male (%) = 42 (81) Sample included people with spasticity = not reported	Construct validity
Ali et al., (2013)	BI	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 3787 Age *(yr),* mean (median IQR) = 71 (60 – 78) Sex, number male (%) = 2715 (55) Sample included people with spasticity = not reported	Construct validity
Anderson et al., (1996)	SF-36	Diagnosis = Stroke Time since diagnosis (*mo*) = greater than 6 n = 90 Age (*yr*), mean (SD) = 72 (12) Sex, number male (%) = 48 (53) Sample included people with spasticity = not reported	Internal consistency Construct validity
Ashford et al., (2015)	ArmA	Diagnosis = Mixed (Stroke n = 15, TBI n = 1) Time since diagnosis (*mo*) = greater than 6 n = 16 Age (yr), mean (SD) = 54.5 (15.7) Sex number male (%) = 9 (56) Sample included people with spasticity = yes	Content validity
Ashford et al., (2016)	ArmA	Diagnosis = Mixed (Stroke n = 48, TBI n = 28, MS n = 6, other n = 10) Time since diagnosis *(mo)* = not reported n = 92 Age *(yr),* mean (SD) = 44.5 (16.7) Sex number male (%) = 54 (59) Sample included people with spasticity = yes	Structural validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Ashford et al., (2014)	ArmA	Diagnosis = Mixed (Stroke n = 30, MS n = 4, TBI n = 22, other n = 2) Time since diagnosis (*mo*) = not reported n = 58 Age (yr), mean (SD) = 47 (17.5) Sex number male (%) = 32 (55) Sample included people with spasticity = yes	Responsiveness
Ashford et al., (2013a)	ArmA	Diagnosis = Stroke Time since diagnosis *(mo)* = not given n = 46 (clinicians), 26 (patient, carers) Age *(yr),* median (range) = 48.5 (30-64) (patients) Sex, number male (%) = 8 (62) (patients) Sample included people with spasticity = yes	Content validity
Ashford et al., (2013)	ArmA	Diagnosis = Mixed (Stroke n = 48, TBI n = 28, MS n = 6, other n = 10) Time since diagnosis *(mo)* = not reported n = 92 Age *(yr),* mean (SD) = 44.5 (16.7) Sex, number male (%) = 54 (59) Sample included people with spasticity = yes	Internal consistency Reliability Structural validity Construct validity Responsiveness Interpretability
Barer & Murphy (1993)	BI (C&W)	Diagnosis = Stroke Time since diagnosis (*mo*) = less than 6 n = 730 Age (*yr*), mean (SD) = 73.2 (not given) Sex number male (%) = 336 (46) Sample included people with spasticity = not reported	Structural validity Construct validity Responsiveness
Barton et al., (2008)	EQ-5D	Diagnosis = Stroke Time since diagnosis (*mo*) = greater than 6 n = 62 Age ≥ 45 years Sex (all sample, not only Stroke), number male (%) = 865 (46.4) Sample included people with spasticity = not reported	Construct validity
Barton et al., (2008)	EQ-5D	Diagnosis = Stroke Time since diagnosis *(mo)* = not reported n = 57 Age (all sample, not only Stroke) *(yr),* mean (range) = 64.7 (45-99) Sex (all sample, not only Stroke), number male (%) = 835 (44.8) Sample included people with spasticity = not reported	Construct validity Interpretability
Beebe & Lang (2009b)	ARAT NHPT	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 33 Age *(yr),* mean (SD) = 53.9 (10.2) Sex, number male (%) = 19 (58) Sample included people with spasticity = yes	Construct validity Responsiveness

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Benedict et al., (2011)	NHPT	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = not reported n = 211 Age *(yr),* mean (SD) = 46.2 (8.9) Sex, number male (%) = 32 (27) Sample included people with spasticity = not reported	Construct validity
Bohannon (1999)	MI	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 10 Age (*yr*), mean (range) = 66.7 (46 – 81) Sex, number male (%) = not given Sample included people with spasticity = not reported	Internal consistency Construct validity
Bovend'Eerdt et al., (2011)	GAS	Diagnosis = Mixed (Stroke n = 27, TBI n =1, MS n = 1) Time since diagnosis *(mo)* = less than 6 n = 29 Age *(yr),* mean (SD) = 50.28 (13.88) Sex, number male (%) = 18 (62) Sample included people with spasticity = not reported	Reliability Measurement error
Brashear et al., (2002)	DAS	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 10 raters Age *(yr),* mean (SD) = 59.9 (16.17) Sex, number male (%) = 5 (56) Sample included people with spasticity = yes	Reliability Content validity
Brock et al., (2009)	GAS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 45 patients 23 carers Age *(yr),* median (range) = 66 (35-87) Sex, number male (%) = (56) Sample included people with spasticity = not reported	Construct validity
Brown et al., (2015)	FIM	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 148 367 Age *(yr),* mean (SD) = 70.6 (13.1) Sex, number male (%) = 71,726 (48) Sample included people with spasticity = not reported	Construct validity Interpretability
Burridge et al., (2009)	ARAT	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 17 Age *(yr),* mean (SD) = 57 (13.4) Sex, number male (%) = 11 (65) Sample included people with spasticity = yes	Construct validity
Carr et al., (1985)	UL-MAS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 5 Age *(yr),* mean (range) = 65 (55-78) Sex, number male (%) = 1 (20) Sample included people with spasticity = not reported	Reliability Content validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Chen et al., (2012)	MAL	Diagnosis = Stroke Time since diagnosis *(mo)* = 3 – 9 n = 116 Age *(yr),* range = Intervention grp 60.98 (13.47) Control grp 63.26 (12.56) Sex, number male (%) = Intervention grp 69 (65) Control grp 73 (63) Sample included people with spasticity = not reported	Measurement error Interpretability
Collin & Wade (1990)	MI RMA – UL	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 20 (reliability), n= 14 (concurrent validity) Age *(yr)* mean (range) = 56.1 (15 - 77) Sex number male (%) = 24 (67) Sample included people with spasticity = not reported	Reliability Construct validity
Collin et al., (1988)	BI (C&W)	Diagnosis = Mixed (Stroke n = 13, Traumatic Brain Injury n = 11, other n = 1) Time since diagnosis (*mo*) = less than 6 n = 25 Age (*yr*), range = 12 – 66 Sex number male (%) = 124 (52) Sample included people with spasticity = not reported	Reliability Content validity
Corrigan et al., (1997)	FIM	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = greater than 6 n = 95 Age *(yr),* mean (SD) = 35.2 (not given) Sex, number male (%) = 67 (70) Sample included people with spasticity = not reported	Construct validity
Costelloe et al., (2008)	NHPT	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = not reported n = 150 Age *(yr),* mean (SD) = not given Sex, number male (%) = not given Sample included people with spasticity = not reported	Construct validity Interpretability
Cullen et al., (2014)	FIM	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = greater than 6 n = 59 Age *(yr),* mean (SD) = drivers 49.77 (15.25) non-driver 51.42 (15.73) Sex, number male (%) = driver 28 (80) non-driver 19 (79) Sample included people with spasticity = not reported	Construct validity
Cuthbert et al., (2015)	FIM	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = greater than 6 n = 64081 Age *(yr),* mean = 76% less than 80 Sex, number male (%) = 41204 (64.3) Sample included people with spasticity = not reported	Construct validity
Dang et al., (2011)	CMSA	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 74 Age *(yr),* mean (SD) = 65.3 (12.4) Sex, number male (%) = 48 (65) Sample included people with spasticity = not reported	Construct validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Demeurisse et al., (1980)	MI	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 100 Age *(yr),* mean (SD) = 69 (not reported) Sex, number male (%) = 59 (59) Sample included people with spasticity = not reported	Content validity
Dennis et al., (2000)	BI (C&W)	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 417 Age *(yr)*, mean (SD) = 64.6 (not given) Sex number male (%) = not reported Sample included people with spasticity = not reported	Construct validity
De Weerdt et al., (1985)	ARAT	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 53 Age *(yr),* mean (SD) = 68.6 (9.3) Sex, number male (%) = 25 (47) Sample included people with spasticity = not reported	Construct validity Responsiveness
Doan et al., (2012)	DAS EQ-5D SA-SIP30	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 279 Age *(yr),* mean (range) = 58.2 (21 – 88) Sex, number male (%) = 150 (54) Sample included people with spasticity = yes	Construct validity
Doig et al., (2010)	GAS	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = greater than 6 n = 14 Age *(yr),* range = 18 – 57 Sex, number male (%) = 12 (86) Sample included people with spasticity = not reported	Construct validity Responsiveness
Donovan et al., (2008)	10MWT	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 30 Age *(yr),* mean (SD) = 61.3 (11.1) Sex, number male (%) = 21 (70) Sample included people with spasticity = not reported	Construct validity
Dorman et al., (1999)	SF-36 EQ-5D	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 531 Age *(yr),* mean (SD) = not reported Sex, number male (%) = not reported Sample included people with spasticity = not reported	Construct validity Interpretability
Dorman et al., (1998)	SF-36 EQ-5D	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 209 Age *(yr),* mean = 70 Sex, number male (%) = 147 (54) Sample included people with spasticity = not reported	Internal consistency Reliability
Dorman et al., (1997)	EQ-5D	Diagnosis = Stroke Time since diagnosis *(mo)* = not reported n = 152 Age % of sample by group <50 = 5%, 50-70 = 46%, >70=49%. Sex, number male (%) = not reported Sample included people with spasticity = not reported	Construct validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Dromerick et al., (2006)	ARAT MAL	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 39 Age *(yr),* mean (SD) = 64.54 (14.13) Sex, number male (%) = 17 (44) Sample included people with spasticity = not reported	Construct validity Interpretability
Duncan et al., (2003)	SIS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 696 Age *(yr),* mean (SD) = 68.6 (12.5) Sex, number male (%) = 386 (55) Sample included people with spasticity = not reported	Content validity Structural validity
Duncan et al., (2002)	SIS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 287 Age *(yr),* mean (SD) = 72.6 (10), 59.8 (15.5) Sex, number male (%) = 135 (47), 78 27.2) Sample included people with spasticity = not reported	Reliability Construct validity
Duncan et al., (2005)	SIS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 26 Age *(yr),* mean (SD) = mail sample 68.48 (11.4) telephone sample 68.84 (12.2) Sex, number male (%) = mail sample 219 (97.8) telephone sample 230 (98.3) Sample included people with spasticity = not reported	Internal consistency Reliability
Duncan et al., (1997)	SF-36	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 200 Age *(yr),* mean (SD) = 63 (13) Sex, number male (%) = 164 (54) Sample included people with spasticity = not reported	Construct validity
Duncan et al., (1999)	SIS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 91 Age *(yr),* mean (SD) = minor stroke 69.2 (10.1) moderate stroke 71.9 (11.7) Sex, number male (%) = 42 (46) Sample included people with spasticity = not reported	Content validity
Edwards et al., (2006)	SA-SIP30	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 219 Age *(yr),* mean (SD) = 64.74 (15.87) Sex, number male (%) = 94 (43) Sample included people with spasticity = not reported	Construct validity
Egan et al., (2014)	FIM	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 55 Age *(yr),* mean (SD) = 64.8 (13.3) Sex, number male (%) = 39 (58) Sample included people with spasticity = not reported	Construct validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Eriksson et al., (2013)	SIS	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 116 Age *(yr),* mean (SD) = 62.4 (12.7) Sex number male (%) = 56 (48) Sample included people with spasticity = not reported	Construct validity Interpretability
Filiatrault et al., (1991)	BI	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 18 Age *(yr)*, mean (SD) = 52.2 (13.5) Sex number male (%) = 12 (67) Sample included people with spasticity = not reported	Construct validity Responsiveness
Fisk et al., (2005)	EQ-5D	Diagnosis = Multiple Sclerosis Time since diagnosis = not given n = 187 Age *(yr),* mean (SD) = 51 (10) Sex, number male (%) = 47 (25) Sample included people with spasticity = not reported	Construct validity
Findler et al., (2001)	SF-36	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = greater than 6 n = 326 Age (yr), mean (SD) = 41.7 (10.8) mild, 35.7 (9.8) moderate-severe Sex, number male (%) = 130 (88) Sample included people with spasticity = not reported	Construct validity
Fleming et al., (2014)	ARAT	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 33 Age *(yr)*, mean (SD) = 61.5 (14.2) Sex, number male (%) = 20 (61) Sample included people with spasticity = yes	Construct validity Interpretability
Freeman et al., (2000)	SF-36	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 149 Age (yr), mean (SD) = 44.6 (10.8) Sex, number male (%) = (32) Sample included people with spasticity = not reported	Internal consistency Construct validity Responsiveness Interpretability
Freeman et al., (1996)	SF-36	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 50 Age *(yr),* mean (SD) = 44.8 (9.8) Sex, number male (%) = 21 (42) Sample included people with spasticity = not reported	Construct validity Interpretability
Gillard et al., (2015)	EQ-5D	Diagnosis = Stroke Time points since diagnosis *(mo)* = greater than 6 n = 460 Age *(yr),* mean (SD) = 67 (14) Sex, number male (%) = 241 (52) Sample included people with spasticity = yes	Construct validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Goodkin et al., (1988)	NHPT	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)*= greater than 6 n = Exp 68, Control 21 Age *(yr),* mean (SD) = Exp 47.16 (11.3) Control 45.24 (16.50) Sex number male (%) = Exp 25 (37) Control 7 (33) Sample included people with spasticity = not reported	Construct validity Interpretability
Gowland (1990)	CMSA	Diagnosis = Stroke Time since diagnosis *(mo)* = not reported n = not reported Age *(yr),* mean (range) = not reported Sex, number male (%) = not reported Sample included people with spasticity = not reported	Content validity
Gowland et al., (1993)	CMSA	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 32 Age *(yr),* mean (range) = 64, (18 – 86) Sex, number male (%) = 14 (44) Sample included people with spasticity = not reported	Reliability Construct validity Responsiveness
Grant et al., (2014)	FIM	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 11983 Age *(yr),* median (25^{th}, 75^{th} percentile) = 72 (61, 81) Sex, number male (%) = 6581 (55) Sample included people with spasticity = not reported	Construct validity
Green et al., (2001)	BI (C&W)	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 22 Age *(yr),* mean (SD) = 71.6 (6.8) Sex number male (%) = 16 (73) Sample included people with spasticity = not reported	Reliability Measurement error
Guilfoyle et al., (2010)	SF-36	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = mixed, mean less than 6 n = 453 Age *(yr),* mean (SD) = 36.6 (16.1) Sex, number male (%) = 392 (76.3) Sample included people with spasticity = not reported	Internal consistency Structural validity Construct validity Interpretability
Hagen et al., (2003)	SF-36	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 136 Age *(yr),* mean (SD) = 70 (11) Sex, number male (%) = 69 (51) Sample included people with spasticity = not reported	Internal consistency Construct validity Responsiveness Interpretability
Hall et al., (1993)	FIM	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = less than 6 n = 332 Age *(yr),* mean (SD) = 34.5 (16) Sex, number male (%) = 259 (78) Sample included people with spasticity = not reported	Structural validity Construct validity Interpretability

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Hamilton & Granger (1994)	FIM	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 1018 Age *(yr),* mean (SD) = 71 (12) Sex, number male (%) = 478 (47) Sample included people with spasticity = not reported	Reliability
Harris & Eng (2007)	MAL	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 93 Age *(yr),* mean (SD) = 68.7 (9.4) Sex, number male (%) = 61 (65) Sample included people with spasticity = yes	Construct validity
Hawthorne et al., (2009)	AqoL	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = greater than 6 n = 56 Age *(yr),* mean (SD) = 39 (15) Sex, number male (%) = 40 (71) Sample included people with spasticity = not reported	Construct validity
Hawthorne et al., (1999)	AqoL	Diagnosis = Mixed (medical and musculoskeletal diagnoses, healthy samples) Time since diagnosis *(mo)* = less than 6 n = 255 Age *(yr),* range = ≤29 – 70+ Sex, number male (%) = 121 (47) Sample included people with spasticity = not reported	Content validity
Heinemann et al., (1997)	FIM	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = less than 6 n = 129 Age *(yr),* mean (SD) = 37.4 (19.5) Sex, number male (%) = (71) Sample included people with spasticity = not reported	Construct validity
Heinemann et al., (1993)	FIM	Diagnosis = Mixed (Stroke n = 10092) Time since diagnosis *(mo)* = less than 6 n = 10092 Age *(yr),* mean (SD) = 62.1 (not given) whole sample Sex, number male (%) = 5349 (53) whole sample Sample included people with spasticity = not reported	Structural validity
Heinemann et al., (1994)	FIM	Diagnosis = Mixed (Stroke n = 9961) Time since diagnosis *(mo)* = less than 6 n = 9961 Age *(yr),* mean (SD) = 70.4 (not reported) Sex, number male (%) = 4781 (48) Sample included people with spasticity = not reported	Structural validity
Heller et al., (1987)	mFAT NHPT	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 10 Age (yr) = not provided Sex, number male (%) = not reported Sample included people with spasticity = not reported	Reliability
Heller et al., (1987)	mFAT NHPT	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 56 Age (yr) = 68.1 (11.4) Sex, number male (%) = 24 (43) Sample included people with spasticity = not reported	Construct validity Interpretability

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Hermann et al., (1996)	SF-36	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 85 Age *(yr),* mean (SD) = 44.6 () Sex, number male (%) = 20 (23) Sample included people with spasticity = not reported	Construct validity
Hirsch et al., (2014)	10MWT	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = less than 6 n = 23 Age *(yr),* mean (SD) = 35.8 (14.2) Sex, number male (%) = 22 (96) Sample included people with spasticity = not reported	Reliability Construct validity
Hobart et al., (2002)	SF-36	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 177 Age *(yr),* mean (SD) = 62 (13) Sex, number male (%) = 126 (71) Sample included people with spasticity = not reported	Internal consistency Structural validity Interpretability
Houlden et al., (2006)	FIM BI (C&W)	Diagnosis = Mixed (Stroke n = 261, Traumatic Brain Injury n = 107) Time since diagnosis (*mo*) = less than 6 n = 368 Age (*yr*), mean (SD) = whole sample not reported Sex number male (%) = 259 (63) Sample included people with spasticity = not reported	Responsiveness Interpretability
Jacob-Lloyd et al., (2005)	MI NHPT	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 58 Age *(yr)* number (%) = 47 (85) older than 60 Sex, number male (%) = 31 (53) Sample included people with spasticity = not reported	Construct validity Responsiveness Interpretability
Jenkinson et al., (2013)	SIS	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 73 Age *(yr)* range = 18 - >75 Sex, number male (%) = 88 (58) Sample included people with spasticity = not reported	Internal consistency Structural validity
Johnson & Selfe (2004)	UL-MAS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 26 Age *(yr)* mean (SD) = 77 (9) Sex, number male (%) = 13 (50) Sample included people with spasticity = not reported	Internal consistency
Jones (1998)	RMA	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 29 Age *(yr)* mean (SD) = 66 (9.4) Sex, number male (%) = 13 (50) Sample included people with spasticity = not reported	Construct validity
Joyce et al., (1994)	GAS	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = less than 6 n = 16 Age *(yr)* mean (range) = 27 (17 - 49) Sex, number male (%) = 9 (56) Sample included people with spasticity = not reported	Reliability Content validity Construct validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Khan et al., (2013)	UL-MAS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 481 Age *(yr)* range = 18-101 Sex, number male (%) = 255 (53) Sample included people with spasticity = not reported	Structural validity Construct validity
Khan et al., (2008)	GAS	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 24 (203 goals) Age *(yr)* mean (SD) = 52 (8.3) Sex, number male (%) =10 (42) Sample included people with spasticity = not reported	Construct validity Responsiveness
Keith et al., (1987)	FIM	Diagnosis = not reported Time since diagnosis *(mo)* = not reported n = not reported Age *(yr),* mean (SD) = not reported Sex, number male (%) = not reported Sample included people with spasticity = not reported	Content validity
Kohn et al., (2014)	EQ-5D	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 3044 Age *(yr),* mean (SD) = 56.8 (9.9) Sex, number male (%) = 600 (20) Sample included people with spasticity = not reported	Construct validity Responsiveness
Kuspinar et al., (2014)	EQ-5D	Diagnosis = MS Time since diagnosis *(mo)* = greater than 6 n = 189 Age *(yr),* mean (SD) = 43 (10) Sex, number male (%) = 49 (26) Sample included people with spasticity = not reported	Construct validity
Kuspinar & Mayo (2013)	EQ-5D	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 185 Age *(yr),* mean (SD) = 42.8 (10) Sex, number male (%) = 48 (26) Sample included people with spasticity = not reported	Content validity Construct validity
Kuys et al., (2009)	10MWT FIM UL-MAS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 105 Age *(yr)* median = 70 (13) Sex, number male (%) = 64 (53) Sample included people with spasticity = not reported	Construct validity
Kwon et al., (2006)	SIS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 95 Age *(yr)* median = 70 (13) Sex, number male (%) = 64 (53) Sample included people with spasticity = not reported	Construct validity Interpretability
Kwon et al., (2004)	BI	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 1680 Age *(yr),* mean (SD) = 70 (11.4) Sex number male (%) = 790 (47) Sample included people with spasticity = not reported	Construct validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Lai et al., (2002)	SIS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 81 Age *(yr),* mean (SD) = 76 (6.56) Sex number male (%) = 48 (59) Sample included people with spasticity = not reported	Construct validity Interpretability
Lang et al., (2008)	ARAT	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 12 Age *(yr),* mean (SD) = 64 (14) Sex, number male (%) = 21 (40) Sample included people with spasticity = not reported	Interpretability
Lang et al., (2006)	ARAT	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 50 Age *(yr),* mean (SD) = 63.7 (13.6) Sex, number male (%) = 21 (42) Sample included people with spasticity = yes	Construct validity Responsiveness
Lannin (2003)	GAS	Diagnosis = mixed (Stroke, Traumatic Brain Injury) Time since diagnosis *(mo*) = greater than 6 n =12 Age *(yr),* mean (range) = 56.5 (26-79) Sex, number male (%) = not reported Sample included people with spasticity = not reported	Responsiveness
Lannin (2004)	UL-MAS	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 27 Age *(yr),* mean (SD) = 67 (10.1) Sex, number male (%) = 15 (50) Sample included people with spasticity = not reported	Internal consistency Structural validity
Lincoln & Leadbitter (1979)	RMA	Diagnosis = Stroke Time since diagnosis *(mo*) = not reported n = 51 Age *(yr),* range = 17 - 65 Sex, number male (%) = not reported Sample included people with spasticity = not reported	Content validity
Loewen & Anderson (1988)	UL-MAS	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 7 Age *(yr),* mean (SD) = 73.6 (8.3) Sex, number male (%) = 2 (29) Sample included people with spasticity = not reported	Reliability
Loewen & Anderson (1990)	UL-MAS	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 50 Age *(yr),* mean (SD) = 68 (10) Sex, number male (%) = 28 (56) Sample included people with spasticity = not reported	Construct validity
Lyle (1981)	ARAT	Diagnosis = Mixed (Stroke n = unknown, Traumatic Brain Injury n = unknown) Time since diagnosis *(mo*) = Greater than 6) n = 20 Age *(yr),* mean (range) = 53.2 (26 – 72) Sex, number male (%) = 13 (65) Sample included people with spasticity = not reported	Content validity Structural validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Mackenzie et al., (2002)	SF-36	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo)* = greater than 6 n = 1197 Age *(yr),* range = 18 – 54 Sex, number male (%) = 790 (66) Sample included people with spasticity = not reported	Structural validity Construct validity
Madden et al., (2006)	SF-36	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 116 Age *(yr),* mean (range) = 70 (10) Sex, number male (%) = 57 (49) Sample included people with spasticity = not reported	Construct validity Responsiveness Interpretability
Mahoney & Barthel (1965)	BI	Diagnosis = not given Time since diagnosis *(mo)* = not given n = not given Age *(yr),* mean (range) = not given Sex, number male (%) = not given Sample included people with spasticity = not reported	Content validity
Malec (1999)	GAS	Diagnosis = Mixed (Traumatic Brain Injury n = 66, Stroke n = 15, other n = 7) Time since diagnosis *(mo*) = greater than 6 (61%) n = 88 Age *(yr),* mean (range) = 33.8 (18 – 69) Sex number male (%) = 64 (72.7) Sample included people with spasticity = not reported	Construct validity
Malec et al., (1991)	GAS	Diagnosis = Traumatic Brain Injury Time since diagnosis *(mo*) = greater than 6 n = 14 Age *(yr),* mean (SD) = 34.3 (12.2) Sex, number male (%) = not reported Sample included people with spasticity = not reported	Construct validity
Miller et al., (2010)	UL-MAS	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 80 Age *(yr),* mean (SD) = 67.4 (15.6) Sex, number male (%) = 46 (58) Sample included people with spasticity = not reported	Internal consistency Structural validity Construct validity Interpretability
Miller et al., (2013)	10MWT	Diagnosis = Stroke Time since diagnosis *(mo*) = greater than 6 n = 77 Age *(yr),* mean (range) = 64.1 (48 – 89) Sex, number male (%) = 58 (75) Sample included people with spasticity = not reported	Construct validity
Moore et al., (2004)	SF-36 EQ-5D	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 114 Age *(yr),* mean (SD) = 45 (11) Sex, number male (%) = 18 (45) Sample included people with spasticity = not reported	Construct validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Moreland et al., (1993)	CMSA	Diagnosis = Stroke Time since diagnosis *(mo*) = not reported n = not reported Age *(yr),* median (range) = not reported Sex, number male (%) = not reported Sample included people with spasticity = not reported	Content validity
Morris et al., (2013)	ARAT NHPT RMA – UL	Diagnosis = Stroke Time since diagnosis *(mo*) = greater than 6 n = 85 Age *(yr),* median (range) = 69 (36 – 88) Sex, number male (%) = 49 (58) Sample included people with spasticity = not reported	Construct validity Interpretability
Mudge & Stott (2009)	10MWT	Diagnosis = Stroke Time since diagnosis *(mo*) = greater than 6 n = 49 Age *(yr),* mean (SD) = 67.4 (12.5) Sex, number male (%) = 29 (59) Sample included people with spasticity = not reported	Construct validity
Murrell et al., (1999)	SF-36	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 22 Age *(yr),* mean (SD) = 52.4 (9.9) Sex, number male (%) = 9 (40) Sample included people with spasticity = not reported	Reliability
Nicholl et al., (2001)	EQ-5D	Diagnosis = Multiple Sclerosis Time points since diagnosis *(mo)* = greater than 6 n = 88 Age *(yr),* mean (SD) = 48.97 (8.9) Sex, number male (%) =24 (25) Sample included people with spasticity = not reported	Construct validity Interpretability
Oczkowski et al., (1993)	FIM	Diagnosis = Stroke Time since diagnosis (*mo*) = less than 6 n =113 Age *(yr),* mean = 65.7 (female) 65.8 (male) Sex, number male (%) = 59 (52.2) Sample included people with spasticity = not reported	Construct validity
O'Mahony et al., (1998)	SF-36	Diagnosis = Stroke Time since diagnosis *(mo)* = not reported n = 104 Age *(yr),* mean (range) = > 45 Sex, number male (%) = not reported Sample included people with spasticity = not reported	Interpretability
Ouellette et al., (2015)	FIM	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 407 Age *(yr),* mean (SD) = 68.2 (13.9) Sex, number male (%) = not given Sample included people with spasticity = not reported	Construct validity
Peters et al., (2014)	EQ-5D	Diagnosis = Stroke Time since diagnosis *(mo)* = not reported n = 102 Age *(yr)* = 78% > 55 Sex, number male (%) = 53 (53) Sample included people with spasticity = not reported	Responsiveness

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Pickard et al., (2005)	EQ-5D	Diagnosis = Stroke Time points since diagnosis *(mo)* = less than 6 n = 96 Age *(yr),* mean (SD) = 67 (15) Sex, number male (%) =51 (52) Sample included people with spasticity = not reported	Responsiveness Interpretability
Pickering et al., (2010)	UL-MAS	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 25 Age *(yr),* mean (SD) = 69.96 (11.97) Sex, number male (%) = 14 (56) Sample included people with spasticity = not reported	Structural validity Interpretability
Pittock et al., (2004)	SF-36	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 185 Age *(yr),* mean (SD) = not given Sex, number male (%) = 56 (30) Sample included people with spasticity = not reported	Construct validity
Poole et al., (2010)	NHPT	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo*) = greater than 6 n = 56 Age *(yr),* mean (SD) = 46.8 (10.48) Sex, number male (%) = 11 (20) Sample included people with spasticity = not reported	Construct validity
Rabadi & Rabadi (2006)	ARAT	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 104 Age *(yr),* mean (SD) = 72.0 (13) Sex, number male (%) = 43 (41) Sample included people with spasticity = not reported	Construct validity Responsiveness
Rabadi & Vincent (2013)	FIM	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 76 Age *(yr),* mean (SD) = 53.6 (10.9) Sex, number male (%) = 63 (83) Sample included people with spasticity = yes	Construct validity Responsiveness
Rand & Eng (2015)	ARAT	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 32 Age *(yr),* mean (SD) = 58.1 (12.4) Sex, number male (%) = 25 (78) Sample included people with spasticity = not reported	Construct validity
Riazi et al., (2003)	SF-36	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 638 Age *(yr),* range = 20 - >60 Sex, number male (%) = 219 (35) Sample included people with spasticity = not reported	Construct validity
Rigby et al., (2009)	OHS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 104 Age *(yr),* mean (SD) = 72.0 (13) Sex, number male (%) = 43 (41) Sample included people with spasticity = not reported	Construct validity

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Robinson et al., (2009)	SF-36	Diagnosis = MS Time since diagnosis *(mo)* = greater than 6 n = 249 Age *(yr),* mean (range) = 39 (10.5) Sex, number male (%) = 75 (30) Sample included people with spasticity = not reported	Construct validity Interpretability
Sabari et al., (2005)	UL-MAS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 (83%) n = 100 Age *(yr),* mean (range) = 54 (18 – 94) Sex, number male (%) = 67 (67) Sample included people with spasticity = not reported	Structural validity Interpretability
Sackley (1990)	RMA RMA-UL	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 52 (R hemiparesis), 38 (L hemiparesis) Age *(yr),* mean (SD) = 63.4 (11.4) (R hemiparesis), 63.2 (11.9) (L hemiparesis) Sex, number male (%) = 33 (64) (R hemiparesis), 23 (61) (L hemiparesis) Sample included people with spasticity = not reported	Construct validity
Salbach et al., (2001)	10MWT	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 50 Age *(yr),* mean (SD) = 68.0 (13) Sex, number male (%) = 31 (62) Sample included people with spasticity = not reported	Responsiveness
Salter et al., (2008)	SF-36 EQ-5D SIS	Diagnosis = Stroke Time since diagnosis (*mo*) = not reported n = not reported Age (*yr*), mean (SD) = not reported Sex, number male (%) = not reported Sample included people with spasticity = not reported	Content validity
Sarker et al., (2012)	BI (C&W)	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 238 Age *(yr),* mean (SD) = 68.6 (14.2) Sex number male (%) = 124 (52) Sample included people with spasticity = not reported	Construct validity Interpretability
Schmid et al., (2012)	10MWT	Diagnosis = Stroke Time since diagnosis *(mo*) = greater than 6 n = 77 Age *(yr),* mean (SD) = 64.06 (8.78) Sex, number male (%) = 58 (75) Sample included people with spasticity = not reported	Construct validity
Schwid et al., (2002)	NHPT	Diagnosis = Multiple Sclerosis Time since diagnosis = unknown n = 27 Age *(yr),* mean (SD) = 51.9 (9.0) Sex, number male (%) = 16 (79) Sample included people with spasticity = not reported	Measurement error
Scrivener et al., (2014)	10MWT	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 190 Age *(yr),* mean (SD) = 76.0 (12.7) Sex, number male (%) = 97 (51) Sample included people with spasticity = not reported	Responsiveness Interpretability

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Sharrack et al., (1999)	BI (C&W) FIM	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 25 – 64 Age *(yr),* median (range) = 40 (42.1 – 77.6) Sex, number male (%) = 22 (34) Sample included people with spasticity = not reported	Internal consistency Reliability Structural validity Construct validity Responsiveness
Simon et al., (2008)	OHS	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 53 Age *(yr),* mean (SD) = 65.6 (12.1) Sex, number male (%) = 14 (28) Sample included people with spasticity = not reported	Construct validity
Stineman et al., (1996)	FIM	Diagnosis = mixed (Stroke = 26, 183, Traumatic Brain Injury = 3, 214) Time since diagnosis *(mo)* = less than 6 n = 29 397 Age *(yr),* mean range = 41.6 – 71.3 Sex, number male (%) = not reported Sample included people with spasticity = not reported	Internal consistency Structural validity
Stone et al., (1993)	MI	Diagnosis = Stroke Time since diagnosis (*mo)* = less than 6 n = 84 Age *(yr),* mean (SD) = 72.37 (12.11) Sex, number male (%) = not given Sample included people with spasticity = not reported	Construct validity
Sturm et al., (2002)	AqoL	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 93 Age *(yr),* mean (range) = 72 (28 – 89) Sex, number male (%) = 42 (45) Sample included people with spasticity = not reported	Construct validity Interpretability
Turner-Stokes et al., (2010)	GAS	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 90 Age *(yr),* mean (SD) = 54.5 (13.2) Sex, number male (%) = 54 (60) Sample included people with spasticity = yes	Construct validity
Uswatte & Taub (2005)	MAL	Diagnosis = not reported Time since diagnosis (*mo)* = not reported n = not reported Age *(yr),* mean (SD) = not reported Sex number male (%) = not reported Sample included people with spasticity = not reported	Content validity
Uswatte et al., (2006)	MAL MAL-28	Diagnosis = Stroke Time since diagnosis (*mo)* = greater than 6 n = 222 Age *(yr),* mean (SD) = 62.2 (13.0) Sex number male (%) = 142 (64) Sample included people with spasticity = not reported	Internal consistency Reliability Content validity Structural validity Interpretability

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Van der Putten et al., (1999)	BI (C&W) FIM	Diagnosis = Mixed (Stroke n = 82, Multiple Sclerosis n = 201) Time since diagnosis *(mo)* = less than 6 n = 283 Age *(yr),* mean (SD) = 52 (16.9) (Stroke), 45 (11.2) (Multiple Sclerosis) Sex number male (%) = 238 (84) Sample included people with spasticity = not reported	Responsiveness Interpretability
Van Straten et al (1997)	SA-SIP30	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 319 Age *(yr),* mean (SD) = 69 (12.6) Sex number male (%) = 175 (55) Sample included people with spasticity = not reported	Content validity
Vickrey et al., (1997)	SF-36	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 171 (internal consistency, hypothesis testing), n = 84 (reliability) Age *(yr),* mean (range) = 45 (20 – 67) Sex, number male (%) = 123 (72) Sample included people with spasticity = not reported	Internal consistency Reliability Construct validity
Vickrey et al., (1995)	SF-36	Diagnosis = Multiple Sclerosis Time since diagnosis *(mo)* = greater than 6 n = 179 Age *(yr),* mean (range) = 45 (20 – 67) Sex, number male (%) = 129 (72) Sample included people with spasticity = not reported	Construct validity
Wade & Hewer (1987)	BI (C&W) MI	Diagnosis = Stroke Time since diagnosis (*mo)* = less than 6 n = 976 Age *(yr),* mean (SD) = not given Sex, number male (%) = not given Sample included people with spasticity = not reported	Structural validity Construct validity
Wallace et al., (2002)	BI	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 372 Age *(yr),* mean (SD) = 69.7 (11.6) Sex number male (%) = 177 (48) Sample included people with spasticity = not reported	Responsiveness
Ware & Sherbourne (1992)	SF-36	Diagnosis = not reported Time since diagnosis *(mo)* = not reported n = not reported Age *(yr),* mean (SD) = not reported Sex number male (%) = not reported Sample included people with spasticity = not reported	Content validity
Wellwood et al., (1995)	BI	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 152 Age *(yr),* mean (SD) = 73 (13.4) Sex number male (%) = 68 (45) Sample included people with spasticity = not reported	Construct validity Interpretability

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Wilkinson et al., (1997)	BI (C&W)	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 106 Age *(yr),* median (range) = 71 (34-79) Sex number male (%) = 57 (54) Sample included people with spasticity = not reported	Construct validity Interpretability
Williams et al., (1999)	SF-36	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 71 Age *(yr),* mean (SD) = 61 (13) Sex, number male (%) = 45 (63) Sample included people with spasticity = not reported	Construct validity
Williams (1990)	EQ-5D	Diagnosis = not reported Time since diagnosis *(mo)* = not reported n = not reported Age *(yr),* mean (SD) = not reported Sex, number male (%) = not reported Sample included people with spasticity = not reported	Content validity
Wolf & Koster et al., (2013)	SIS	Diagnosis = Stroke Time since diagnosis (*mo)*= greater than 6 n = 96 Age *(yr),* median (range) = Grp 1 64.2 (13.4), Grp 2 60.5 (12.8) Sex, number male (%) = Grp 1 28 (52), Grp 2 31 (55) Sample included people with spasticity = not reported	Construct validity
Xie et al., (2006)	EQ-5D	Diagnosis = Stroke Time since diagnosis *(mo)* = not reported n = 1040 Age *(yr)* = ≥18 Sex, number male (%) = 447 (43.9) Sample included people with spasticity = not reported	Construct validity
Yozbatiran et al., (2008)	ARAT	Diagnosis = Stroke Time since diagnosis (*mo)* = greater than 6 n = 12 (validity) n = 9 (interrater reliability) n = 8 (intra rater) Age *(yr),* mean (SD) = 61.0 (15.0) Sex, number male (%) = 6 (50) Sample included people with spasticity = not reported Rater characteristics Rater n =2 Clinical experience *(yr)* = 8 Observations n = 58	Reliability Construct validity

RMA = Rivermead Motor Assessment, RMA-UL = Rivermead Motor Assessment – Upper Limb, BI (C&W) = Barthel Index Collin & Wade version, EQ-5D = EuroQol -5 dimension, SIS = Stroke Impact Scale, SF-36 = Medical Outcome Study 36-Item Short-Form Health Survey, ArmA = Arm Activity Measure, ARAT = Action Research Arm Test, NHPT = Nine Hole Peg Test, MI = Motricity Index, GAS = Goal Attainment Scale, DAS = Disability Assessment Scale, FIM = Functional Independence Measure, UL-MAS = Upper Limb – Motor Assessment Scale, CMSA = Chedoke-McMaster Stroke Assessment, SA-SIP30 = Stroke-Adapted Version of the Sickness Impact Profile, 10MWT = Ten Metre Walk Test, MAL = Motor Activity Log, BI = Barthel Index, AQoL = Assessment of Quality of Life, mFAT = modified Frenchay Arm Test, OHS = Oxford Handicap Scale, MAL-28 = Motor Activity log – 28.

Table 4.3. *Synthesis of evidence*

Measurement tool	Sample	Content validity	Structural validity	Internal consistency	Cross cultural validity	Reliability			Measurement error	Construct validity	Responsiveness
						Inter	Intra	Retest			
ARAT	*Spasticity n = 4*									*Moderate - (13/21)*	*Low + (4/4)*
	Whole sample n = 12	Very Low	Very Low +			Very Low +	Very Low +			Moderate - (19/30)	Moderate + (6/6)
ArmA	*Spasticity n = 5*	*High*	*High* +	*Moderate* +				*Low* +		*Very Low* +	*Moderate + (4/4)*
	Whole sample n = 5	High	High +	Moderate +				Low +		Very Low +	Moderate + (4/4)
AQoL	*Spasticity n = 0*										
	Whole sample n = 3	Very Low								High + (3/3)	
BI	*Spasticity n = 0*										
	Whole sample n = 6	Very Low								High + (5/6)	Very Low - (0/1)
BI (C&W)	*Spasticity n = 0*										
	Whole sample n = 9	Very Low	Low +			Very Low ?		Very Low ?	Very Low +	Moderate +	Low - (2/3)
CMSA	*Spasticity n = 0*										
	Whole sample n = 4	Very Low				Moderate + Low +*	Moderate +	Low +		Moderate + (5/6)	Very Low + (1/1)
DAS	*Spasticity n = 2*	*Very Low*				*Low ?*	*Low -*			*moderate + (2/2)*	
	Whole sample n = 2	Very Low				Low ?	Low -			moderate + (2/2)	

Measurement tool	Sample	Content validity	Structural validity	Internal consistency	Cross cultural validity	Reliability			Measurement error	Construct validity	Responsiveness
						Inter	Intra	Retest			
EQ-5D	*Spasticity n = 2*									*High + (3/3)*	
	Whole sample n = 19	Moderate ?						Moderate +^ Very Low - ^^		Moderate + (24/34)	Low - (11/15)
FAT	*Spasticity n = 0*										
	Whole sample n = 0										
mFAT	*Spasticity n = 0*										
	Whole sample n = 2					Very Low ?		Very Low ?		Very Low - (0/1)	
FIM	*Spasticity n =1*									*Moderate + (1/1)*	*Very Low + (1/1)*
	Whole sample n = 20	Very Low	High +	High +		Moderate +	Low +			High + (23/29)	Moderate - (5/7)
Global Ax	*Spasticity n = 0*										
	Whole sample n = 0										
GAS	*Spasticity n = 1*									*Very Low – (3/7)*	
	Whole sample n = 9					Low -			Low ?	Moderate – (14/23)	Low + (4/4)
GAS-10pt	*Spasticity n = 0*										
	Whole sample n = 0										

Measurement tool	Sample	Content validity	Structural validity	Internal consistency	Cross cultural validity	Reliability			Measurement error	Construct validity	Responsiveness
						Inter	Intra	Retest			
Klein-Bell	*Spasticity n = 0* Whole sample n = 0										
LASIS	*Spasticity n = 0* Whole sample n = 0										
MAL	*Spasticity n = 1* Whole sample n = 5	Very Low	Very Low ?						Low ?	*Low - (3/7)* Moderate - (4/9)	
MAL-5	*Spasticity n = 0* Whole sample n = 0										
MAL-28	*Spasticity n = 0* Whole sample n = 1	Very Low	Very Low ?	Very Low +**				Moderate +^ Low -^^		Very Low + (3/4)^ Very Low – (2/4)^^	
MI	*Spasticity n = 0* Whole sample n = 6	Very Low		Very Low ?		Very Low ?				Moderate - (4/6)	Very Low - (0/1)
NHPT	*Spasticity n = 1* Whole sample n = 10					Very Low ?		Very Low ?	Very Low +	*Very Low - (3/5)* Moderate - (21/32)	*Very Low + (2/2)* Low + (3/3)

Measurement tool	Sample	Content validity	Structural validity	Internal consistency	Cross cultural validity	Reliability			Measurement error	Construct validity	Responsiveness
						Inter	Intra	Retest			
OHS	*Spasticity n = 0* Whole sample n = 2									Low - (2/3)	
PDS / CBS	*Spasticity n = 0* Whole sample n = 0										
RMA	*Spasticity n = 0* Whole sample n = 5	Very Low	Very Low -							High + (2/2)	
RMA – UL	*Spasticity n = 0* Whole sample n = 6	Very Low	Very Low +, - ^^^							High + (3/4)	
SF-36	*Spasticity n = 0* Whole sample n = 24	Very Low	Moderate ?	High +				Moderate +^ Low -^^		Moderate – (25/44)	Very Low – (0/4)
SA-SIP	*Spasticity n = 1* Whole sample n = 4	Moderate								*Moderate* *+ (1/1)* High + (3/3)	
SIS	*Spasticity n = 0* Whole sample n = 10	Moderate	High +	Moderate +		Low ?		Low +		High + (18/19)	

Measurement tool	Sample	Content validity	Structural validity	Internal consistency	Cross cultural validity	Reliability			Measurement error	Construct validity	Responsiveness
						Inter	Intra	Retest			
10MWT	*Spasticity n = 0* Whole sample n = 8							Moderate +		Moderate - (4/8)	Low - (3/6)
UL-MAS	*Spasticity n = 0* Whole sample n = 10	Very Low	Moderate +	Moderate +**		Low ?	Low ?			Moderate - (3/8)	

High = Very confident that the true measurement property lies close to that of the estimate of the measurement property. *Moderate* = Moderate confidence in the measurement property estimate. *Low* = Limited confidence in the measurement property estimate. *Very low* = Little confidence in the measurement property estimate, full definition of ratings reported in (Prinsen et al., 2018). + = sufficient, - insufficient, ? indeterminant (Prinsen et al., 2018).

*Moderate + Impairment Inventory, Low + Activity Inventory **Internal consistency evidence strength cannot exceed structural validity as per COSMIN guidelines and has been reduced accordingly. ^Patients reports ^^ proxy reports ^^^ '+' acute sample, '-' subacute sample.

ARAT = Action Research Arm Test, ArmA = Arm Activity Measure, AQoL = Assessment of Quality of Life, BI = Barthel Index, BI (C&W) = Barthel Index - Collin & Wade version, CMSA = Chedoke-McMaster Stroke Assessment, DAS = Disability Assessment Scale, EQ-5D = EuroQol – 5 dimension, FAT = Frenchay Arm Test, mFAT = modified Frenchay Arm Test, FIM = Functional Independence Measure, GAS = Goal Attainment Scale, GAS – 10pt = Goal Attainment Scale – 10 point, Global Ax = Global Assessment Scale, KleinBell ADL = Klein-Bell Activities of Daily Living scale, LASIS = Leeds Adult Spasticity Impact Scale, SF-36 = Medical Outcome Study 36-Item Short-Form Health Survey, MAL = Motor Activity Log, MAL-5 = Motor Activity Log - 5, MAL-28 = Motor Activity Log - 28, MI = Motricity Index, NHPT = Nine Hole Peg Test, OHS = Oxford Handicap Scale, PDS/CBS = Patient Disability Scale / Carer Burden Scale, RMA = Rivermead Motor Assessment, RMA-UL = Rivermead Motor Assessment - Upper Limb, SA-SIP = Stroke-Adapted Version of the Sickness Impact Profile, SIS = Stroke Impact Scale, 10MWT = Ten Metre Walk Test, UL MAS = Upper Limb Motor Assessment Scale.

4.5.3. Characteristics of Each Measurement Tool

The number of studies examining each measurement tool is presented, together with findings for all participants and then for participants with upper limb spasticity. The synthesis of evidence for each measurement tools is presented in Table 4.3. Due to the volume of data, summaries of individual study results and psychometric properties tested are tabulated within Appendix F. The following summarizes the appraisal of each tool. These have been placed in alphabetical order.

4.5.4. Action Research Arm Test

The Action Research Arm Test (ARAT) (Lyle, 1981) is an observational performance test that evaluates a person's ability to use their upper limb to handle objects using grasp, grip, pinch and gross motor movements. Twelve studies evaluated the psychometric properties of the ARAT (Beebe & Lang, 2009; Burridge et al., 2009; De Weerdt & Harrison, 1985; Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, et al., 2006; Fleming et al., 2014; Lang et al., 2008; Lang et al., 2006; Lyle, 1981; Morris et al., 2013; Rabadi & Rabadi, 2006; Rand & Eng, 2015; Yozbatiran et al., 2008), four of those studies specifically identified participants with upper limb spasticity (Beebe & Lang, 2009; Burridge et al., 2009; Fleming et al., 2014; Lang et al., 2006). The majority of studies included participants post-stroke with a single study including a mixed sample, post-stroke and TBI (Lyle, 1981).

Content validity: The Upper Extremity Function Test (UEFT) (Carroll, 1965) was modified by Lyle (1981) to produce the ARAT. No further content validity studies were identified. The ARAT was found to have sufficient relevance, but indeterminant ratings for comprehensiveness and comprehensibility and no participants were interviewed regarding those properties.

Results for whole sample: Research supports hierarchical ordering of items (Lyle, 1981) and reliability within (ICC = 0.99) and between raters (ICC 0.99) (Yozbatiran et al., 2008). The ARAT was found to correlate highly with other like-tests of activity and dexterity (r = 0.65 – 0.95) (De Weerdt & Harrison, 1985; Dromerick, Lang, Birkenmeier, Hahn,

Sahrmann, et al., 2006; Morris et al., 2013; Rabadi & Rabadi, 2006; Rand & Eng, 2015; Yozbatiran et al., 2008) and weak to moderately with the Functional Independence Measure (FIM), a more global measure of function (r = 0.47) (Rabadi & Rabadi, 2006). ARAT scores were not, however, a predictor of overall quality of life (Morris et al., 2013). The ARAT was found to be responsive over time in acute as well as chronic stroke and TBI samples (Beebe & Lang, 2009; De Weerdt & Harrison, 1985; Lang et al., 2006; Rabadi & Rabadi, 2006). ARAT was found to be equally sensitive to change as like measures when used with participants less than 6 months post-stroke (De Weerdt & Harrison, 1985; Rabadi & Rabadi, 2006). Mixed results have been reported with respect to ceiling effect in stroke populations (Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, et al., 2006; Fleming et al., 2014) and there is one study which has reported a minimal, clinically important change of 12 points (dominant) and 17 (non-dominant) (Lang et al., 2008).

Results pertaining to sample with upper limb spasticity. The ARAT correlated strongly with like measures of activity and dexterity (r = 0.69 - 0.95) (Beebe & Lang, 2009) and less with a global measure of function (Functional Independence Measure (FIM) r = 0.2 – 0.6) (Lang et al., 2006) and impairments, including grip and pinch strength, spasticity and AROM (r = - 0.28 – 0.86) (Beebe & Lang, 2009; Burridge et al., 2009; Lang et al., 2006). The ARAT was moderate to highly responsive to capture change in participants less than 6 months post-stroke (ES = 0.55 – 1.018) (Beebe & Lang, 2009; Lang et al., 2006), being as equally responsive as like measures (NHPT and Jebsen-Taylor test of hand function), more responsive than measures of impairment (pinch and grip strength), but less responsive than the SIS-Hand (ES = 0.55 – 1.018) (Beebe & Lang, 2009). Neither a floor nor ceiling effects were found in a sample of participants greater than 6 months post-stroke (Fleming et al., 2014).

4.5.5. Arm Activity Measure

The Arm Activity measure (ArmA) is a 20-item self-report tool which includes 7 passive and 13 active items to capture real arm activity in neurological populations (Ashford, Slade, et al., 2013a). Five studies (Ashford et al., 2015; Ashford et al., 2016; Ashford et al., 2014; Ashford, Slade, et al., 2013; Ashford, Turner-Stokes, et al., 2013) evaluated the

psychometric properties of the ArmA, the majority of studies included a mixed sample including participants post-stroke, TBI and MS. All included studies specifically identified participants with upper limb spasticity.

Content validity: The ArmA was developed based on goal analysis, systematic literature review and a modified Delphi survey which demonstrated relevance, comprehensiveness and comprehensibility (Ashford et al., 2015; Ashford, Slade, et al., 2013).

Results pertaining to sample with upper limb spasticity: The ArmA subscales demonstrated internal consistency (passive subscale $\alpha = 0.85$, active subscale $\alpha = 0.96$) and retest reliability (quadratic weight kappa 0.90 (CI 0.68 – 1.12), active subscale 0.93 (CI 0.71 – 1.15)) in a sample with upper limb spasticity (Ashford, Turner-Stokes, et al., 2013). The ArmA demonstrated convergent and divergent validity with passive and active items of the Leeds Adult Spasticity Scale (LASIS) and Disabilities of Arm Shoulder and Hand (DASH) (convergent: Rho 0.48; p = 0.01 to 0.63; p = 0.01; divergent: Rho 0.02; p = 0.9 to 0.23; p = 0.078) (Ashford, Turner-Stokes, et al., 2013) and was found to be responsive (Ashford et al., 2014; Ashford, Turner-Stokes, et al., 2013). Preliminary analysis suggests clinically meaningful change is indicated by 2.5 or 3 point improvement (passive subscale) and 1.1 or 2.5 point improvement (active subscale) (Ashford, Turner-Stokes, et al., 2013). The ArmA active function subscale suffered a ceiling effect (37%), however no floor effect was observed for either subscale (Ashford, Turner-Stokes, et al., 2013).

4.5.6. Assessment of Quality of Life

The Assessment of Quality of Life (AQoL) is a generic HRQoL measure that assesses independent living, social relationships, physical senses, psychological wellbeing and illness (Hawthorne et al., 1999). Three studies evaluated the psychometric properties of the AQoL, one included participants greater than 6 months post TBI (Hawthorne et al., 2009) and two less than 6 months post-stroke (Hawthorne et al., 1999; Sturm et al., 2002). Neither study specifically identified participants with upper limb spasticity.

Content validity: Development research underpinning the AQoL (Hawthorne et al., 1999) demonstrated sufficient relevance, but indeterminant ratings for comprehensiveness and comprehensibility. No other content validity studies conducted in a neurological sample were identified.

Results for whole sample: The AQoL discriminated between participants with and without TBI (effect size (ES) = 0.80), with participants post TBI scoring 2.0 utilities lower than participants without (Hawthorne et al., 2009). The AQoL correlated more strongly with measures of handicap (London Handicap Scale (LHS) r = 0.83) than disability (Barthel Index (BI) r = 0.77) or impairment (National Institute of Health Stroke Scale (NIHSS) r = -0.69) in the first 6 months post-stroke and was a significant predictors of death or institutionalization at 12 months (Sturm et al., 2002). No floor or ceiling effects (1-2%) were found in a stroke population (Sturm et al., 2002).

4.5.7. Barthel Index

The Barthel Index (BI) was initially developed to score the abilities of participants to care for themselves (Mahoney & Barthel, 1965). The BI evaluates 10 activity areas, with a maximum score of 100 indicating independence in all included areas. Six studies evaluated the psychometric properties of the BI (Ali et al., 2013; Filiatrault et al., 1991; Kwon et al., 2004; Mahoney & Barthel, 1965; Wallace et al., 2002; Wellwood et al., 1995). Five studies were completed with participants post-stroke, 4 included participants less than 6 months post-stroke (Ali et al., 2013; Filiatrault et al., 1991; Kwon et al., 2004; Wallace et al., 2002), 1 greater than 6 months post-stroke (Wellwood et al., 1995) and 1 discussed tool development with a non-specific sample (Mahoney & Barthel, 1965). No included studies specifically identified participants with upper limb spasticity.

Content validity: No research on the development of the BI was located.

Results for whole sample: The BI correlated moderately with measures of upper limb function (Fugl-Meyer Rho = 0.60) (Functional Test for the Hemiplegic/Paretic Upper Limb Rho = 0.61) (Filiatrault et al., 1991) and global measures function (FIM r_s = 0.95, p<0.0001; Modified Rankin Scale (MRS) r_s = 0.89, p<0.0001; Office of Population

Censuses and Surveys (OPCS) disability instrument r = 0.73, p<0.001) (Kwon et al., 2004; Wellwood et al., 1995). The BI was equally responsive to change within the first three months post-stroke as like global measures (FIM) (Wallace et al., 2002) and a measure of motor function (Fugl-Meyer Test) (Filiatrault et al., 1991), however determined responsiveness was low. Evidence of a ceiling effect was found in a sample greater than 6 months post-stroke (Wellwood et al., 1995).

4.5.8. Barthel Index (Collin & Wade)

The Barthel Activities of Daily Living Index (BI C&W) (Collin et al., 1988) is a modification of the original BI measurement tool, with all 10 areas of activity included but is scored in increments of 1 rather than 5 as per the original BI (Mahoney & Barthel, 1965). Nine studies evaluated the psychometric properties of the BI(C&W) (Barer & Murphy, 1993; Collin et al., 1988; Dennis et al., 2000; Green et al., 2001; Houlden et al., 2006; Sarker et al., 2012; van der Putten et al., 1999; Wade & Hewer, 1987; Wilkinson et al., 1997), 6 studies included participants post-stroke (Barer & Murphy, 1993; Dennis et al., 2000; Green et al., 2001; Sarker et al., 2012; Wade & Hewer, 1987; Wilkinson et al., 1997) and 3 included mixed samples (stroke, MS, TBI) (Collin et al., 1988; Houlden et al., 2006; van der Putten et al., 1999). No studies specifically identified participants with upper limb spasticity.

Content validity: No information presenting the methodology used to revise the original BI was found, only justification from revised test authors who felt the original five-point incremental scoring was misleading in accuracy (Collin et al., 1988).

Results for whole sample: Research supports use of a summed BI(C&W) score due to a single factor (68% of variance) underlying the scale (Wade & Hewer, 1987). While the hierarchical nature of the BI(C&W) was supported by Wade and Hewer (1987), Barer and Murphy (1993) reported a failure to meet Guttman scaling criteria. Test-retest reliability results appear mixed, with high agreement (75%) between scores but variations in kappa (-0.99 to 0.81) (Green et al., 2001). Inter-rater reliability between self-report, family, nursing staff and skilled observers was acceptable (agreement within 2 points or less for 72% of participants) (Collin et al., 1988). The BI(C&W) was strongly associated with

measures of upper limb activity (r = 0.729 – 0.826) (Motricity Index Upper Limb (MI UL) and Motricity Index (MI) total, Frenchay Activity Index (FAI)), complex daily activities (r ≥ 0.80), and disability (r_s=0.726 – 0.80) (London Handicap Scale, Modified Rankin Scale (MRS)), and less with measures of psychological wellbeing and impairments (depression, anxiety, pain) (r = 0.2 – 0.423) (Dennis et al., 2000; Sarker et al., 2012; Wade & Hewer, 1987; Wilkinson et al., 1997). Research suggests that BI(C&W) is at least equally responsive to FIM (Houlden et al., 2006; van der Putten et al., 1999). However, BI(C&W) suffered from floor and ceiling effects across the acute through to community continuum in a mixed neurorehabilitation sample (Houlden et al., 2006; Sarker et al., 2012; van der Putten et al., 1999; Wilkinson et al., 1997).

4.5.9. Chedoke-McMaster Stroke Assessment

The Chedoke-McMaster Stroke assessment (CMSA) is comprised of two parts; the impairment inventory and the activity inventory (formerly known as the disability inventory) (Gowland et al., 1993). The CMSA impairment inventory classifies participants into subgroups based on the stages of motor recovery, while the CMSA activity inventory provides a measure of activity performance. Four studies evaluated the psychometric properties of the CMSA, two included participants less than 6 months post-stroke (Dang et al., 2011; Gowland et al., 1993), two did not report on the length of time post-stroke for participants (Gowland, 1990; Moreland et al., 1993) and no study specifically identified participants with upper limb spasticity.

Content validity: Evidence located for the development of the CMSA (Gowland, 1990; Moreland et al., 1993), did not indicate participants were consulted on the comprehensiveness or comprehensibility of included items. Relevance of items for the intended purpose of assessment of patient post-stroke within the neurorehabilitation setting was sufficient, however further content validity studies were not identified.

Results for whole sample: Evidence supports the reliability of the CMSA; inter-rater (ICC 0.88 (95% CI 0.76 – 0.94) to 0.99 (95%CI 0.98 – 1.00)), intra-rater (ICC 0.93 (95% CI 0.85 – 0.96) to 0.98 (95% CI 0.95 – 0.99)), test retest (ICC 0.98 (95% CI 0.95 – 0.99)) (Gowland et al., 1993). Consistent with the definition of the CMSA, strong correlations with both

subscales and total scores for like measures of upper limb activity performance (Fugl-Meyer r = 0.95, p<0.001) and global measures of function (FIM r = 0.79, p<0.05) were demonstrated (Gowland et al., 1993). The predictive validity through use of the Gowland's predictive equations, however, were not supported due to large error associated with the predicted value (Dang et al., 2011). The CMSA was found to be more responsive than the FIM when used with participants less than 6 months post-stroke (Gowland et al., 1993).

4.5.10. Disability Assessment Scale

The Disability Assessment Scale (DAS) is a brief measure of functional disability (Brashear, Zafonte, et al., 2002). Two studies were included, both identified participants with upper limb spasticity (Brashear, Zafonte, et al., 2002; Doan et al., 2012).

Content validity: Brashear et al. (2002) reported the development of the DAS to fill the identified gap within the evaluation of functional impairment commonly seen in participants with post-stroke upper limb spasticity (i.e. dressing, hygiene, limb position, pain). No additional research underpinning measurement tool development was reported.

Results pertaining to sample with upper limb spasticity identified: Good to excellent intra-rater reliability (78% of evaluations weighted kappa ≥ .4) and good inter-rater reliability (Kendall W 0.49 (95% CI 0.30 – 1.00, p<.001) to 0.77 (95% CI 0.37 – 1.00, p<.001) was reported when used by professionals (neurologists, physiatrists, occupational therapists and physical therapists) with a mean of 6 years clinical experience (Brashear, Zafonte, et al., 2002). Greater DAS scores were found to be associated with Stroke-Adapted Version of the Sickness Impact Scale (SA-SIP) scores (P<.05), reduced quality of life and caregiver burden (P<.05) (Brashear, Gordon, et al., 2002; Doan et al., 2012).

4.5.11. EuroQol-5 Dimension

The EuroQol-5 dimension (EQ-5D) is a generic measure of health-related quality of life (Fisk et al., 2005; Gillard et al., 2015; Williams, 1990). Nineteen studies evaluated the

psychometric properties of the EQ-5D, including participants with MS (n = 6), (Fisk et al., 2005; Kohn et al., 2014; Kuspinar et al., 2014; Kuspinar & Mayo, 2013; Moore et al., 2004; Nicholl et al., 2001) a mixed neurological sample (n = 1) (Alderman et al., 2001) and post-stroke (n = 12) (Barton, Sach, Avery, et al., 2008; Barton, Sach, Doherty, et al., 2008; Doan et al., 2012; Dorman et al., 1999; Dorman et al., 1998; Dorman et al., 1997; Gillard et al., 2015; Peters et al., 2014; Pickard et al., 2005; Salter et al., 2008; Xie et al., 2006). Two studies specifically identified participants with upper limb spasticity (Doan et al., 2012; Gillard et al., 2015).

Content validity: During the development of the EQ-5D there is no evidence that participants were consulted on the comprehensiveness or comprehensibility of included items. Relevance of items for the intended purpose was sufficient (Kuspinar & Mayo, 2013). The EQ-5D contains 6 of 9 recommended dimensions for patient-based, health related quality of life measures and is less comprehensive than the Stroke Impact Scale (SIS) (Salter et al., 2008).

Results for whole sample: Test-retest reliability of the patient-reported EQ-5D was moderate to good for VAS and the mobility domain (ICC ≥0.70) (Dorman et al., 1998; Fisk et al., 2005), test-retest reliability was lower in proxy-reported scores (Dorman et al., 1998). The EQ-5D correlated moderately with global measures of function such as the EDSS (r = -0.66) (Fisk et al., 2005), but was less sensitive than disease-specific quality of life scales and the generic SF-36 when used with participants with MS (Nicholl et al., 2001). A single study found a moderate inverse relationship between the EQ-5D and the Nine Hole Peg Test, a specific measure of upper limb use (r = -0.56) (Fisk et al., 2005). When used with participants post-stroke, the EQ-5D correlated with global measures of function including the SF-6D, a classification for describing health from a selection of SF-36 items (r = 0.77) (Barton, Sach, Doherty, et al., 2008) and the SF-36 (r = 0.57 – 0.63) (Dorman et al., 1999). Evidence of the discriminant ability was found between participants post-stroke and those who had not suffered a stroke (Barton, Sach, Doherty, et al., 2008; Xie et al., 2006), between stroke type and severity (Dorman et al., 1997), and between participants with and without spasticity (Gillard et al., 2015). The EQ-5D Index had the greatest change score when compared to like generic HRQoL measures

less than 6 months post-stroke (Pickard et al., 2005), was more responsive to changes in disability (MRS r = -0.36) and daily activities (BI r = 0.57) in comparison to the EQ-5D VAS (Pickard et al., 2005). Contrarily, neither the EQ-5D Index or VAS was responsive to change over a one year period post-stroke despite 23.8% of participants reporting improvement and 23.2% deterioration (Peters et al., 2014). The EQ-5D did not demonstrate either floor and ceiling effects when used with acute participants post-stroke (Pickard et al., 2005).

Results pertaining to sample with upper limb spasticity identified: The EQ-5D index scores were found to correlate with measures of disability ($p<.002$) and carer burden ($p<.05$) (Doan et al., 2012) and to distinguish between participants with and without upper limb spasticity post-stroke, with mean differences (-0.07, 95% CI -0.12 to -0.33) equivalent to the MCID established for the EQ-5D for other health conditions (MCID is yet to be established for post-stroke populations) (Gillard et al., 2015).

4.5.12. Modified Frenchay Arm Test

The modified Frenchay Arm Test (mFAT), reduces the 25 clinical tests to 5 so as to measure arm function after stroke (Heller et al., 1987). Two studies evaluated the psychometric properties of the mFAT (Heller et al., 1987); no studies specifically identified participants with upper limb spasticity.

Content validity: No studies were identified providing information targeting measurement tool development and/or content validity.

Results for whole sample: There was evidence for the reliability of the mFAT (inter-rater (Rho = 0.75 – 0.99), test-retest (Rho = 0.68-0.90 and 0.83-0.99)) when administered to participants 18 months post-stroke (Heller et al., 1987). The mFAT was found to be less sensitive than the NHPT in participants less than 6 months post-stroke with mild impairments (Heller et al., 1987). Floor effects (30%) and ceiling effects (34%) were evident within acute stroke (Heller et al., 1987).

4.5.13. Functional Independence Measure

A total of 20 studies evaluated the psychometric properties, in participants post-stroke (n = 9) (Brown et al., 2015; Egan et al., 2014; Grant et al., 2014; Hamilton & Granger, 1994; Heinemann et al., 1993, 1994; Kuys et al., 2009; Oczkowski & Barreca, 1993; Ouellette et al., 2015), TBI (n = 5) (Corrigan et al., 1997; Cullen, Krakowski, & Taggart, 2014; Cuthbert et al., 2015; Hall et al., 1993; Heinemann et al., 1997), MS (n = 2) (Rabadi & Vincent, 2013; Sharrack, Hughes, Soudain, & Dunn, 1999) and a mixed neurological sample (n=3) (Houlden et al., 2006; Stineman et al., 1996; van der Putten et al., 1999). One study specifically identified participants with upper limb spasticity in a sample with MS (Rabadi & Vincent, 2013).

Content validity: The FIM was found to have sufficient relevance, but indeterminant ratings for comprehensiveness and comprehensibility during development, as nil information was located to determine if participants were interviewed regarding those properties (Keith et al., 1987).

Results pertaining to whole sample: A two factor structure was identified for the FIM by a number of researchers, with separate motor and cognitive domains accounting for 89.4 to 97.9% of variance (Hall et al., 1993; Heinemann et al., 1993, 1994; Sharrack et al., 1999). Evidence for internal consistency has been reported across a number of sample populations (complete FIM $\alpha = 0.94 – 0.98$, FIM motor $\alpha = 0.93 – 0.97$ and FIM cognitive $\alpha = 0.93 – 0.94$ for stroke, MS, traumatic and non-traumatic samples (Sharrack et al., 1999; Stineman et al., 1996)). And between-rater reliability has been demonstrated for both the motor and cognitive domains of the FIM in acute stroke (ICC 0.96, 0.91) respectively (Hamilton & Granger, 1994) and with participants with MS (FIM total inter-rater ICC = 0.99, FIM total intra-rater ICC = 0.94) (Sharrack et al., 1999). Predictive associations between FIM scores and length of stay, discharge destination, minutes of assistance and supervision required on discharge and return to driving were identified (Brown et al., 2015; Corrigan et al., 1997; Cullen et al., 2014; Grant et al., 2014; Heinemann et al., 1997; Oczkowski & Barreca, 1993; Ouellette et al., 2015). When used with participants with MS, FIM was found to be a valid measure of disability (Rabadi & Vincent, 2013), strongly correlating with like global measures (BI r = 0.88), activity

measures (Ambulation Index r = - 0.73) and moderate to strongly with specific activity measures including housework (r = 0.64, p<0.001), work (r = -0.59 p<0.001), independence (r = -0.44, p = 0.001), and disability r = -0.96, p< 0.001) (Sharrack et al., 1999). The FIM total score was at best only moderately responsive to change in a neurorehabilitation sample (ES 0.52 – 0.72), but the FIM cognitive was not (ES = 0.35 – 0.43) (Houlden et al., 2006). In comparison to other measures, the FIM was found to be less responsive than the original BI, equally responsive to BI(C&W) in stroke and more responsive than EDSS in MS, yet still only weak to moderately responsive to change (FIM ES = 0.46, FIM SRM 0.53, EDSS 0.15) (Rabadi & Vincent, 2013; Sharrack et al., 1999; van der Putten et al., 1999). Evidence of floor and ceiling effects for FIM were also found (Brown et al., 2015; Sharrack et al., 1999; van der Putten et al., 1999).

Results pertaining to sample with upper limb spasticity identified: FIM scores correlated with a measures of disability (Kurtkze Expanded Disability Status Scale (EDSS) r_s = -0.69) (Rabadi & Vincent, 2013) and was found to be responsive when capturing change in participants with MS (SRM = 0.53) (Rabadi & Vincent, 2013).

4.5.14. Goal Attainment Scaling

Goal Attainment Scaling (GAS) was first introduced by Kirusek and Sherman (1968) and provides a structured approach to defining and measuring individualized patient centred and/or program based goals. A total of 9 studies evaluated the psychometric properties, in post-stroke (n = 2) (Brock et al., 2009; Turner-Stokes et al., 2010), MS (n = 1) (Khan et al., 2008), TBI (n = 3) (Doig et al., 2010; Joyce et al., 1994; Malec et al., 1991) and mixed ABI (n = 3) samples (Bovend'Eerdt et al., 2011; Lannin, 2003; Malec, 1999). Only one study met inclusion criteria that specifically identified participants with upper limb spasticity (in a sample greater than 6 months post-stroke) (Turner-Stokes et al., 2010).

Content validity: Not assessed, as GAS identifies goal content particular to individual participants and programs (i.e. high face validity).

Results for whole sample: There were conflicting results in inter-rater reliability within a mixed neurological sample, while Joyce, Rockwood and Mate-Kole (1994) report high

reliability (r = 0.92, r = 0.94) between an individual rater familiar with GAS and the treating team, Bovend'Eerdt, Dawes, Izadi and Wade (2011) found a fair level ($ICC_{A,k}$ 0.478) and low agreement (LOA -1.52 ± 25.54) between a therapist and masked assessor. When used with participants with MS, GAS change score correlated weakly with the BI (r_s = -0.25) and FIM (r_s = -0.6) (Khan et al., 2008). In a sample of participants with ABI secondary to trauma and stroke, GAS also correlated strongly with global clinical impressions (r = 0.81) (Khan et al., 2008), weak to strongly with measures of daily activity, participation, disability, vocational outcome and quality of life (r = 0.34 - 0.81) but not with length of stay (Joyce et al., 1994; Malec, 1999; Malec et al., 1991). In the same sample, GAS at 2 months predicted final GAS scores at the completion of a rehabilitation program ranging from 7 to 42 weeks (Malec et al., 1991). Ratings between participants and significant others agreed on 70% of occasions (Doig et al., 2010). GAS was more responsive than the FIM and BI (ES 9.0 SRM: 2.4 t value 10.0 z value 1.4) in MS (Khan et al., 2008) and was responsive to patient centred outcomes and program change in a mixed neurological sample (Lannin, 2003).

Results pertaining to sample with upper limb spasticity: GAS was found to have moderate correlations with self-reported benefit (rho = 0.46, p<.001), low correlations with quality of life (rho = 0.07, p = 0.52), disability (rho = 0.19, p = 0.08), carer burden (rho = 0.14, p = 0.26), measures of pain (rho = 0.03, p = 0.77), mood (rho = 0.06, p = 0.61) and spasticity (rho = 0.35, p = 0.001 (Turner-Stokes et al., 2010).

4.5.15. Medical Outcome Study 36-Item Short-Form Health Survey

The Medical Outcome Study 36-Item Short-Form Health Survey (SF-36) is a global scale assessing eight health concepts (Mackenzie et al., 2002; Ware & Sherbourne, 1992). A total of 24 studies investigated the psychometric properties of the SF-36, 10 included participants with MS (Freeman et al., 2000; Freeman et al., 1996; Herrmann et al., 1996; Moore et al., 2004; Murrell et al., 1999; Pittock et al., 2004; Riazi et al., 2003; Robinson Jr et al., 2009b; Vickrey et al., 1997; Vickrey et al., 1995), 10 post-stroke (Anderson et al., 1996; Dorman et al., 1999; Dorman et al., 1998; Duncan et al., 1997; Hagen et al., 2003; Hobart et al., 2002; Madden et al., 2006; O'Mahony et al., 1998; Salter et al., 2008; Williams et al., 1999), 3 post TBI (Findler et al., 2001; Guilfoyle et al., 2010; Mackenzie et

al., 2002) and 1 discussed tool development with nil specific sample (Ware & Sherbourne, 1992). No studies specifically identified participants with upper limb spasticity.

Content validity: The development of the SF-36 (Ware & Sherbourne, 1992) did not appear to consult participants on the comprehensiveness or comprehensibility of included items (Ware & Sherbourne, 1992). Relevance of items for the intended purpose was sufficient. The SF-36 contains 6 of 9 recommended dimensions for patient-based, health-related quality of life, less comprehensive than the SIS (Salter et al., 2008).

Results for whole sample: The SF-36 was found to have a two-factor structure; with the eight dimensions falling within the two constructs of physical and mental health (Mackenzie et al., 2002). Mixed results were found for the use of the domain scores, with scaling assumptions met in the TBI population (Guilfoyle et al., 2010) but only 6 of 8 scales meeting the scaling assumptions in stroke (Hobart et al., 2002). Evidence for internal consistency of the 8 dimensions, Cronbach alpha >0.70 in majority of studies (Anderson et al., 1996; Dorman et al., 1998; Findler et al., 2001; Freeman et al., 2000; Guilfoyle et al., 2010; Vickrey et al., 1997), however dimensions of vitality and general health did not meet this criteria ($\alpha = 0.68$, $\alpha = 0.66 – 0.68$) (Hagen et al., 2003; Hobart et al., 2002). Test-retest reliability varied; higher for patient reported scores (ICC = 0.30 – 0.81) than proxy reported scores (ICC = 0.25 to 0.76) (Dorman et al., 1998; Murrell et al., 1999; Vickrey et al., 1995). Individual domains of the SF-36 correlated with like subscales of global measures (all $r = \geq 0.50$) post-stroke (EQ-5D) (Dorman et al., 1999) post TBI (Symptom Checklist, Health Problem List, Beck Depression Inventory) (Findler et al., 2001) and with participants with MS (LHS, FIM, general health questionnaire) (Freeman et al., 2000). Correlations, however, were not as strong as hypothesized between individual domains and like dimensions for the BI, CNS and FIM post stroke (Hagen et al., 2003; Madden et al., 2006) nor with the MSFC in a MS population ($r = 0.16 – 0.51$) (Robinson Jr et al., 2009). The SF-36 physical and mental summary scores had weak to moderate correlations with participants rating of severity of symptoms ($r = 0.38$, $r = 0.18$) and quality of life ($r = 0.47$, $r = 0.29$) (Moore et al., 2004; Williams et al., 1999). The ability to discriminate between subgroups of participants with varying levels of function

across post-stroke, TBI and MS populations was demonstrated (Herrmann et al., 1996; Pittock et al., 2004; Riazi et al., 2003; Vickrey et al., 1997; Vickrey et al., 1995). The SF-36 was more responsive in the first three months post-stroke (Hagen et al., 2003) but less responsive in comparison to other tools measuring associated constructs in MS (ES = 0.01 – 0.30) (Freeman et al., 2000). SF-36 did not correlate with FIM change scores, suggesting the change captured within a HRQoL measure was not reflected in a global measure of activity (Madden et al., 2006). There was evidence of significant floor and ceiling effects within MS (Freeman et al., 2000; Freeman et al., 1996) and TBI (Guilfoyle et al., 2010), and varied reports post-stroke (Dorman et al., 1999; Hagen et al., 2003; Hobart et al., 2002; Madden et al., 2006; O'Mahony et al., 1998). The minimal important clinical change varied across dimensions, reported to be 4-9 points within physical functioning, 6-8 within role physical, 6-7 social functioning and 6 points within the physical summary score (Robinson Jr et al., 2009).

4.5.16. Motor Activity Log

The Motor Activity Log (MAL) is a structured interview designed to capture use of the affected upper limb on two scales, Amount of Use (AOU) and Quality of Movement (QOM) (Uswatte et al., 2006). Five studies evaluated the psychometric properties of MAL; all involved participants post-stroke (S. Chen et al., 2012; Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, et al., 2006; Harris & Eng, 2007; Uswatte & Taub, 2005; Uswatte et al., 2006), and one specifically identified participants with upper limb spasticity (Harris & Eng, 2007).

Content validity: The MAL was developed based on the non-use model to capture real-world arm function (Uswatte & Taub, 2005). Item analysis suggests 2 items (put on makeup and write on paper) had greater than 20% missing data, with participants rating as not applicable, and had lower item-total correlations and reliability coefficients (Uswatte et al., 2006).

Results for the whole sample: The self-reported QOM scale correlated with performance based measures (ARAT r = 0.61, WMFT r = 0.65) with the AOU scale correlating less strongly with the WMFT r = 0.40 (Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, et al.,

2006; Uswatte et al., 2006). The minimal detectable change was defined as 16.8% for the AOU and 15.3% for the QOM scales, but the minimal important change was not defined (Chen et al., 2012).

Results pertaining to sample with upper limb spasticity: The MAL correlated strongly with measures of activity (Chedoke Arm and Hand Activity Inventory (CAHAI) $r = 0.82$ $p<0.01$), weakly with measures of participation (Reintegration to Normal Living Index (RNL) $r = 0.23$ $p<0.05$) and of varying strengths (weak to moderate) with impairments, stronger than expected (spasticity $r = -0.71$, strength $r = 0.61$ to 0.84, pain $r = -0.06$, sensation $r = -0.43$, all $p<0.01$) (Harris & Eng, 2007).

4.5.17. Motor Activity Log-28

The Motor Activity Log-28 (MAL-28) is a revision of the MAL-30 with removal of redundant items 'write on paper' and 'put makeup/shaving cream on face' (Uswatte et al., 2006). A single study evaluated the psychometric properties of this measurement tool involving participants greater than 6 months post-stroke, and without any participants with upper limb spasticity (Uswatte et al., 2006).

Content validity: Content analysis indicated appropriate range of items to cover basic (63%) and instrumental (41%) daily activities in addition to items that require finger movement, bimanual and unimanual tasks (Uswatte et al., 2006).

Results for the whole sample: Item analysis indicated that 98% of participants encountered included items in daily life (Uswatte et al., 2006). There was evidence for internal consistency ($\alpha = 0.94 – 0.95$) and increased test-retest reliability with self-ratings rather than proxy (Uswatte et al., 2006). The MAL-28 held convergent validity with real life measure of hand performance and less with overall physical activity, patient ratings stronger than proxy (Uswatte et al., 2006).

4.5.18. Motricity Index

The Motricity Index (MI) is a brief scale of motor recovery (Demeurisse et al., 1980). Six studies evaluated the psychometric properties of MI (Bohannon, 1999; Collin & Wade, 1990; Demeurisse et al., 1980; Jacob-Lloyd et al., 2005; Stone et al., 1993; Wade & Hewer, 1987); all involved participants post-stroke, and none specifically identified participants with upper limb spasticity.

Content validity: Demeurisse et al. (1980) detailed the development of the MI with mixed results regarding its relevance and no evidence supporting either comprehensiveness nor comprehensibility.

Results for whole sample: There was evidence of the internal consistency of this tool (α = 0.97) (Bohannon, 1999) and high inter-rater reliability between an experienced and junior doctor (rho = 0.88) rating 20 participants six weeks post-stroke (Collin & Wade, 1990). The Upper Limb MI (UL MI) correlated strongly with like measures of upper limb activity (RMA arm r = 0.73 – 0.76) (Collin & Wade, 1990) and with global measures of activity (BI r = 0.77) (Wade & Hewer, 1987) whilst correlating moderately with measures of dexterity (NHPT r = 0.36 – 0.56) (Jacob-Lloyd et al., 2005). The UL MI correlated strongly with impairments also, including grip strength (r = 0.74 – 0.94) (Bohannon, 1999). The MI, when combined with the visual neglect recovery index and age at 2-3 days post-stroke was a significant predictor of independence at 3 months (β = 0.042, p<.001) and 6 months (β = 0.038, p<.001) (Stone et al., 1993). Evidence of a ceiling effect was noted, with 18% of the sample scoring the maximum score within the UL component of the MI on discharge from a rehabilitation ward post-stroke (Jacob-Lloyd et al., 2005). There was no evidence of a floor effect.

4.5.19. Nine-Hole Peg Test

The Nine-Hole Peg Test (NHPT) is a timed measure of unilateral upper limb dexterity through the placing and removal of nine pegs in/out of a board (Kellor, Frost, Silberberg, Iversen, & Cummings, 1971). Ten studies evaluated the psychometric properties; 5 post-stroke (Beebe & Lang, 2009b; Heller et al., 1987; Jacob-Lloyd et al., 2005; Morris et al., 2013b) and 5 included participants with MS (Benedict et al., 2011; Costelloe et al., 2008;

Goodkin et al., 1988; Poole et al., 2010; Schwid et al., 2002). One study specifically identified participants with upper limb spasticity (J. A. Beebe & C. E. Lang, 2009b).

Content validity: The NHPT was first discussed as being used in a study in 1985 (Mathiowetz, Weber, Kashman, & Volland, 1985); no information was reported to inform the development nor content validity of the NHPT.

Results for whole sample: The NHPT when used with participants post-stroke correlated with both observed (r = 0.36 – 0.95) (Beebe & Lang, 2009; Goodkin et al., 1988; Heller et al., 1987; Jacob-Lloyd et al., 2005; Poole et al., 2010) and self-reported measures of activity and hand use (r = 0.53 – 0.66) (Beebe & Lang, 2009), was more sensitive than the FAT (Heller et al., 1987), had poor predictive validity in comparison to like measures, and did not predict HRQoL (Morris et al., 2013). The NHPT correlated highly with measures of tremor and dexterity in MS, common activity limitation features (r = -0.62 - -0.87 $p<0.005$) (Alusi, Worthington, Glickman, Findley, & Bain, 2000). There was evidence for the reliability of the NHPT (inter-rater Rho = 0.75 – 0.99 and test-retest Rho = 0.68-0.90 and 0.83-0.99) when administered to participants 18 months post-stroke (Heller et al., 1987). The NHPT was moderate to highly responsive within the first 6 months post-stroke (ES = 0.52-0.66) (Beebe & Lang, 2009; Jacob-Lloyd et al., 2005), was more responsive than the upper limb MI (Jacob-Lloyd et al., 2005) and measures of strength, equally responsive to the ARAT, Jebsen-Taylor test of hand function and less responsive than the SIS-hand (Beebe & Lang, 2009). True change was indicated by a change of 20% when administered to participants with MS (Schwid et al., 2002). There were no floor or ceiling effects found in the MS population.

Results pertaining to sample with upper limb spasticity identified: Strong correlations with measures of hand use, grip and dexterity were reported in stroke populations (r_s = 0.61 – 0.95) and with measures of strength (r_s = 0.61 – 0.82) (Beebe & Lang, 2009) despite the NHPT being a simulated task performance measure. The NHPT was found to be equally responsive as like measures of upper limb activity performance (ARAT and Jebsen-Taylor test of hand function) (ES 0.52 - 0.66), more responsive than measures of

impairment (pinch and grip strength) but less responsive than the SIS-Hand (ES = 0.55 – 1.018) in the first 6 months post-stroke (Beebe & Lang, 2009).

4.5.20. Oxford Handicap Scale

The Oxford Handicap Scale (OHS) is a simple tool modified from the Rankin Scale to grade the ability of a person and the level of daily assistance required to live independently (Bamford, Sandercock, Warlow, & Slattery, 1989). Two studies evaluated the psychometric properties of the OHS, both including participants less than 6 months post-stroke (Rigby et al., 2009; Simon et al., 2008). Neither study specifically identified participants to have upper limb spasticity.

Content validity: No published information regarding the development nor content validity of the OHS was located.

Results for whole sample: The OHS was not a predictor of caregiver burden (Rigby et al., 2009) but was found to predict both the number of services and amount of time required from services on discharge (Simon et al., 2008).

4.5.21. Rivermead Motor Assessment

The Rivermead Motor Assessment (RMA) (Lincoln & Leadbitter, 1979) is comprised of three sections; for this review studies were separated into two categories 1) 'RMA' all three sections (upper limb, trunk and leg) administered and reported and 2) 'RMA UL' upper limb section of the RMA only administered and reported. A total of 7 studies were included (Adams, Pickering, Ashburn, et al., 1997; Adams, Pickering, & Taylor, 1997; Collin & Wade, 1990; Lincoln & Leadbitter, 1979; Morris et al., 2013b; Sackley, 1990), all studies included participants post-stroke, 4 of the 7 studies included participants less than 6 months post-stroke (Adams, Pickering, & Taylor, 1997; Collin & Wade, 1990; Jones, 1998; Sackley, 1990). When separated into the two categories, evidence for the 'complete RMA' was drawn from 5 studies (Adams, Pickering, Ashburn, et al., 1997; Adams, Pickering, & Taylor, 1997; Jones, 1998; Lincoln & Leadbitter, 1979; Sackley, 1990) and evidence for the 'RMA UL' section was drawn from 6 studies (Adams, Pickering,

Ashburn, et al., 1997; Adams, Pickering, & Taylor, 1997; Collin & Wade, 1990; Lincoln & Leadbitter, 1979; Morris et al., 2013; Sackley, 1990).

Content validity: Test authors Lincoln and Leadbitter (1979) detail the measurement tool development. This was completed via selecting a preliminary series of items ranging widely in difficulty ordered into the three sections: gross, leg and trunk and arm. All individual sections were found to have mixed results regarding relevance, reduced due to methods used to create items and nil information regarding comprehensiveness nor comprehensibility.

Results for whole sample: The hierarchical scale of the RMA in an acute and non-acute stroke sample found varying results. Evidence to support the scalability of the RMA was found for the gross function and arm section in acute stroke only (Adams, Pickering, & Taylor, 1997). Scalability was supported in the gross function section only, when used with participants 6 and 12 months post-stroke (Adams, Pickering, Ashburn, et al., 1997; Sackley, 1990). The RMA correlated with ADL performance (r = 0.51) and balance (r = -0.45) (Sackley, 1990), a related construct. Agreement between clinician and participants predicted scores with achieved scores was found (clinician ICC 0.965 Bland Altman 96.6; participants ICC 0.908 Bland Altman 79.3) (Jones, 1998). The hierarchical scale of the RMA UL section was supported only when administered to participants in the acute phase post-stroke (Guttman scaling criteria met) (Adams, Pickering, & Taylor, 1997), the scalability criteria was not met when used with participants 6 and 12 months post-stroke (Adams, Pickering, Ashburn, et al., 1997). The UL section of the RMA was found to correlate strongly with measures of upper limb activity at 6, 12 and 18 weeks post-stroke (r = Rho 0.73 – 0.76) (Collin & Wade, 1990) and greater than six months post-stroke (r = - 0.80) (Morris et al., 2013). The RMA UL correlated moderately with perceived physical activity (r = -0.47) and did not predict overall HRQoL (Morris et al., 2013).

4.5.22. Stroke-Adapted Version of the Sickness Impact Profile

The Stroke-Adapted Version of the Sickness Impact Profile (SA-SIP30) was derived from the original Sickness Impact Profile and contains the following 8 subscales: body care

and movement, mobility, ambulation, social interaction, emotional behavior, alertness behavior, communication and household management (van Straten et al., 1997). Four studies evaluated the psychometric properties of the SA-SIP30 (Doan et al., 2012; Edwards, Hahn, Baum, & Dromerick, 2006; Salter et al., 2008; van Straten et al., 1997), all involved participants post-stroke, and only one study specifically identified participants with upper limb spasticity (Doan et al., 2012).

Content validity: Test authors detailed the methodology applied to create the SA-SIP, based on statistical relevancy and homogeneity (van Straten et al., 1997). The scale was found to be relevant, however to lack comprehensiveness (as only 5 of 9 recommended dimensions for patient-based, health related quality of life measures were included) (Salter et al., 2008). No information regarding comprehensibility was provided.

Results for whole sample: The SA-SIP accounted for 53% of variance in predicting participation (R^2 =0.63, P<0.001) and was more sensitive to detecting stroke related changes impacting on independence at 6 months post-stroke (Edwards et al., 2006).

Results pertaining to sample with upper limb spasticity: The SA-SIP30 was significantly associated with greater disability in hygiene, dressing, limb posture and pain (P<.05) (Doan et al., 2012).

4.5.23. Stroke Impact Scale

The Stroke Impact Scale (SIS) is a stroke-specific measure of global health outcome (Duncan et al., 2003) and comprises of eight domains: strength, hand function, activities of daily living, instrumental activities of daily living, mobility, communication, emotion, memory and thinking, and participation. The SIS was found to be reported as either individual or collective domains which are administered and reported separately. To maintain consistency across all measures within this review, the SIS was required to be administered in full and in the form of version 3 to meet inclusion criteria. Ten studies evaluated the psychometric properties of version 3 of the SIS (Duncan et al., 2003; Duncan et al., 2002; Duncan et al., 2005; Duncan et al., 1999; Eriksson et al., 2013; Jenkinson et al., 2013; Kwon et al., 2006; Lai et al., 2002; Salter et al., 2008; Wolf &

Koster, 2013), all included participants post-stroke and none specifically identified participants with upper limb spasticity.

Content validity: The SIS was originally developed following a comprehensive iterative process with the use of participants, caregivers and standardized instrument development guidelines implemented but specific details are not available (unpublished information) (Duncan et al., 1999). Rasch analysis led to revision of the measure (Duncan et al., 2003) demonstrating comprehensiveness (containing 7 of 9 recommended dimensions for patient-based, health related quality of life) and to be more comprehensive than EQ-5D, SA-SIP and SF-36 (Salter et al., 2008).

Results for whole sample: Rasch analysis refined the SIS into version 3 producing unidimensional domains ranging in item difficulty and with the ability to discriminate (Duncan et al., 2003). A single index was proposed, aggregated from the 8 domains (α = 0.93) accounting for 68.76% of the variance (Jenkinson et al., 2013). These 8 domains were each found to be internally consistent ($\alpha \geq 0.86 - 0.96$) (Duncan et al., 2005; Jenkinson et al., 2013), suggesting possible item redundancy and further investigations of shorter forms. Agreement between patient and proxy ratings were fair to excellent, being stronger in the observable physical domains (ICC 0.50 to 0.83) (Duncan et al., 2002). The tool was reliable between testing sessions when administered via mail (ICC 0.77 - 0.99) and telephone modes (ICC 0.90 - 0.99) (Duncan et al., 2005). The individual and related domains of the SIS were found to correlate with global measures of independence, activity and participation, both patient and proxy reported, ($r = 0.69 – 0.78$) (Duncan et al., 2002; Kwon et al., 2006; Wolf & Koster, 2013). The SIS was able to discriminate between participants deemed recovered by the BI (Lai et al., 2002) and held superior ability to discriminate between varying levels of disability compared to the FIM and SF-36V (modified version of the SF-36) when tools were administered via phone (Kwon et al., 2006). Floor and ceiling effects were varied ranging from nil floor effect and 0 – 32% ceiling effect (Eriksson et al., 2013; Kwon et al., 2006).

4.5.24. Ten Metre Walk Test

The 10 metre walk test (10MWT) is a common measure of gait speed. Tools evaluating gait speed are clinically used to measure the effect of upper limb spasticity on lower limb activity performance as involuntary and/or impaired arm movements may impact on balance and walking ability. This measurement tool was found to vary in administration methods with respect to (1) the length of the track, 6 or 10m, (2) whether acceleration and deceleration was allowed and/or timed and (3) the pace walked (fastest or maximum, self-selected, or comfortable). All variations were included in this review. Eight studies evaluating the psychometric properties of the 10MWT were included (Donovan et al., 2008; Hirsch et al., 2014; Kuys et al., 2009; Miller et al., 2013; Mudge & Stott, 2009; Salbach et al., 2001; Schmid et al., 2012; Scrivener et al., 2014). Seven studies were completed post-stroke (Donovan et al., 2008; Kuys et al., 2009; Miller et al., 2013; Mudge & Stott, 2009; Salbach et al., 2001; Schmid et al., 2012; Scrivener et al., 2014) and one less than 6 months post TBI (Hirsch et al., 2014). No studies specifically identified participants with upper limb spasticity.

Content Validity: Nil specific information was located regarding the development nor content validity of this measurement tool.

Results for whole sample: 10MWT was reliable between retests when assessed using two speeds; self-selected and fastest pace (ICC ranged from 0.946 - 0.979) (Gowland et al., 1993). No relationship was identified between the different gait velocities (self-selected pace or fastest pace) (Hirsch et al., 2014), suggesting that clinically it may not matter which velocity is chosen. The 10MWT correlated weakly with the ICF Measure of Participation and Activities (IMPACT) (participation r = -0.21, activity r = -0.31, P<.05) post-stroke (Schmid et al., 2012), but did not correlate with impairment ratings (fatigue r = -0.18, P .128 and pain r = 0.04, P = .706) (Miller et al., 2013). 10MWT was not able to predict ambulatory activity in natural environments (Mudge & Stott, 2009). The 10MWT was found to be a responsive measurement tool (maximum speed ES 0.55 - 1.44, SRM 0.83 -0.93, mES 0.45; comfortable speed ES = 0.74 SRM = 0.92), however less responsive than the 5MWT, Berg Balance, BI and Motor Assessment Scale (items 3-5) in acute stroke (Salbach et al., 2001; Scrivener et al., 2014). There was evidence of a large floor

effect (66.8%) on admission and on discharge (25.7%) within an acute stroke sample but no evidence for a ceiling effect on admission or discharge (Scrivener et al., 2014).

4.5.25. Upper-Limb Motor Assessment Scale

The Upper Limb - Motor Assessment Scale (UL-MAS) is a subscale of items 6, 7 and 8 of the Motor Assessment Scale, and it provides a task orientated performance-based measure of upper limb activity (Carr et al., 1985). Ten studies evaluating the psychometric properties of the UL-MAS were included (Carr et al., 1985; Johnson & Selfe, 2004; Khan et al., 2013; Kuys et al., 2009; Lannin, 2004; Loewen & Anderson, 1988; Loewen & Anderson, 1990; Miller et al., 2010; Pickering et al., 2010; Sabari et al., 2005), all involved participants less than 6 months post-stroke, and no studies specifically identified participants with upper limb spasticity.

Content validity: Evidence located for the development of the MAS and subsequent UL-MAS did not indicate participants were consulted on the comprehensiveness or comprehensibility of included items (Carr et al., 1985). Relevance of items for the intended purpose was sufficient.

Results for whole sample: There was evidence to support the production of a single composite score from the UL-MAS items, which may be interpreted as a total score for UL function (Lannin, 2004). Inconsistencies were identified within the hierarchical scoring (Miller et al., 2010; Pickering et al., 2010; Sabari et al., 2005) with clinical recommendations to attempt and score every item (Miller et al., 2010). Furthermore, task 2 within the Hand Movements item may not be indicative of upper limb motor recovery in adults aged 65 years and older (Miller et al., 2010). The UL-MAS is a unidimensional scale measuring a single construct, upper limb motor performance, (α = 0.83 to 0.95, and with removal of wrist deviation 0.93) (Johnson & Selfe, 2004; Lannin, 2004; Miller et al., 2010). It was reliable between (Kendall Tau = 0.74 – 1.00) and amongst assessors (kappa 0.93 – 1.0, 88 – 85 % agreement) (Carr et al., 1985; Loewen & Anderson, 1988). The UL-MAS was able to discriminate between differing levels of motor recovery both in the acute and subacute phase, with Rasch based scoring more precise (Khan et al., 2013). Varying levels of floor and ceiling effects have been reported for the

UL-MAS (floor effect 0 – 38%, ceiling effect 0 – 67%) (Miller et al., 2010; Pickering et al., 2010; Sabari et al., 2005).

4.6. Discussion

This systematic review located, appraised and synthesised published literature investigating the psychometric properties of measurement tools which assess upper limb function in the context of everyday activities. Across the included 30 measurement tools, there was wide variability in the quality of evidence in relation to participants with neurological conditions, but overall, tools with the greatest number of psychometric publications demonstrated the strongest evidence. While the FIM™ had the highest quality evidence supporting its validity and reliability, it suffered from both floor and ceiling effects. On consideration of specific constructs measured by the tools, wide variability across quality of evidence remained. Both patient-reported measures, the ArmA and DAS, and performance-based measures, the UL-MAS and ARAT, demonstrated evidence within the measures specifically targeting upper limb activity. Evidence supported use regardless of whether upper limb spasticity was present or not, except for the UL-MAS, which is replaced with the MAL for patients with identified upper limb spasticity. Despite the BI and BI(C&W) holding high to moderate levels of evidence for construct validity, the FIM held the strongest level of evidence for global measures of activity, regardless of whether or not upper limb spasticity was present. The SIS, a patient-reported measure, held the strongest level of evidence across a greater number of properties and demonstrated higher correlations with measures of upper limb performance and activity of the global health-related quality of life measures. The EQ-5D and SA-SIP were the only health-related quality of life measures with evidence supporting construct validity for participants with upper limb spasticity. In light of mixed findings without a clearly superior measurement tool, findings highlights the need for further research into the psychometric properties of measurement tools which capture upper limb activity and/or participation performance.

The search yielded psychometric studies primarily conducted between 2000 and 2010, with an even split of additional evidence located in the 10 years either side of that decade. It was interesting that few papers have been published in the more recent years

– this may reflect publication preferences of journals in rehabilitation, or a potential assumption by clinicians that the psychometric properties have been well established, or that the prevalence of spasticity is not as common as clinically thought nor impacting upper limb use as significantly as recent attention suggests (Ada, O'Dwyer, & O'Neill, 2006). Most studies were completed with participants post-stroke in the acute to subacute phase, and as such, findings from these studies may not apply to a more chronic population or a group of neurological patients who have not suffered a stroke. Individual study sample sizes were commonly small (less than n=100 in over half (57%) of studies), which is a common limitation highlighted by other reviews of functional measurement tools (Dobson et al., 2012; Wales, Clemson, Lannin, & Cameron, 2016). This finding strengthens earlier calls for continued investment in appropriately powered psychometric studies, inclusion of psychometric evaluation in both routine data collection and longitudinal studies, and a need for scientific journals or outcome tool publishers to publish such research.

The construct validity and responsiveness, followed by reliability properties of measurement tools, were most commonly evaluated across the different tools, but rarely was content validity or measurement error tested. The methodological quality of included studies was wide ranging, from 'inadequate' to 'very good', suggesting that making decisions between measures may be difficult, since there was little consistent data to guide decisions. Detailed data was often lacking within studies such as those reporting on the reliability of tools where information failed to describe testing conditions, stability of patients between sessions and evidence for systematic change occurrence. The COSMIN process recommends that an 'a priori' hypothesis be developed when evaluating construct validity and responsiveness, however in our review only a very small number of studies clearly defined hypotheses about the expected results. The majority of studies were found to report generic hypotheses, where hypotheses were assigned based on interpretations by the authors. Furthermore, the quality of statistical approaches used were low, for example often reporting on statistical significance of findings rather than expected strengths and direction of correlations. Consistent with Zaki, Bulgiba, Nordin and Ismail (2013), our review also suggests that the quality of research in psychometrics is unlikely to improve without

education and clear guidelines on analysis. The COSMIN checklist may provide such guidance; the COSMIN process separates the statistical methods based on Classical Test Theory (CTT) or on Item Response Theory (IRT) and an understanding of these methods is likely key to improving the psychometrics of scales where multiple items contribute to an overall score.

The review identified very limited evidence useful for the clinical selection of a single tool to evaluate upper limb activity when upper limb spasticity is present. Inadequate representation of the intended population within the sample of a psychometric study can lead to erroneous assumptions about the psychometrics of a tool (Sikorskii & Noble, 2013). In the context of instrument development, internal and external validity are important for application of an instrument in assessing new target populations (in this case, adults with upper limb spasticity). The DAS, EQ-5D, FIM™, NHPT and SA-SIP had evidence supporting both internal and external validity and responsiveness, however no single measurement tool had identified psychometric evidence for all properties in a sample of participants with upper limb spasticity. This gap in available research is acknowledged, and is both a limitation to this systematic review and a recommendation for further research. The evidence located to guide selection for the broader neurorehabilitation sample was larger in comparison primarily due to additional numbers of contributing studies. However, despite large numbers of contributing studies, we could still not conclude that any of the identified measurement tools from the Ashford and Turner-Stokes (2013) review have published psychometric evidence for all relevant psychometric properties.

In this review, despite selecting the most recent and comprehensive set of tools at the time of registering our protocol, we acknowledge a potential limitation in range of tools included and that other existing tools had not been used in clinical trials or cohort studies of patients with spasticity, and therefore were not synthesized in the Ashford and Turner-Stokes (2013) review. Emerging technologies such as kinematic assessments and accelerometers were not included in this review and may provide future developments in this space. The limited psychometric testing of the tools that were included was a further limitation, making it difficult to compare the psychometric

properties of tools across different pathologies. This may mean that the preferred assessments of a reader may not appear in this extensive review, and where included, it may have only been tested in a single diagnostic population. Only one additional measurement tool beyond the initial systematic review was recommended in the recent national guidelines (Royal College of Physicians et al., 2018), that tool being the Arm Activity Measure (ArmA). Psychometric studies not published in English were also excluded for pragmatic reasons; formal translations have not yet occurred in many of the measurement tools (e.g. ARAT and UL-MAS) and therefore studies conducted in languages other than English were excluded as per COSMIN guidelines.

4.7. Conclusions

This systematic review of included measurement tools provides a comprehensive systematic synthesis of evidence for the psychometric properties of the upper limb measurement tools used to evaluate the dimensions of activity and/or participation. The findings may provide guidance for clinicians on evidence-based measurement tool selection. Together, 30 measurement tools met the inclusion criteria and of these, 8 demonstrated at least a moderate level of confidence in the measurement property estimate in two or more standards. While no reviewed tool had at least moderate estimates for all standards (i.e. content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for construct validity and responsiveness), and thus, could not be recommended as the gold-standard for assessment of upper extremity function and activity in research and clinical practice, the review was able to suggest which measurement tools should continue to be researched and refined for use. Future research needs to investigate the psychometric properties of these measurement tools, in each neurological population as well as with a subsample with spasticity in the upper limb.

4.8. Chapter Synopsis

This Chapter provided an in-depth, systematic synthesis of available psychometric evidence for tools available to measure upper limb activity outcomes. One particular outcome measurement tool, the Action Research Arm Test (ARAT) which is

recommended in recent clinical guidelines (Kwakkel et al., 2017) and is commonly used within neurorehabilitation, will now be extracted and discussed in greater detail. This demonstrated the type of detailed examination that could and should occur for all outcome measurement tool presented in Chapter 4 in the absence of a clear gold standard.

Chapter 5:

Psychometric Properties and Clinical Utility of the Action Research Arm Test

This study has been published as:

Pike, S., Lannin, N. A., Wales, K., & Cusick, A. (2018). A systematic review of the psychometric properties of the Action Research Arm Test in neurorehabilitation. *Australian Occupational Therapy Journal, 65*(5), 449-471. doi:10.1111/1440-1630.12527

Trial Registration: Prospective Register of Systematic Reviews (PROSPERO) registration number CRD42014013190.

See Appendix A for publication permission, Appendix C for trial registration, Appendix F for supplementary material, and Appendix I for published manuscript.

5.0. Chapter Overview

The previous chapter systematically identified, appraised and synthesized the available evidence for the psychometric properties of 30 outcome measurement tools (Ashford & Turner-Stokes, 2013) using the revised COSMIN methodology (Mokkink et al., 2018; Terwee et al., 2018). The Action Research Arm Test (ARAT) was one of the appraised tools. The ARAT is a standardised outcome measurement tool commonly used both nationally and internationally within clinical practice and research to measure upper limb activity. Despite the popularity of use, existing evidence regarding the psychometric properties of the ARAT has not been examined using consensus standards. This tool was therefore selected for extraction and in-depth review to fill this information gap.

5.1. Background

Neurorehabilitation outcome measurement is complicated by highly variable clinical presentations, along with diverse individualised and person-centred intervention goals. Clinicians seek outcome measurement tools that can accommodate the diverse and individualised nature of neurorehabilitation at the same time as providing meaningful, sensitive and reliable data on which to base decisions and plans. Clinicians working in neurorehabilitation are particularly interested in the effect of interventions on attainment of 'real life' activity and participation goals. Outcome measures thus need to capture the complexity of factors contributing to performance of everyday tasks and participation in 'real life' contexts.

The Action Research Arm Test (ARAT) (Lyle, 1981) is a standardised observational performance measure that evaluates a person's ability to use their upper limb to handle objects using grasp, grip, pinch and gross motor movements. These movements are needed to perform many everyday tasks. For this reason, an inability to perform test items is thus proposed to be a valid indicator of upper limb activity limitation (Kwakkel et al., 2017). The ARAT has been demonstrated to be unidimensional, measuring the single construct of upper limb function related to everyday activities (Koh et al., 2006; Van der Lee, Roorda, Beckerman, Lankhorst, & Bouter, 2002). Furthermore, it has been found to hold concurrent validity with other tests of activity limitation, including the Wolf Motor Function Test (WMFT), Motor Activity Log (MAL), Stroke Impact Scale (SIS)

hand function items (Lin et al., 2009; Lin, Chuang, Wu, Hsieh, & Chang, 2010). To date, no review has used consensus standards to examine existing evidence regarding psychometric properties of the ARAT when used with adults who have neurological conditions and are undergoing neurorehabilitation and experience spasticity. This study aims to fill this information gap.

5.1.1. Development of the ARAT

The ARAT was developed by Ronald Lyle in 1981 and is based on Carroll's Upper Extremity Function Test (UEFT) (Carroll, 1965). Theoretically, the UEFT assumed that "upper extremity movements used in daily activities can be reduced to certain patterns", and that observation of these patterned movements could provide information to monitor upper extremity function related to everyday activities (Carroll, 1965, p. 479). The patterns in the UEFT were grouped as grip; lateral prehension; pinch; placing; pronation and supination; and writing. In the ARAT, Lyle sought to adapt the UEFT to shorten administration from 60 minutes. Lyle questioned Carroll's grouping of items into five patterns and performed correlational analysis, item reduction and hierarchical scoring order to produce what became known as the ARAT – a "rapid yet reliable" measurement tool to measure "changes in upper limb function" (Lyle, 1981, p.483).

5.1.2. Description of the ARAT

There are a number of adaptations of the original (1981) ARAT. This study of psychometric properties examines only the original version, hereby called "original ARAT" in the remainder of the paper. This is a 19-item performance test where participants are asked to complete movements and handle objects with each upper limb, starting with the less affected limb. Item performance is not timed. The items are organised into four subscales:

- Grasp (of differing sized blocks, a ball and stone);
- Grip (pouring water between glasses, moving differing diameter tubes placed vertically, placing washer over bolt);
- Pinch (of various sized marbles and ball bearing between thumb and individual digit combinations and place on shelf); and

- Gross movement (of hand behind and on top of head and to mouth without objects).

The test is standardised through the description of the size and nature of each object to be handled and the action to be performed. The test can be administered anywhere by anyone and no test certification is required. The test kit can be purchased through the test site (http://www.aratest.eu/) or self-assembled using guidelines.

5.1.3. ARAT Scoring

Original ARAT items are scored on observation of movement performance using an ordinal four-level scale ranging from 0 to 3. Each limb is scored separately. A score of 0 indicates that no part of the test can be performed; 1 indicates partial test performance; 2 indicates with abnormally long time or great difficulty; 3 indicates normal item performance. Possible total scores range from 0 to 57.

5.1.4. ARAT Structure

Items in each subtest are ordered hierarchically, with the most difficult item presented first. If the highest score is obtained for the first item (score of 3), it is inferred that all items less difficult in the subscale could be completed and the person is not required to attempt remaining items and moves to the next subtest. If <3 is obtained in a subtest, the second and easiest item is attempted. If 0 is scored, the person is deemed to score 0 on all other subtest items. If 1 or 2 is scored, they must attempt all remaining subtest items. This hierarchical nature of the original ARAT speeds administration for people with high or low performance in subtests.

5.1.5. ARAT Implementation

The original ARAT provided clinicians and researchers with instructions that were limited in detail giving few specifications in what to observe in administration and how to score in a consistent and standardised manner. This resulted in variable in administration, scoring and interpretation of results across clinicians, researchers and sites. Subsequent guidelines and manuals included additional operational definitions to increase standardisation in administration and scoring (e.g. defining 'abnormally long' into

specific time frames) (Hsueh, Lee, & Hsieh, 2002; Platz, Pinkowski, van Wijck, & Johnson, 2005; Platz, Pinkowski, van Wijck, Kim, et al., 2005; Wagenaar et al., 1990; Yozbatiran et al., 2008). None of these post-1981 ARAT versions have been identified as gold-standard; thus, only studies using the original ARAT are included in this review.

5.1.6. ARAT Uptake

The original ARAT and adaptations are commonly cited as a primary outcome measure within clinical trials and intervention studies (Santisteban et al., 2016). Apart from its popularity, the robust and sensitive nature of the ARAT are indicated by its use as a criterion measure in validation studies for new, existing and/or modified assessments (Barreca, Stratford, Masters, Lambert, & Griffiths, 2006; Blennerhassett, Avery, & Carey, 2010; Edwards, Lang, Wagner, Birkenmeier, & Dromerick, 2012; Page, Hade, & Persch, 2015). The ARAT has also been included in clinical practice guidelines and consensus statements relevant to neurorehabilitation, stroke (Kwakkel et al., 2017; Sullivan et al., 2013) and upper limb spasticity (Sheean et al., 2010).

Internationally, the ARAT has widespread practice use, including translated versions, however, like many other outcome measures, uptake as measured by published intervention studies particular to stroke using this tool varies from country to country (Santisteban et al., 2016). Its published use is most common in versions presented in English, with the United Kingdom and Australia being highest (Santisteban et al., 2016). When the ARAT is used in translation, these versions lack validation and translation technique information; thus, only studies using the original ARAT citing Lyle (1981) and in the English language are included in this review.

5.1.7. ARAT Sensitivity in Practice

An important milestone in research and clinical application of the original ARAT in neurorehabilitation came with the identification of a minimum clinically important change (MCD), although it was specific only to acute and chronic post-stroke samples (Lang et al., 2006; Van der Lee et al., 2001). The score was 5.7 points or 10% of the total score. The threshold score gave a sound base to inform research and practice decisions regarding interpretation of intervention impacts.

5.1.8. Psychometric Properties of the ARAT

Psychometric properties of the original ARAT have been evaluated in studies with patient's post-stroke. Many of these studies are included in results of this review and are thus not cited here. Studies excluded on the basis of language but relevant to this introduction show that the original ARAT is unidimensional, measuring the single construct of upper limb function related to everyday activities (Koh et al., 2006; Van der Lee et al., 2002). Furthermore, it has been demonstrated to hold predictive validity and concurrent validity with similar tests of activity limitation, including the WMFT, MAL and SIS hand function (Lin et al., 2009; Lin et al., 2010). The ARAT has been demonstrated to be a reliable and responsive measure (Chen, Lin, Wu, & Chen, 2012; Hsueh & Hsieh, 2002; Rabadi & Rabadi, 2006).

While a handful of assessment systematic reviews with various foci have included the ARAT, no systematic review has yet synthesised the psychometric properties of any version of the ARAT or the original ARAT specifically using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) approach to evaluate the methodological quality of included studies or their conformity with consensus-based measurement standards. There has not been a synthesis of the evidence of psychometric properties of the original ARAT in neurorehabilitation. Furthermore, despite the high incidence of patients in neurorehabilitation with upper limb spasticity, there is limited guidance on the use of original ARAT with this clinical population.

5.2. Study Aim

This review will identify and synthesise published evidence regarding psychometric properties of the original ARAT when used with adults who have a neurological condition and are undergoing neurorehabilitation and experience spasticity.

5.3. Research Questions

1. What are the psychometric properties; internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing (construct validity), cross-cultural validity, responsiveness and interpretability

of the original ARAT when used with adults with a neurological condition undergoing neurorehabilitation?

2. Are psychometric properties of the original ARAT different when the presence of upper limb spasticity is reported in the study samples?

5.4. Method

This systematic review applied the COSMIN methodological approach (Mokkink et al., 2010) supplemented by a quality appraisal proposed by Terwee (Terwee et al., 2017; Terwee et al., 2007).

5.4.1. Identification and Selection of Studies

To identify relevant articles, searches were conducted of the following from inception until December 2017: Medical Literature Analysis and Retrieval System Online (MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL) and Excerpta Medica dataBASE (EMBASE). The search strategy is reported in a larger evaluation of upper limb function measurements (Pike et al., 2015) and presented in Appendix F. Studies were included: if the original version of the ARAT was used with no modifications, in full, with all items administered and data reported; if the study was conducted and reported in English; if any of the psychometric properties defined by COSMIN were investigated; if it was an original study that collected data; if the original ARAT was either the primary outcome measure or was used in such a way that its psychometric properties were still evaluated and reported; and if reports were in peer-reviewed literature. Studies needed to have participants who were adults (>18 years), who were undergoing neurorehabilitation with a study sample where there was no less than 90% of participants with a neurological condition diagnosis of stroke, multiple sclerosis, cerebral palsy, traumatic brain injury or anoxia.

Studies reviewed in the subset which included upper limb spasticity were identified by explicit documentation of the presence of upper limb spasticity in participants whose data were reported. For example, Page et al. (2012) reported: (a) a Modified Ashworth Scale score ≥3 as an exclusion criterion and (b) nil report of participants with spasticity

≤3 in the study. Page et al. (2012) were thus deemed a study with nil upper limb spasticity participants present.

The thesis author screened all titles and abstracts; potential exclusions were examined by one of the two thesis authors supervisor and those with agreement were excluded. Others required inspection of the full text by two of three investigators and a consensus decision on inclusion or exclusion was made. Full text was obtained for all included papers, and following full text inspection, final exclusion decisions were made by consensus.

5.4.2. Data Collection

An author developed data extraction form in Excel™ was used to record information. The thesis author examined full text and entered data into the form, referring uncertain aspects for consensus decision by two and/or three of the investigators, whereupon the data were entered into the form. Aspects recorded were study design; participants; description of neurorehabilitation programmes; outcome measures used; the ICF classification; psychometric properties; and inclusion/exclusion decisions.

5.4.3. Data Analysis

The quality of included papers was evaluated using the COSMIN checklist with 4-point scale. This checklist was applied to determine whether each study met the standards for methodological quality with regard to internal consistency, reliability (test–retest, inter-rater and intra-rater), measurement error, content validity (including face validity), structural validity, hypothesis testing, cross-cultural validity, criterion validity, responsiveness, interpretability and generalizability (Pike et al., 2015; Terwee et al., 2012). A rating of 'excellent', 'good', 'fair' or 'poor' was assigned for each measurement property.

Terwee's quality criteria (Terwee et al., 2017; Terwee et al., 2007) were applied to individual studies to analyse the measurement properties of the original ARAT. The study design, methods and outcomes, for content validity, internal consistency, construct validity, structural validity, Item Response Theory/Rasch analysis, reliability,

responsiveness, measurement error, floor and ceiling effects and interpretability were evaluated. Criterion validity was considered in relation to the measurement of upper limb activity performance – the decision was made not to evaluate this as no agreed gold standard measure of upper limb activity performance exists. Measurement properties were rated as positive '+', indeterminant '?', negative '–' or no information '0'.

The sample size of individual studies was not assessed within the COSMIN data extraction and rating phase. Rather sample size was considered in the evidence synthesis stage where sample sizes from included individual studies were combined – in line with Dobson et al. (2012) approach.

A best evidence synthesis was completed for each psychometric property based on: methodological quality of reporting studies (COSMIN); rating and consistency of the rating assigned for measurement properties meeting the quality criteria (the 'Terwee' criteria); and overall sample size (evidence was assigned 'strong' when total sample size of combined eligible studies was ≥100, 'moderate' with total samples between 50 and 99, 'limited' with total samples between 25 and 49 and 'unknown' with total sample less than 25 (Dobson et al., 2012). The synthesis of best evidence approach was based on that applied by Wales et al. (2016) and Dobson et al. (2012). and adapted from Terwee et al. (2007). Studies with poor methodological quality were *not* included in the best evidence synthesis.

5.5. Results

The search strategy identified 711 studies (excluding duplicates). After screening titles, abstracts and full text, 28 of these 711 (4%) were deemed eligible and included for appraisal. Figure 5.1 presents the flow of papers through the review. The included studies are detailed in Table 5.1. A summary of study results is detailed in Table 5.2; the synthesis of best evidence for psychometric properties is within Table 5.3; and COSMIN and Terwee ratings are outlined in Appendix F.

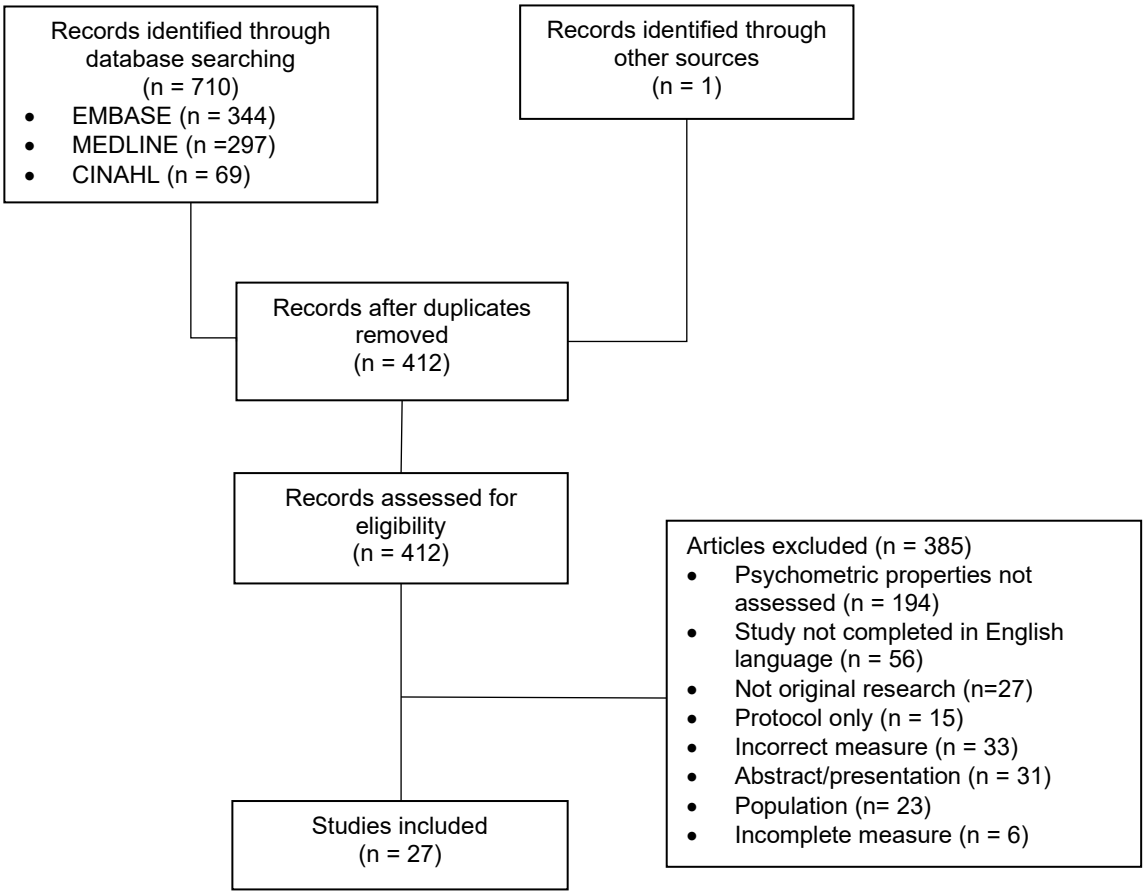


Figure 5.1. *Study inclusion - exclusion process*

5.5.1. Study Participants

Twenty-five studies included only post-stroke participants and three were mixed samples, including post- stroke and traumatic brain injury (TBI). There was a total of 1005 participants, 985 with stroke, 15 TBI and 20 participants who could not be differentiated as either stroke or TBI.

Chronicity post brain injury was extracted from studies and assigned >6 or <6 months post brain injury. The split was relatively even, with 46% of the studies, including participants >6 months post their initial brain injury. This percentage included all studies with both mixed sample and TBI diagnoses. Six of the included studies specifically identified 199 participants with upper limb spasticity; MAS scores ranged from 1 to 3 (182 post-stroke; 15 TBI).

Table 5.1. *Characteristics of included studies*

Studies included	Summary of study participants	Psychometric property tested	
Barden, Baguley, Nott, & Chapparo (2014)	Diagnosis = Mixed (Stroke n = 22, TBI n = 6) Time since diagnosis *(mo)* = greater than 6 n = 28 Age *(yr)*, mean (SD) = 51 (17) Sex, number male (%) = 15 (54) Sample included people with spasticity = yes		Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Barden, Nott, Heard, Chapparo, & Baguley (2012)	Diagnosis = Mixed (Stroke n = 29, TBI n = 9) Time since diagnosis *(mo)* = greater than 6 n = 38 Age *(yr)*, median (range) = 50 (18 - 81) Sex, number male (%) = 22 (58) Sample included people with spasticity = yes	✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Barreca, Stratford, Lambert, Masters, & Streiner (2005)	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 39 Age *(yr),* mean (SD) = acute grp 71.4 (50.9 - 90.0) chronic grp 64.0 (44.7 - 76.6) Sex, number male (%) = 20 (51) Sample included people with spasticity = not reported	✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Barreca, Stratford, Masters, Lambert, & Griffiths (2006)	Diagnosis = Stroke Time since diagnosis (*mo*) = less than 6 n = 105 Age *(yr),* quartiles = mild-mod impairment 66, 76, 81, severe impairment 59, 69, 77 Sex, number male (%) = 54 (51) Sample included people with spasticity = not reported	✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability

Studies included	Summary of study participants		Psychometric property tested
Barreca, Stratford, Masters, Lambert, Griffiths, et al. (2006)	Diagnosis = Stroke Time since diagnosis *(mo)* = mixed (62% less than 6) n = 39 Age *(yr),* median (1st, 3rd quartiles) = acute grp 71 (51, 90) chronic grp 64 (45, 77) Sex, number male (%) = 20 (51) Sample included people with spasticity = not reported	✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Beebe & Lang (2009)	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 33 Age *(yr),* mean (SD) = 53.9 (10.2) Sex, number male (%) = 19 (58) Sample included people with spasticity = yes	✓ ✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Blennerhassett et al. (2010)	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 22 Age (*yr*), median (IQR), [range] = 63 (50 - 69), [23 - 80] Sex, number male (%) = 17 (77) Sample included people with spasticity = not reported	✓ ✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Burridge et al. (2009)	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 17 Age *(yr),* mean (SD) = 57 (13.4) Sex, number male (%) = 11 (65) Sample included people with spasticity = yes	✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability

Studies included	Summary of study participants		Psychometric property tested
Celik et al. (2010)	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 9 Age *(yr)*, range = 48 – 67 Sex, number male (%) = 7 (78) Sample included people with spasticity = not reported	 ✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
De Weerdt & Harrison (1985)	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 53 Age *(yr),* mean (SD) = 68.6 (9.3) Sex, number male (%) = 25 (47) Sample included people with spasticity = not reported	 ✓ ✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, et al. (2006)	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 39 Age *(yr),* mean (SD) = 64.54 (14.13) Sex, number male (%) = 17 (44) Sample included people with spasticity = not reported	 ✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Edwards et al. (2012)	Diagnosis = Stroke Time since diagnosis *(mo)* = less than 6 n = 40 Age *(yr),* mean (SD) = 63.7 (13.6) Sex, number male (%) = 21 (42) Sample included people with spasticity = not reported	 ✓ ✓ ✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability

Studies included	Summary of study participants		Psychometric property tested
Fleming et al. (2014)	Diagnosis = Stroke Time since diagnosis *(mo)* = greater than 6 n = 33 Age *(yr)*, mean (SD) = 61.5 (14.2) Sex, number male (%) = 20 (61) Sample included people with spasticity = yes	 ✓ ✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Lang et al. (2008)	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 12 Age *(yr),* mean (SD) = 64 (14) Sex, number male (%) = 21 (40) Sample included people with spasticity = not reported	 ✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Lang et al. (2006)	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 50 Age *(yr),* mean (SD) = 63.7 (13.6) Sex, number male (%) = 21 (42) Sample included people with spasticity = yes	 ✓ ✓ 	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Lyle (1981)	Diagnosis = mixed (Stroke n = unknown, TBI n = unknown) Time since diagnosis *(mo*) = Mixed (mean greater than 6) n = 20 Age *(yr),* mean (range) = 53.2 (26 - 72) Sex, number male (%) = 13 (65) Sample included people with spasticity = not reported	 ✓ 	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability

Studies included	Summary of study participants		Psychometric property tested
McDonnell, Hillier, Ridding, & Miles (2006)	Diagnosis = Stroke Time since diagnosis *(mo*) = less than 6 n = 17 Age *(yr),* range = 45 - 94 Sex, number male (%) = 9 (53) Sample included people with spasticity = not reported Rater characteristics Nil provided		Internal consistency ✓ Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross cultural validity Responsiveness Interpretability
Morris et al. (2013)	Diagnosis = Stroke Time since diagnosis *(mo*) = greater than 6 n = 85 Age *(yr),* median (range) = 69 (36 - 88) Sex, number male (%) = 49 (58) Sample included people with spasticity = not reported		Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross cultural validity Responsiveness Interpretability
Notley, Turk, Pickering, Simpson, & Burridge (2007)	Diagnosis = Stroke Time since diagnosis (*mo*) = greater than 6 n = 10 Age *(yr),* mean (SD) = 63 (13.8) Sex, number male (%) = 6 (60) Sample included people with spasticity = not reported		Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross cultural validity Responsiveness Interpretability
O'Dell et al. (2013)	Diagnosis = Stroke Time since diagnosis *(mo*) = greater than 6 n = 32 Age *(yr),* mean (SD) = 56 (12.4), Sex, number male (%) = 23 (72) Sample included people with spasticity = not reported		Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross cultural validity ✓ Responsiveness Interpretability

Studies included	Summary of study participants		Psychometric property tested
Page et al. (2015)	Diagnosis = Stroke		Internal consistency
	Time since diagnosis *(mo*) = greater than 6	✓	Reliability
	n = 32		Measurement error
	Age *(yr),* mean (SD) = 56.6 (10.1)		Content validity
	Sex, number male (%) = 15 (47)		Structural validity
	Sample included people with spasticity = not reported	✓	Hypothesis testing
	Rater characteristics		Cross cultural validity
	Rater n =1 Clinical experience *(yr)* = 8		Responsiveness
	Observations n = 64		Interpretability
Page, Levine, & Hade (2012)	Diagnosis = Stroke		Internal consistency
	Time since diagnosis *(mo*) = greater than 6	✓	Reliability
	n = 29	✓	Measurement error
	Age *(yr),* mean (SD) = 60.8 (12.3)		Content validity
	Sex, number male (%) = 23 (79)		Structural validity
	Sample included people with spasticity = not reported	✓	Hypothesis testing
	Rater characteristics		Cross cultural validity
	Rater n =1 Clinical experience *(yr)* = 8		Responsiveness
	Observations n = 58		Interpretability
Rabadi & Rabadi (2006)	Diagnosis = Stroke		Internal consistency
	Time since diagnosis *(mo)* = less than 6		Reliability
	n = 104		Measurement error
	Age *(yr),* mean (SD) = 72.0 (13)		Content validity
	Sex, number male (%) = 43 (41)		Structural validity
	Sample included people with spasticity = not reported	✓	Hypothesis testing
			Cross cultural validity
		✓	Responsiveness
			Interpretability
Rand & Eng (2015)	Diagnosis = Stroke		Internal consistency
	Time since diagnosis *(mo)* = less than 6		Reliability
	n = 32		Measurement error
	Age *(yr),* mean (SD) = 58.1 (12.4)		Content validity
	Sex, number male (%) = 25 (78)		Structural validity
	Sample included people with spasticity = not reported	✓	Hypothesis testing
			Cross cultural validity
			Responsiveness
			Interpretability

Studies included	Summary of study participants		Psychometric property tested
Stinear, Barber, Petoe, Anwar, & Byblow (2012)	Diagnosis = Stroke Time since diagnosis (*mo)* = less than 6 n = 40 Age *(yr),* median (range) = 70 (31 - 91) Sex, number male (%) = 16 (40) Sample included people with spasticity = not reported	✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Urbin, Bailey, & Lang (2015)	Diagnosis = Stroke Time since diagnosis (*mo)* = Mixed (77% greater than 6) n = 35 Age (yr), mean (SD) = 56 (10.4), 62 (9.4) Sex number male (%) = 6 (75), 20 (74) Sample included people with spasticity = not reported	✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability
Yozbatiran et al. (2008)	Diagnosis = Stroke Time since diagnosis (*mo)* = greater than 6 n = 12 (validity) n = 9 (interrater reliability) n = 8 (intra rater) Age *(yr),* mean (SD) = 61.0 (15.0) Sex, number male (%) = 6 (50) Sample included people with spasticity = not reported Rater characteristics Rater n =2 Clinical experience *(yr)* = 8 Observations n = 58	✓ ✓	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross cultural validity Responsiveness Interpretability

5.5.2. Measurement Properties and Synthesis of Best Evidence

Information regarding original ARAT measurement properties was extracted and appraised. This included reliability, measurement error, validity (both structural and construct), responsiveness, interpretability and floor–ceiling effects. Interpretability was examined according to COSMIN guidelines whereby data are extracted only if the study explicitly aimed to assess interpretability through floor and ceiling effects, minimal important change (MIC) and distribution of scores in subgroups. No COSMIN score is

assigned for interpretability. Terwee's quality criteria were thus applied to consider interpretability as the qualitative meaning of quantitative scores and floor and ceiling effects.

5.5.2.1. Reliability

Five studies evaluating ARAT reliability were located. Retest reliability of the original ARAT was examined in one study (McDonnell et al., 2006). This showed that it was reliable between testing sessions ($ICC_{(3,1)}$ = 0.93 ± 0.05) in 17 participants, 2–7 months post-stroke. This study was only of fair methodological quality due to limited methodological detail being available and thus uncertainty regarding independent administrations of repeat measures. *The evidence synthesis resulted in an unknown level of retest reliability* due to a sample size <25 in participants post-stroke with no evidence located for people with a neurological condition with upper limb spasticity.

Three studies evaluated the intra-rater reliability (Page et al., 2012); all included participants >6 months post-stroke with nil identified upper limb spasticity. The studies found a high level of reliability, ICC ranging from 0.71 (95% CI 0.53–0.89) to 0.99 (95% CI 0.98, 0.99) within raters. The methodological quality of ranged from fair to excellent with two studies receiving final ratings of good and all three had positive quality criteria. *This synthesis of evidence found moderate positive evidence to support intra-rater reliability* when used with people >6 months post-stroke. There was no evidence located to support or refute intra-rater reliability when the original ARAT was used with people with upper limb spasticity.

One study examined inter-rater reliability (Yozbatiran et al., 2008). This was high when two blinded raters scored the original ARAT within the same session with nine participants who had a mean 34 months post-stroke (ICC 0.96). The methodological rating of this paper was good with a positive Terwee rating; ratings were reduced due to limited methodological detail provided*. This synthesis of evidence found unknown evidence for inter-rater reliability* due to the small sample size of the single study.

5.5.2.2. Responsiveness

A total of 10 studies were appraised with only nine considered in the best evidence synthesis stage of this study. These studies had a methodological quality ranging from fair to excellent. Ratings were reduced due to a lack of clearly specified hypotheses, application of less than optimal statistical approaches, including effect sizes, and uncertainty as to what occurred in the interim period between measurements. Three of the nine studies included participants with upper limb spasticity (Barden, Baguley, Nott, & Chapparo, 2014; Beebe & Lang, 2009; Lang et al., 2006). Synthesis of best evidence found that, for all nine studies, there was a *positive moderate level of evidence for responsiveness*.

Studies including participants with no identified limb spasticity used a range of statistical approaches to evaluate responsiveness and found the original ARAT to be responsive to change over time in acute through to chronic stroke and in chronic TBI. In comparison to 'like' measures of upper limb activity performance, the original ARAT performed well but was less responsive than the Chedoke Arm and Hand Activity Inventory (Barreca et al., 2005) when used with people within the first 6 months post- stroke. When used within a sample of people with upper limb spasticity, which included participants post-stroke and TBI, the synthesised evidence rating was reduced to 'limited' as a result of only three studies contributing to the evidence. All studies were of fair methodological quality and only one (Lang et al., 2006) had a positive Terwee criteria due to a responsiveness ratio (RR) >1.96. The original ARAT was found to have moderate to high effect sizes ranging from 0.55 to 0.78 across studies (Barden et al., 2014; Beebe & Lang, 2009).

5.5.2.3. Measurement Error

A single study which did not include participants with upper limb spasticity was located (Page et al., 2012). This reported a smallest detectable change (SDC) of 22.54 – the smallest amount of change attributed to true change and not random measurement error. COSMIN notes the close relationship and influence of measurement error and reliability, between and within raters and over time, on estimated SDCs. This study's methodological quality rating was 'fair' due to a lack of clear description of test conditions and independent administration; the assumption was that this was similar

and independent. The criteria for good measurement properties thus received an 'indeterminant' rating because the MIC was not defined and no convincing explanation was given that agreement was acceptable, particularly given the SDC was 40% of the total score of 57. This evidence synthesis concluded that the original ARAT has a *conflicting level of evidence for measurement error*, further reduced due to the small sample size.

5.5.2.4. Structural Validity

Lyle's, 1981, study detailing the development of the ARAT was the sole paper which evaluated structural validity. Lyle reported that the ARAT met Guttman Scaling criteria within each subscale achieving coefficients of scalability greater than 0.6 and coefficients of reproducibility greater than 0.9. This study had an excellent level of methodological quality and met requirements for a positive rating for good measurement properties. The small sample size ($n < 25$), how- ever, resulted in an '*unknown' level for structural validity*.

5.5.2.5. Hypothesis Testing (Construct Validity)

A total of 26 studies evaluated the psychometric property of construct validity. Only four of these had a primary study purpose of evaluating psychometric properties; 16 included the original ARAT to validate new or modified outcome measurement tools and the final six studies evaluated predictors of the use and recovery of the upper limb with the original ARAT as the measure of activity limitation. Methodological quality ranged from poor to excellent within individual studies. A lack of clearly stated hypotheses reduced ratings for the majority of studies. Less commonly, ratings were reduced due to limited details regarding blinding of assessors and handling of missing data.

The original ARAT was found to have a high correlation with other like-tests, including the Brunnstrom Fugl Meyer test (Fugl-Meyer, Jaakso, Leyman, Olsson, & Steglind, 1975), the WMFT (Wolf, Lecraw, Barton, & Jann, 1989), all versions of the Chedoke Arm & Hand Activity Inventory (Gowland et al., 1993) and the Arm Motor Ability Test-9 (McCulloch, Cook, Fleming, Novack, & Taub, 1988). Activity limitation as measured by the original ARAT was a predictor of grasp and release but not of overall quality of life.

The synthesis of evidence found the original ARAT to have s*trong positive evidence to support construct validity*. The synthesis of evidence considering participants with upper limb spasticity only, reduced in strength to a level of limited evidence with only six studies contributing. There is a *moderate level of positive evidence to support construct validity* in a sample of participants without upper limb spasticity with a total of 18 papers contributing to the final synthesis.

5.5.2.6. Interpretability

The original ARAT had highly variable results regarding floor and ceiling effects across the six studies that discussed this property. The variation overall ranged from 0 to 100% of participants for floor effects and 0 – 41% ceiling effects. COSMIN does not assign a rating for floor and ceiling effects – however, Terwee's criteria for good measurement properties assigns a rating. Positive ratings are assigned to studies if ≤15% of respondents achieved the highest or lowest possible score and a negative for ≥15% despite adequate design and methods. This is now applied in the evidence synthesis.

Dromerick et al. (2006) reported up to 41% of participants with a moderate degree of upper limb motor dysfunction with no identified spasticity scored the maximum 90 days post-stroke in comparison to 36% for the WMFT (Wolf et al., 1989) functional ability scale whilst still recording limitations on the FIM (Granger, Hamilton, Linacre, Heinemann, & Wright, 1993) and MAL (Uswatte & Taub, 1999). None of the participants in the study by Fleming, Newham, Roberts-Lewis and Sorinola (2014) who were greater than 6 months post-stroke and had identified upper limb spasticity were scored maximum or minimum in the original ARAT. Fleming's study found, however, that a score of 54 (out of a possible 57) was required to score a 2.5 on the MAL. Conversely, a floor effect was noted in Barden et al. (2012) study including patients with TBI and upper limb spasticity; and O'Dell et al. (2013) study sample which included people with severe functional limitations and no upper limb spasticity. Edwards et al. (2012) found no floor effects within the first 90 days post-stroke. These findings indicate that it is more the level of upper limb activity limitation which impacts on the 1981-ARAT's ability to detect change rather than the time post-diagnosis. The synthesis revealed conflicting evidence regarding interpretability for all studies and for studies when stratifying by the presence

of upper limb spasticity. Interpretability can also be affected by the MIC. MIC was reported in a single study in an acute setting by Lang et al. (2008). MIC was 12 for the dominant affected limb and 17 for the non-dominant limb. Further studies are required, but evidence for *MIC was synthesised to be positive*.

Table 5.2. *Summary of study results*

Study author, year	Internal Consistency	Inter-rater reliability	Intra - rater reliability	Retest reliability	Measurement error	Content Validity	Structural validity	Hypothesis Testing (Construct validity)	Responsiveness	Interpretability	Floor and ceiling effect
Barden et al., (2014)								r=0.50 – 0.63 DCD	ES=0.78		Floor 100%
Barden et al., (2012)								Predictive of grasp/release p= -0.43 - 0.73	ROC curve=0.88 (95% CI 0.76-1.00)		Floor 26% Ceiling 5%
Barreca et al., (2005)								CMSA r=0.81-0.93	ROC curve 0.88 (95% CI 0.76-1.00)		
Barreca et al., (2006)								CAHAI -9,13 r=0.93–0.95	ROC curve 0.72.		
Barreca, Stratford, Masters, Lambert, Griffiths, et al., (2006)								CMSA, CAHAI-7, 8, 9 l wk 0 r=0.87-0.95 (one sided 95% CI 0.68 -0.91), Wk 2-6 r=0.92-0.94 (one sided 95% CI 0.81-0.90)			
Beebe & Lang (2009)								1, 3, 6 mo r_s= 0.61–0.95, grip, dexterity.	ES=0.55,0.63		
Blennerhassett, Avery, & Carey (2010)								HFS baseline r=0.96 (95% CI 0.90-0.98) 4-6 wk r=0.95 (95%CI 0.87-0.98).	Rho_c=062, 95% CI 0.35-0.90, Kw=0.65		
Burridge et al., (2009)								Negative UMN r = 0.025 – 0.710, Positive UMN r = -0.008 – 0.231			
Celik et al., (2010)								Robotics r =-0.83 - 0.51			Floor 0% Ceiling 22%
De Weerdt & Harrison (1985)								B-FM 2 wk r=0.91, 8 wk r=0.94	T=1, n=28, z=4.60		

Study author, year	Internal Consistency	Inter-rater reliability	Intra - rater reliability	Retest reliability	Measurement error	Content Validity	Structural validity	Hypothesis Testing (Construct validity)	Responsiveness	Interpretability	Floor and ceiling effect
Dromerick et al., (2006)								FIM r=0.47, MAL QOM r=0.61, WMFT time, functional ability r=0.65, 0.95.			Ceiling 41%
Edwards et al., (2012)									ES=1.018, 1.390		Floor 2 - 5.9% Ceiling 3.9 - 33%
Fleming et al., (2014)								54=MAL 2.5 Predictive validity R_2=0.6; F1, 17=25.518; P<0.001.			Floor 0% Ceiling 0%
Lang et al., (2008)										MCID 12, 17 (dominant, non-dominant)	
Lang et al., (2006)								spasticity r=-0.28 to -0.49, disability r=0.2–0.6	RR=5.200, 7.067 ES=1.018		
Lyle (1981)							Guttman Scaling criteria met				
McDonnell et al., (2006)				ICC $_{(3,1)}$ 0.93 +/- 0.05 range 0.83-0.90				FMA r=0.75, grip strength r=0.73, tapping speed r=0.61.			
Morris et al., (2013)								RMA r=-0.80, NHP r=-0.25. Greater UL dysfunction associated with poorer HRQOL. Not predictive of overall HRQoL.			
Notley et al., (2007)								Tracking accuracy r = -0.441 - 0.829			

Study author, year	Internal Consistency	Inter-rater reliability	Intra - rater reliability	Retest reliability	Measurement error	Content Validity	Structural validity	Hypothesis Testing (Construct validity)	Responsiveness	Interpretability	Floor and ceiling effect
O'Dell et al., (2013)								AMAT-9 r=0.79, SIS hand r=0.40	SRM 0.89.		
Page et al., (2015)			ICC 0.99 (95%CI 0.98, 0.99					w/h UE FM pre-test r=1 0.74 (P<0.001) post-test 2 r=0.67 (P<0.001)			
Page et al., (2012)			ICC 0.71 (95% CI 0.53-0.89)		SDC 22.54			w/h UE FM r_s=0.72			
Rabadi & Rabadi (2006)								FMA r=0.77 p<0.001 - 0.87 p<0.001	SRM 0.68		
Rand & Eng (2015)								Discharge ARAT and 12 mo MAL r=0.78, p< 0.001, accelerometer r=0.58, p< 0.001, predicting 12mo UL function R_2=0.776, P<.001 (MAL), 0.470, P.001			
Stinear et al., (2012)								PREP algorithm predicted 12-week ARAT			
Urbin et al., (2015)								sensor acceleration r=0.73 - 0.85, P<.001			
Yozbatiran et al., (2008)		p 0.96, ICC 0.9986	p 0.99 ICC 0.99					arm Fugyl-Meyer r=0.94 (P<0.01)			

Study author, year	Internal Consistency	Inter-rater reliability	Intra - rater reliability	Retest reliability	Measurement error	Content Validity	Structural validity	Hypothesis Testing (Construct validity)	Responsiveness	Interpretability	Floor and ceiling effect
Summary of results	N/A	ICC 0.9986	ICC 0.71 0-0.99	ICC $_{(3,1)}$ 0.93 +/- 0.05	MDC 22.54	N/A	Guttman Scaling criteria met	Measures UL function as intended. r = 0.,25 – 0.95, not predictive HRQoL	ROC curve=0.72 – 0.88 ES=0.52–1.390 RR=5.20–7.067 SRM=0.68–0.89 Rho_c=0.62 (95%CI 0.35–0.90) Kw=0.65 Z=4.60-4.85	MCID 12, 17	Floor 0 - 100% Ceiling 0 - 41%

Table 5.3. *Synthesis of Best Evidence and Criteria*

	Internal consistency	Reliability (Inter)	Reliability (intra)	Reliability (retest)	Measurement Error	Content Validity	Structural validity	(Construct validity)	Cross cultural validity	Responsiveness	Interpretability	Floor ceiling effect
Whole sample	0	?	++	?	+/-	0	?	+++	0	++	+	+/-
UL spasticity present	0	0	0	0	0	0	0	+	0	+	0	+/-
UL spasticity not present	0	?	++	?	+/-	0	?	++	0	++	+	+/-

Level	Rating	Criteria
Strong	+++ or --- total sample size ≥ 100)	Consistent findings in multiple studies of good methodological quality or one study of excellent methodological quality
Moderate	++ or – (total sample size 50 – 99)	Consistent findings in multiple studies of fair methodological quality or in one study of good methodological quality
Limited	+ or – (total sample size 25 – 49)	One study of fair quality
Conflicting	±	Conflicting findings
Unknown	? (total sample < 25)	Only studies of poor methodological quality

5.6. Discussion

This systematic review considered methodological quality of the original ARAT when used within neurorehabilitation. The original ARAT is a tool commonly used in clinical practice and trials to measure the ability to perform activities with the upper limb. The original ARAT is one of the earliest measures of upper limb performance (Page et al., 2012), yet despite such history, this review located very few studies that specifically investigated performance of this tool. This is surprising because it has been used as a presumably psychometrically acceptable comparator to validate or evaluate other existing, new or modified tools.

Study results demonstrate clinicians and researchers alike can be confident that when using the original ARAT with people post-stroke and with TBI without the presence of upper limb spasticity that a *moderate to strong level of evidence supports the intra-rater reliability, construct validity and responsiveness* of this tool. This review identified *several areas where only limited or inconclusive evidence exists*, specifically: inter and test–retest reliability (do different raters administer and score differently? Is there an influence on performance of the test items on repeat session over time?); measurement error (what is true change in performance and not the result of systematic or random error?); content validity; structural validity; and floor and ceiling effects (does the type of patient or the timing of use of the ARAT matter?).

This systematic review demonstrates differences within the psychometric properties of the original ARAT when the presence of upper limb spasticity is apparent in study samples. The level of evidence is significantly reduced and/or missing across various properties with only *limited positive evidence identified for construct validity and responsiveness*. An *inconclusive level of evidence was identified for floor and ceiling effects* and nil other evidence for remaining properties located.

This review highlights the importance of study or practice purpose and the clinical or research context in determining whether or not selection of the original ARAT is appropriate. This is because ratings of best evidence synthesis vary across participants

and contexts. Practically, conflicting evidence drawn from a single study for the SDC (22.54) casts doubt on assumptions about utility of this measurement tool.

The paucity of studies contributing to best evidence synthesis indicates that more research is needed to provide the evidence needed for sensible interpretation of results. Interpretation decisions should also reflect change deemed important by patients. The MIC has been reported to be 10% of the total score or approximately a change of 6 in the subacute and chronic phase (Van Der Lee et al., 1999). In contrast, this review identified one acute setting study where the MIC was found to be 12 for the dominant affected limb and 17 for the non-dominant limb. A higher MIC in the acute setting has been proposed by Lang et al. (2008) to be due to the large portion of recovery that can occur during this period and strong expectations for continued recovery. Recovery expectations exemplify differences in MIC and SDC, the influence of both on the utility of the ARAT and need for clinicians to be aware of this measurement property in interpretation of results.

Studies frequently did not report detailed information regarding the methodology, the manner in which missing items were handled, a lack of clearly defined hypotheses and less than optimal statistical analyses particularly for construct validity and responsiveness. This reduced quality ratings. It is possible that if these had been reported, stronger ratings could have been made. As highlighted by Kennedy et al. (2013), COSMIN does not differentiate between poor methodological quality and poor reporting; thus, lower ratings may be a reflection of either poor methodological quality or underreporting of study characteristics.

Studies were excluded from this review because they did not meet inclusion criteria. The exclusion of studies where the original ARAT was not administered in English or the report was not in English is one example. This exclusion was made because of a lack of cross-cultural or translation validation studies. Thus studies completed in the Netherlands, Taiwan and China were excluded which would otherwise have been eligible. Another exclusion that led to a narrowing of the evidence base was the decision to use only those studies where the original ARAT had been administered in full with no

modification. This exclusion was made because of the variability in potential modifications and the lack of transparency and consistency that modifications introduce. These exclusions together meant a large body of evidence were sacrificed to attain study results from relatively homogenous studies. The review method itself holds limitations. Despite COSMIN criteria being quite explicit, implementation of the appraisal tool requires a level of individualisation for each application. COSMIN replicability in systematic reviews is an inherent and unavoidable limitation.

5.7. Conclusions

The ARAT is a frequently used outcome measure in clinical neurorehabilitation practice and research (Santisteban et al., 2016). In terms of suitable clinical populations, this review of the original 1981 version provides evidence that it is appropriate to use with people post-stroke and there may be potential for use within TBI populations (although the small sample size of included participants post-TBI means further work is required in a TBI population). In terms of psychometric properties, the original ARAT has been shown to measure what it seeks to measure, that is, upper limb activity limitation. It is able to detect change in upper limb activity performance. More evidence is needed to understand the smallest detectable change (SDC) and minimal important change (MIC) at different recovery time points and in different neurorehabilitation populations including those people with upper limb spasticity. The presence of upper limb spasticity significantly reduces recommendations regarding routine use in clinical practice due to limited evidence found only for construct validity and responsiveness. There is a need for further work to apply the ARAT to neurorehabilitation post-stroke populations and other populations with upper limb hemiparesis and spasticity with more rigorous research methodology and meticulous reporting to build evidence about use of the original version of the ARAT in neurorehabilitation.

5.8. Chapter Synopsis

Chapter 5 synthesised evidence for the psychometric properties of the ARAT, and demonstrated acceptable properties for its use with adults post-stroke. Further validation is required for its use with adults post-stroke who have upper limb spasticity, and those who have suffered a TBI. Together Chapters 3, 4 and 5 have presented

extensive psychometric evidence about upper limb activity outcome measurement tools used in neurorehabilitation. To tailor measurement processes to meet the needs of each patient, goal attainment scaling is recommended as an adjunct to standardised measurement. Thus providing an individual focus on measurement beyond what tools such as the ARAT can provide which evaluates upper limb activity within a clinical setting. The following Chapters (6, 7) compare two goal attainment scaling methods, the original GAS and GAS-Light, to provide clinicians with evidence to support practice decisions.

Chapter 6:

Comparison of Goal Attainment Scaling (GAS) and GAS-Light Performance

The work covered in this chapter has been submitted as:

Pike, S., Cusick, A., Turner-Stokes, L., Buckley, D., Li Teng Han, M., & Lannin NA (2020 under review). Comparison of Goal Attainment Scaling (GAS) and GAS-Light performance in neurorehabilitation.

See Appendix B for ethics approval and Appendix G for supplementary material.

6.0. Chapter Overview

Goal directed neurorehabilitation targeting outcomes that are meaningful and specific to the ABI survivor is considered best practice. Significant variations in the tools and approaches used to set, document and evaluate goal achievement exist. Consensus on how to best complete goal setting to direct upper limb neurorehabilitation programs and capture individualised outcomes is lacking. Goal attainment scaling (GAS) is a structured method to set, document and evaluate individual goals. It is used extensively in healthcare, rehabilitation and has been recommended for and used in neurorehabilitation for many years. This study compares the performance of two GAS methods to provide clinicians with psychometric evidence to support their selection of the most appropriate tool for their clinical practice context and adherence to best practice.

6.1. Background

Goal setting is an integral component of best practice in neurorehabilitation (Hurn et al., 2006; Levack et al., 2006; Sugavanam et al., 2013b; Wade, 2020), with the goals set then used by the clinical team to direct neurorehabilitation programs. Setting neurorehabilitation goals is, however, complex (Levack & Siegert, 2014) and this complexity likely contributes to a lack of consensus on how goals should be set, recorded and evaluated. While there is a wide range of approaches to goal setting adopted across services, and in fact between clinicians within the same service, goal setting is perceived to improve patient engagement in and satisfaction with neurorehabilitation, team communication regarding the neurorehabilitation program and the ability for the team to tailor each patient's individual program to meet their needs. The evidence of how effectively goal setting achieves this has been synthesised by three systematic reviews (Levack et al., 2015; Levack et al., 2006; Sugavanam et al., 2013), which have all concluded that heterogeneity and low methodological quality of included studies means that whether or not goal setting achieves these outcomes remains unknown. While it is not yet known which approach to goal setting should be used, whether goal setting does improve outcomes attained during neurorehabilitation, goal setting has been recommended in national clinical practice guidelines internationally for a number of decades.

While in some clinical settings, goals may be set without any evaluative reference back to goal achievement, there is a growing acknowledgement that likely benefits of setting goals comes with the reflection on the patient's achievements; further that such reflection can inform potential changes to the neurorehabilitation program as it progresses (Hurn et al., 2006; Turner-Stokes, 2009). One such measure of individualised goal achievement is that of goal attainment scaling (GAS) (Kiresuk & Sherman, 1968). Following the introduction of GAS by Kiresuk and Sherman (1968) in a mental health clinical setting, GAS has been adopted within other areas of healthcare including geriatrics (Bouwens, van Heugten, & Verhey, 2008; Burnes, Connolly, Hamilton, & Lachs, 2018; Stolee, Rockwood, Fox, & Streiner, 1992), pediatric rehabilitation (Harpster et al., 2019; Sakzewski, Boyd, & Ziviani, 2007), chronic pain (Zaza, Stolee, & Prkachin, 1999), focal spasticity management (Ashford & Turner-Stokes, 2006; Sheean et al., 2010; Turner-Stokes et al., 2010) and adult neurorehabilitation (Hurn et al., 2006; Khan et al., 2008). In neurorehabilitation, GAS allows clinicians to document and evaluate the extent of goal achievement (Malec, 1999), complimenting other standardised outcome measurement tools (Turner-Stokes, 2009). The ability of the tool to capture progress for patient-specific outcomes allows measurement of the achievement of individualized neurorehabilitation goals. This ability of GAS to capture diverse, individualized goals with a single measure has contributed to the growing use of GAS across neurorehabilitation (Turner-Stokes, 2009).

In neurorehabilitation, GAS has been widely used and is well-validated (Bovend'Eerdt et al., 2011; Brock et al., 2009; Doig et al., 2010; Hurn et al., 2006; Joyce et al., 1994; Khan et al., 2008; Lannin, 2003; Malec, 1999; Malec et al., 1991; Nott et al., 2014; Turner-Stokes, 2009; Turner-Stokes et al., 2010). There is a substantial body of research outlining the advantages of using GAS from both clinical and research perspectives in neurorehabilitation (Ertzgaard, Ward, Wissel, & Borg, 2011; Stevens, Beurskens, Köke, & van Der Weijden, 2013; Turner-Stokes, 2009; Turner-Stokes, Williams, & Johnson, 2009). However, clinicians have also highlighted challenges using GAS within the clinical setting, most notably the utility of the tool (Ertzgaard et al., 2011; Grant & Ponsford, 2014; Stevens et al., 2013; Turner-Stokes, 2009). The GAS has been criticised for being time-consuming to administer and score, that the use of zero and negative numbers in the scale can be discouraging to patients, and that there is no capacity to record partial goal

achievement despite clearly demonstrated progress and the benefits of being able to record individualised attainment for outcome evaluation and service evaluation (Turner-Stokes, 2009). In response to these clinical challenges, Turner-Stokes (2009) proposed the GAS-Light method, a version of the GAS which is brief, uses a different scoring scale, but retains the key attributes of individualised goals and scaling of goal attainment by the patient. Clinical choice has been reported by Australian clinicians but there is a lack of guidance in the literature to date regarding the equivalence of the two tools. Importantly, the psychometric properties of GAS-Light and how it compares to GAS when used to measure goal attainment in adult neurorehabilitation have not been systematically studied.

6.2. Study Aim

Therefore, the aim of this study was to examine the reliability and validity of the GAS-Light (a briefer version of the original GAS), in adults who are receiving neurorehabilitation for an upper limb motor impairment. GAS-Light was hypothesised to provide a reliable measure of attainment of upper limb activity and participation goals, reflected by strong internal consistency. To examine concurrent validity of the GAS-Light, goal attainment levels were tested against the original GAS, and the sensitivity of GAS-Light was determined by examining its ability to detect real life changes and the attainment of goals in adult patients attending neurorehabilitation for upper limb motor impairment.

6.3. Research Questions

1. What are the psychometric properties of GAS-Light when compared to GAS?

    i. Is GAS-Light valid in an Australian neurorehabilitation setting when determining upper limb activity and/or performance goals?

    ii. Is GAS-Light sensitive enough to detect real-life changes and the attainment of goals in adult patients undergoing neurorehabilitation for upper limb motor impairment?

6.4. Method

Institutional ethical approval was obtained from the University of Wollongong/Illawarra Shoalhaven Local Health District Medical Human Research Ethics Committee (approval number HE12/077), La Trobe University Faculty Human Ethics Committee (approval number FHEC12/152), The Alfred Human Research Ethics Committee (approval number 38/13) with local site-specific approval obtained from the Murrumbidgee Local Health District (MLHD) (approval number SSA/12/MLHD/107). Informed written consent was provided by all participants or their nominated person responsible where participants were clinically deemed unable to provide informed consent.

A repeated measures design (Verma, 2015) was adopted to compare GAS-Light (Turner-Stokes, 2009) to the original GAS (Kiresuk & Sherman, 1968). A multi-site recruitment method was used: one site was a regional rehabilitation service (inpatient and community), the other site was a metropolitan spasticity clinic. Adult patients were invited to participate if they: presented to either service with an upper limb motor impairment limiting engagement in daily activities or, in the absence of disruption of functional use, moderate spasticity (as indicated by a score of 3 or greater on the Modified Ashworth Scale); and were naïve to GAS (that is, they had not previously set goals using the GAS method). We included participants whose upper limb motor and/or spasticity impairment was as a result of an acquired brain impairment due to stroke, traumatic brain injury, multiple sclerosis, or cerebral palsy. Persons under 18 years of age and persons identified as belonging to vulnerable groups were excluded, they were patients highly dependent on medical care or pregnant, so as to maximise follow-up of all participants.

At baseline, the following participants data was collected: demographic information, a brief screen of cognitive impairments (as measured using the Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975)), the National Institutes of Health Stroke Scale (Lyden et al., 1999), and upper limb muscle strength using manual muscle testing.

Recruited participants engaged in usual clinical measurement by the neurorehabilitation team and as part of usual practice identified their neurorehabilitation goals. The treating

clinician was either an occupational therapist or physiotherapist and alone or in combination they facilitated these measurement and goal setting usual care sessions. As part of the study protocol, these therapists completed the GAS-Light tool as the framework for the goal setting process. The independent assessor concurrently completed the GAS tool. This process is depicted over page in Figure 6.1. The treating clinician and the independent assessor were present at the same time and goal-scale development occurred simultaneously but without consultation between the treating clinician and the independent assessor (treating clinician was blind to the independent assessor scaling). This approach ensured consistency of information provided by participants and/or family/carers and allowed the treating clinician and independent assessor to observe task performance and equally receive any other information to inform goal setting and evaluation.

Treating clinicians were occupational therapists and physiotherapists employed within recruiting services. All treating clinicians received standardised training in both the GAS and GAS-Light methods prior to participation in the study, see Appendix G for training program outline. Two independent assessors (one at each site) were experienced occupational therapists with 15 years and 16 years respectively of clinical experience.

Participants then engaged in usual care, which included participation in a tailored, goal-directed neurorehabilitation program; information regarding intervention provided was not collected in this study. Data were collected at baseline (goal setting session) and follow-up (pre-determined date for goal evaluation session at neurorehabilitation program end). At the participants pre-determined follow up appointment, the treating clinician directed the session, including repeating any baseline standardised assessments and/or observing functional task performance as required, to complete the GAS-Light tool. The independent assessor attended this appointment and completed the GAS tool independently with neither result compared nor discussed between the treating clinician and independent assessor. When required, the independent assessor sought further clarification or information from the participant to determine the correct level of attainment for completing the GAS follow up guide; this occurred after the GAS-Light scoring was finalised.

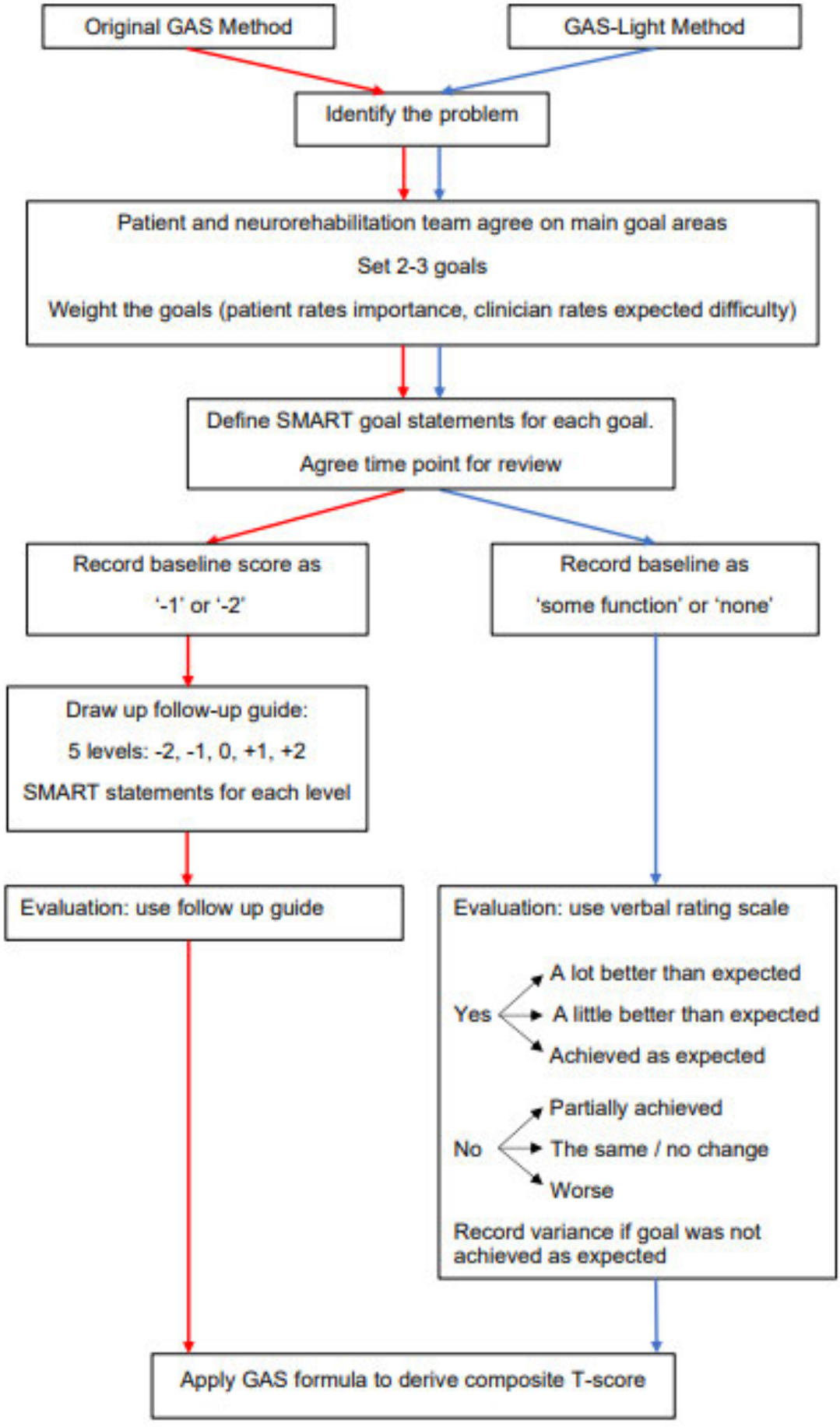


Figure 6.1. *GAS and GAS-Light process applied in the study*

6.4.1. Goal Attainment Scale

GAS (Kiresuk & Sherman, 1968) provides a structured approach to define and measure individualised patient and/or program based goal attainment. GAS has demonstrated reliability (Bovend'Eerdt et al., 2011; Joyce et al., 1994) concurrent validity (Brock et al., 2009) and responsiveness (Khan et al., 2008; Lannin, 2003) when used within neurorehabilitation. The GAS method used in this study was published by Bovend'Eerdt et al. (2009). When determining goal attainment, all information provided by participants and/or family/carers, and/or observed task performance and/or any other objective information is considered to inform the level of goal attainment as per the follow up guide.

6.4.2. Goal Attainment Scale-Light

The GAS-Light method published by Turner-Stokes (2009) was used in this study, as depicted in Figure 6.1. When assessing goal attainment, all information provided by participants and/or family/carers, and/or observed task performance and/or any other objective information is considered to inform the level of goal attainment as per the verbal rating scale. While GAS-Light has minimal published psychometric data, it has demonstrated a correlation with measures of function (Arm Activity Scale $r = 0.63$, $p<0.001$), impairment (Tardieu $r = 0.43$, $p<0.001$, active range of motion $r = 0.41$, $p<0.001$, passive range of motion $r = 0.43$, $p<0.001$, associated reaction rating scale $r = 0.76$, $p<0.001$), symptoms and carer report (visual analogue scale $r = 0.46$, $p<0.001$) within spasticity management (Turner-Stokes, Fheodoroff, Jacinto, & Maisonobe, 2013). Furthermore, preliminary evaluation (unpublished) suggested acceptable accuracy (86-92%) and time savings when compared against a pre-prepared follow up guide (Turner-Stokes, 2009).

6.4.3. Data Analysis

A priori power calculations indicated a sample size of 70 participants was required to determine the validity and reliability of GAS-Light (Dunn, 1989). Analyses were performed using SPSS version 25 for Windows, with statistical significance set at 0.05. GAS and GAS-Light data did conform to criteria for normality (Kolmogorov -Smirnov

<0.05, Shapiro-Wilk <0.05) therefore parametric statistical tests were used where indicated.

Participant goals were linked to the International Classification of Functioning, Disability Health (ICF) (World Health Organisation, 2001) by the thesis author following linking rules proposed by Cieza (Cieza et al., 2019; Cieza et al., 2005). Descriptive statistics were generated to show frequency of goals categorised into ICF domains and categories.

The GAS formula is designed to generate a single aggregated T-score with a mean of 50 and standard deviation of 10 (Kiresuk & Sherman, 1968). Box plots were used to compare overall neurorehabilitation program baseline and achieved summative T-scores determined by each tool. Reliability of the GAS-Light approach was computed via the intra-class correlation coefficient, two-way random effect models and 'single rater' unit (Barton & Peat, 2014) (results interpreted as per Koo and Li) (Koo & Li, 2016). Agreement between the tools was investigated through the calculation of measurement error, error range and assessing regression. To determine whether the level of agreement between GAS and GAS-Light was affected by relationships with participant characteristics or goal types, mutual information score and Chi Square score were calculated. The Bland Altman method also tested agreement between the tools in addition to determining the presence of systematic bias. The Bland Altman method plotted the average achieved T-score of both methods against the difference between the means (Barton & Peat, 2014). Effect size was calculated using Cohens's *d* (mean change/standard deviation of baseline score) and standardised response mean (SRM) (mean change / standard deviation of change score), two commonly used methods providing differing results (Khan et al., 2008; Rockwood et al., 2003). It was hypothesised that the effect size would be equal across approaches. We used the criteria suggested by Koo and Li (2016) for judging the strength of the correlations obtained based on the 95% confidence intervals of the estimate: <0.5 for poor, 0.5-0.75 for moderate, 0.75-0.9 for good, and >0.90 for excellent.

6.5. Results

A total of 61 participants provided informed consent over the study period; one participant did not proceed with their neurorehabilitation following the initial goal

setting session and this incomplete data was excluded. The scatterplot of GAS and GAS-Light T-scores showed that two participants had differences between the achieved T-score for both methods exceeding the mean difference by greater than 2.5 standard deviations; for this reason their data was excluded as outliers (criteria published by Hough, 1999). The remaining n = 58 participant data is presented in Table 6.1.

Table 6.1. *Participant and clinician demographics*

Participant characteristics	Diagnosis				Total (n=58)
	Stroke (n=51)	TBI (n=3)	CP (n=3)	MS plus Stroke (n=1)	
Age *(yr)*, mean (SD)	61.94 (13)	45 (12)	23 (7)	65 (0)	59.09 (15)
Sex, number male (%)	31 (52)	2 (67)	1 (33)	1 (100)	33 (57)
NIHSS mean (SD) [range]	5.32 (4.07) [0 – 15]*			11 (0) [11]	
MMSE, number <19, (%)	5 (10)	0 (0)	2 (67)	1 (100)	8 (13)
MMSE≥19, mean (SD)	27.02 (3.23)	26.33 (4.62)	30 (0)	0 (0)	27.04 (3.26)
UL spasticity, n (%)	16 (31)	3 (100)	3 (100)	1 (100)	23 (40)
Affected UL					
n right (%)	22 (43)	3 (100)	0 (0)	1 (100)	26 (45)
n left (%)	28 (55)	0 (0)	1 (33)	0 (0)	29 (50)
n both (%)	1 (2)	0 (0)	2 (67)	0 (0)	3 (5)
Baseline UL Function n (%)					
No weakness	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Mild weakness	27 (53)	2 (67)	1 (33)	0 (0)	30 (52)
Significant weakness	17 (33)	1 (33)	2 (67)	1 (100)	21 (36)
Total paralysis	7 (14)	0 (0)	0 (0)	0 (0)	7 (12)

Clinician characteristics	Occupational therapist (n = 11)	Physiotherapist (n = 1)	Total (n = 12)
Clinical experience *(yr),* mean (SD)	7 (3.2)	19 (0)	8 (4.6)

*n = 50, TBI = Traumatic Brain Injury, CP = Cerebral Palsy, MS = Multiple Sclerosis, NIHSS = National Institutes of Health Stroke Scale, MMSE = Mini Mental State Examination, UL = Upper Limb

Participants were between the ages of 33 and 87 years (mean 59, SD 15.44), 33 (57%) were male with the majority of participants post-stroke, n = 51 (88%). Twenty-three (40%) participants had identified upper limb spasticity, 30 (52%) participants were determined to have mild weakness and 21 (36%) significant weakness. Treating clinicians included occupational therapists (n = 11) and physiotherapists (n = 1) with 2 – 19 years (mean = 8 (SD 4.6)) clinical experience.

Participants identified a total of 132 goals; these were set and evaluated at the pre-determined follow up session. On average participants identified two goals (range 1-4) as part of their neurorehabilitation program. The majority of goals were categorised within the Activity and Participation domain (71%) of the ICF, with the remaining coded to the Body Functions domain (29%). The majority of goals concerned 'mobility' followed by 'self-care' and 'neuromusculoskeletal, and movement related functions', see Figure 6.2.

Baseline and achieved summative T-scores for GAS-Light and GAS are presented in box plots in Figure 6.3 over page. At baseline, the two tools collected goal data equally. At follow up, results show that the median achieved T-score was slightly higher for the GAS tool in comparison to GAS-Light.

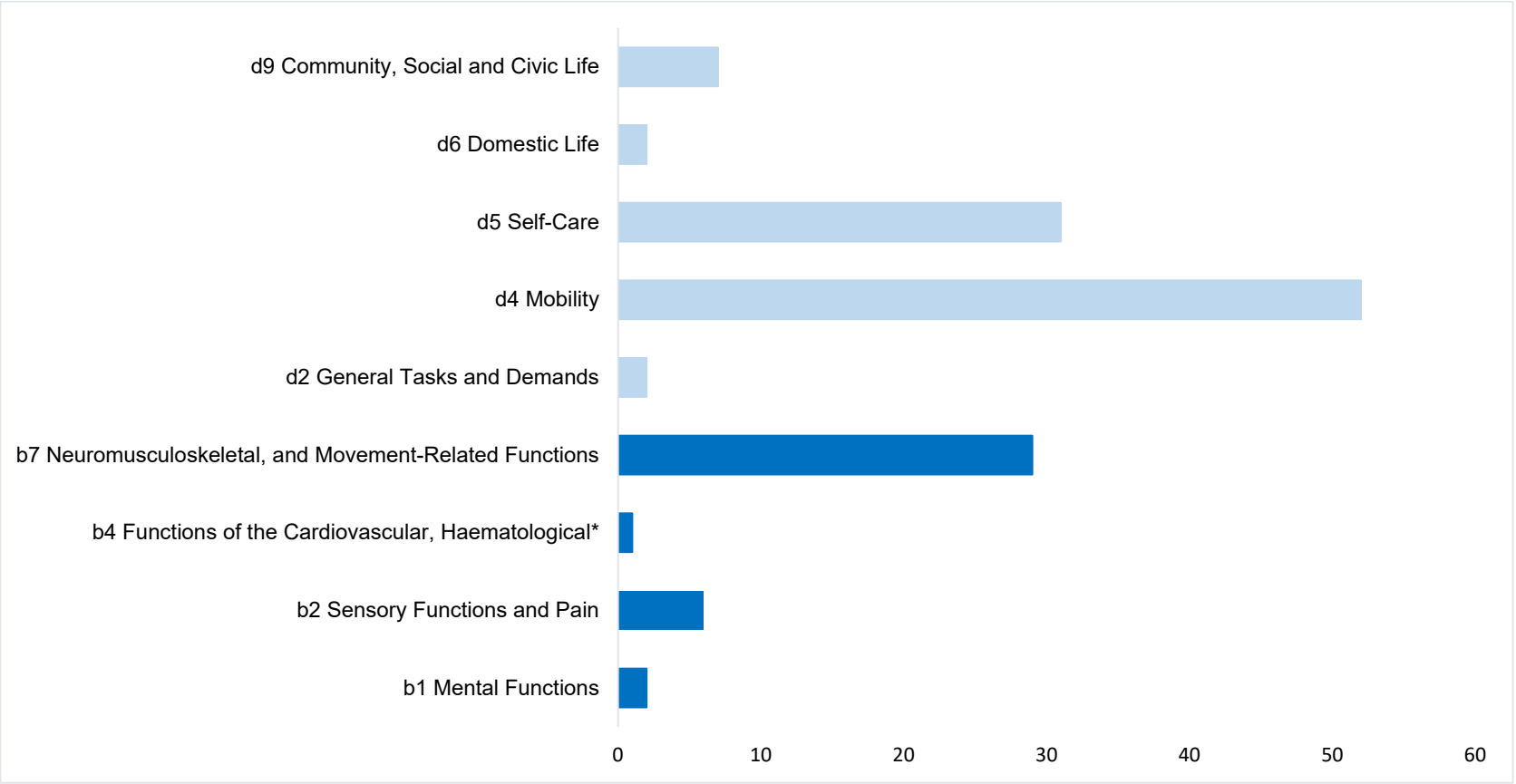


Figure 6.2. *ICF chapter goal frequency*

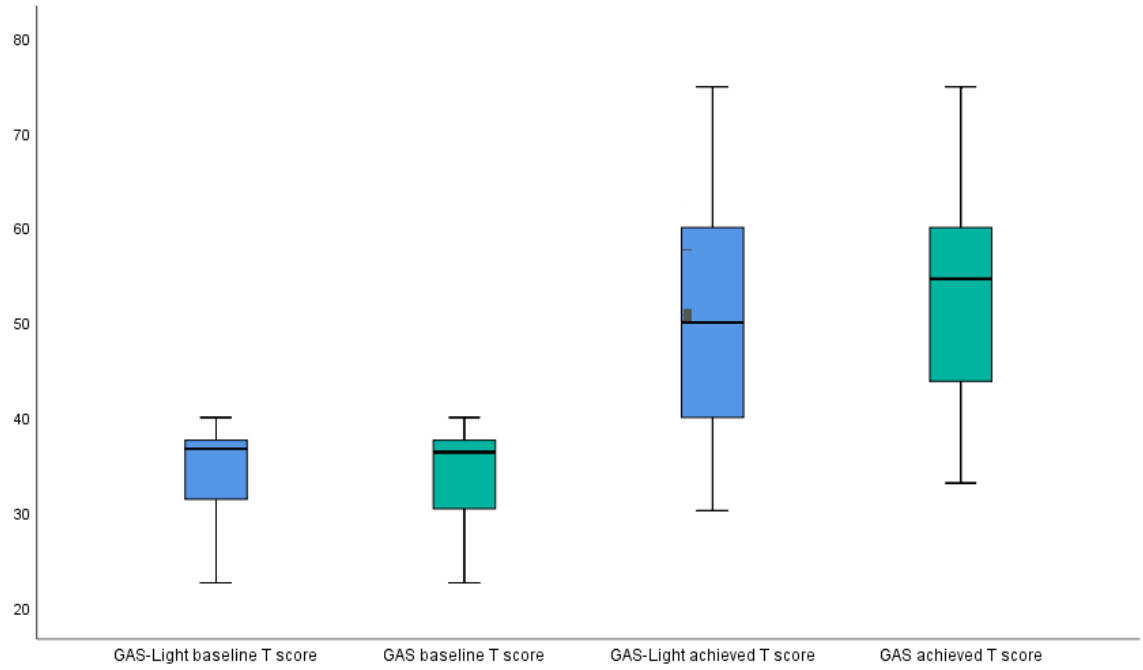


Figure 6.3. *Box plot*

The intraclass correlation coefficient (ICC) confidence interval suggests good to excellent correlation between GAS and GAS-Light T-scores (ICC (2,k) = 0.91, 95% CI 0.84 to 0.95, P<.001), suggesting that goal attainment was measured similarly by the two tools. The Bland Altman plot shows excellent agreement between GAS and GAS-Light measurements with >95% of the data points within ± 2SD of the mean difference. Agreement and measurement error are shown in Table 6.2 and Figure 6.4.

Table 6.2. *Reliability of goal attainment measured by GAS-Light and GAS*

	Mean difference	Limits of agreement	Error range	ICC
Goal attainment	-1.8897	-10.30, 6.52	5.95	0.91

A small mean systematic difference of -1.89 between the two methods was found, suggesting a small amount of systematic bias, with 8.14 random error and limits of agreement -10.30 to 6.52. The regression with GAS-Light (achieved) as the dependent

variable, revealed GAS (achieved) explained 85.7% variance in GAS-Light. Error range indicates that the range within which the true GAS-Light score lies is 5.95 above and 5.95 below GAS ratings. The mutual information score for each patient attribute and goal types was negligible with an average of 10^{-2} for both baseline T-score and change T-score suggesting no correlation with GAS-Light performance. This finding was reinforced by the negligible Chi-square score of 10^{-3}. Together these findings suggest the differences across the GAS-Light and GAS are the result of randomness.

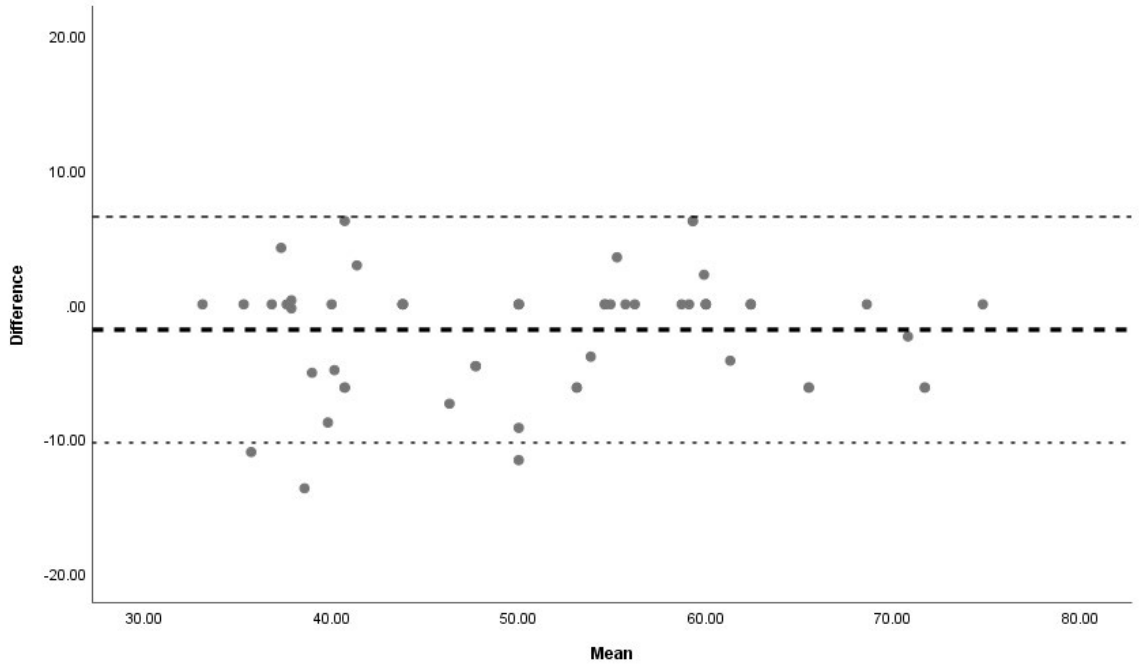


Figure 6.4. *Bland-Altman plot*

Table 6.3 presents the effect sizes of the two tools, showing they were both large and similar across GAS and GAS-Light, acknowledging that GAS was found to be slightly more sensitive to change in the study sample.

Table 6.3. *Effect size of GAS-Light and GAS in sample n=58*

	Cohen's *d*	SRM
GAS-Light	3.42	1.60
GAS	3.68	1.81

6.6. Discussion

This study shows the GAS-Light tool to be both valid and reliable when used with adults within a neurorehabilitation program. GAS-Light can be considered as a suitable tool to not only capture pre-defined expected levels of performance but then to ascertain when that performance exceeds expectations (by a little or a lot), deteriorates or fails to change. Furthermore, this study shows that GAS-Light is a suitable alternative to GAS in the neurorehabilitation clinical setting, providing similar standardised ratings of change, with the benefit of being able to capture partial goal achievement, a limitation of original GAS.

Setting goals in neurorehabilitation has been suggested to assist with both motivation and engagement in neurorehabilitation programs (Levack et al., 2006) and thus having methods to discuss goal attainment may assist with clinical decision-making and directing ongoing neurorehabilitation (Hurn et al., 2006; Turner-Stokes, 2009). GAS is a well-established method to track goal attainment, and in this study, it was demonstrated that GAS-Light can also be used to measure patient change in neurorehabilitation. In addition, GAS-Light captured partial goal achievement in our study for 19% of the set goals, a capability unique to the GAS-Light tool. The impact on patient experience of neurorehabilitation and being able to report partial achievement (as opposed to no-change) with patients while discussing further neurorehabilitation was not evaluated in this study, but is an area recommended for future research. It is acknowledged that the rating of partial goal achievement within GAS-Light may have resulted in an underreporting of goal achievement in this study; specifically, GAS-Light achieved mean T-score 50.50 (11.37) while GAS was 52.39, (11.00). Turner-Stokes and Williams (2010) in comparison, identified overestimation, or over reporting of goal achievement with modified scoring achieving a mean T-score of 48.7 (6.5) and standard scoring achieving a mean T-score of 48.2 (6.8). Variations in results between our GAS-Light findings and that study may be due to the inclusion of different participant diagnoses (Turner-Stokes & Williams, 2010). Regardless, both studies highlight that slight differences in T-scores between scoring methods may arise. Whether or not these small differences are clinically relevant will remain up to each clinician to determine.

Although this is the first study which aimed to test the psychometric properties of GAS-Light by comparing GAS-Light to GAS in a neurorehabilitation context, other studies have provided insights to: interrater consistency of GAS-Light (Bovend'Eerdt et al., 2011); and the use of different scoring methods to permit inclusion of partial goal achievement in GAS-Light outcomes (Turner-Stokes & Williams, 2010). This study sheds new light on the issue of rater agreement in goal attainment scaling adding to earlier work by Bovend'Eerdt, Botell and Wade (2011). They found poor agreement between treating clinician and independent assessor ratings on the GAS, whilst we found a high reliability and a high level of agreement between GAS and GAS-Light and between raters. Our raters were trained by the thesis author prior to study implementation, and supported in administering the GAS and GAS-Light during the study via mentoring from the thesis author throughout study implementation. Particular benefits included peer review of goal statements for increased clarity, ensuring performance or self-ratings could not fall between levels affecting the ability to determine the level of attainment, and encouraging a focus on activity and participation-based rather than impairment-based goals, when appropriate. This may have contributed to higher administration consistency which translated to higher agreement in scores. While the psychometric properties of GAS-Light have not been previously investigated, it has been used in research studies to measure goal attainment (Ashford, Williams, Nair, Orridge, & Turner-Stokes, 2019; Turner-Stokes et al., 2016; Turner-Stokes, Fheodoroff, Jacinto, & Maisonobe, 2013; Turner-Stokes, Fheodoroff, Jacinto, Maisonobe, et al., 2013). The ULIS study (Turner-Stokes et al., 2016), for example, used a structured approach to goal setting combined with targeted standardised outcome measures (including GAS-Light) in spasticity management (Turner-Stokes et al., 2016). The findings from this study therefore build on the research utility of the GAS-Light and suggest that with training, it has clinical equivalence to the original GAS.

A number of features in this study suggest caution may be required when generalising findings, the main being study sample size. The recruitment target of n=70 participants was not met. A limitation is therefore our sample size (n=12 below target). An extension of recruitment time was not practically feasible. Given that *a*priori sample size calculations were based on estimates, an observed power analysis may provide

assurance as to whether our recruited sample size was sufficient to detect effects. The observed power for this study sample size of n=58 was 0.999 suggesting the smaller sample size provided sufficient power to undertake the test statistics (see Appendix G). Our participants were primarily post-stroke with arm and hand weakness which may have influenced the type of goals set, with a large number being in the ICF chapters of mobility and movement-related functions. The likely impact of neurological condition on goal types limits applicability to other neurological samples.

With those limitations acknowledged, these findings do provide support for the use of GAS-Light in neurorehabilitation as a measure of choice, or as an alternative to GAS. We found very little systematic difference between T-scores calculated from GAS and GAS-Light (within the 10% minimal clinical change (Khan et al., 2008)), and the calculated effect sizes show both methods were sensitive in detecting change, consistent with other studies using GAS in rehabilitation (Khan et al., 2008). We also found that the level of goal attainment determined by GAS-Light was not influenced by goal types when categorised using the ICF; for example, a goal based on reduced pain compared to a goal related to the performance of an activity such as dressing or meal preparation. Further, in our study, GAS-Light performed in a manner equivalent to GAS regardless of patient sex, cognitive ability, upper limb strength or in the presence of limb spasticity. GAS-Light therefore, may be considered a suitable choice for use with a variety of clinical presentations within upper limb neurorehabilitation for adults following ABI.

To date, the uptake of GAS within clinical practice has been limited with explanations including: the application of negative numbers and the inability to capture partial goal achievement unless baseline is set at -2 which then eliminates the potential to detect performance deterioration (Turner-Stokes, 2009; Turner-Stokes et al., 2010). Our study suggests GAS-Light addresses these implementation barriers thus providing a suitable alternative to the original GAS within neurorehabilitation. The follow-up guide does not use negative numbers and captures partial goal achievement.

6.6.1. Practice Recommendations

Further, the consistency of T-scores calculated from either tool in our study suggests GAS and GAS-Light methods may be considered interchangeable for neurorehabilitation clinical practice purposes. Interchanging the two tools within clinical research is not recommended because of the spread of the limits of agreement within our population. While each is sensitive, valid and reliable, in clinical research studies investigators should select either GAS or GAS-Light for use throughout the study (Ashford et al., 2019; Turner-Stokes, 2009).

6.7. Conclusions

The GAS-Light tool demonstrates a high level of agreement, reliability, and sensitivity to change when compared with the original GAS method in adult neurorehabilitation. Goal attainment as measured using the GAS-Light was found to be consistent with the estimation measured using the GAS, but the GAS-Light was also found to be sensitive to deterioration and partial goal achievement which is beneficial in neurorehabilitation.

6.8. Chapter Synopsis

This Chapter has presented psychometric evidence for GAS and GAS-Light when used within neurorehabilitation for upper limb motor impairment post-ABI. The following Chapter will further compare the two goal attainment scaling tools to investigate their clinical utility and whether adults post-ABI perceive the GAS-Light to be acceptable for use.

Chapter 7:

Clinical Utility and Patient Acceptance of Goal Attainment Scaling (GAS) and GAS-Light

See Appendix B for ethics approvals and Appendix H for supplementary material.

7.0. Chapter Overview

The importance of using psychometrically sound outcome measurement tools and goal setting practices within neurorehabilitation have been presented in the preceding chapters. Chapter 6 presented the results of a study comparing two GAS methods providing evidence about psychometric properties. Knowledge of psychometric evidence forms part of the clinical reasoning process used by clinicians when selecting tools for practice. The clinical utility of a tool is another component influencing this process. The clinical utility of GAS-Light is as yet unknown. This chapter presents the findings of a clinical utility study nested within the comparative study presented in Chapter 5. This study explored the clinical utility of GAS and GAS-Light from the clinicians perspective, and the acceptance of GAS-Light from the patients perspective when used in upper limb neurorehabilitation.

7.1. Background

Patient-centred goal setting and measurement of intervention outcomes are recommended components of best practice neurorehabilitation (Hebert et al., 2016; Intercollegiate Stroke Working Party, 2008; Lees, 2009; Ottawa Panel, 2006; Winstein et al., 2016). Previous chapters in this thesis have presented evidence regarding the psychometric properties of outcome measurement tools used in neurorehabilitation for people with ABI related upper limb motor impairment and its functional consequence. Whilst psychometric properties are integral to clinical reasoning to select the most suitable tool for the purpose (De Vet et al., 2011), the clinical utility and feasibility of using the tools in clinical practice is also key.

Clinical utility is a term commonly used within health care, yet there is not a universally agreed formal definition (Lesko, Zineh, & Huang, 2010; Smart, 2006). In the absence of an encompassing definition, clinical utility of outcome measurement tools can be articulated. It is described as the ability of a tool to provide clinical information to inform interventions (Donnelly & Carswell, 2002; Law et al., 1990; Van Herk, Van Dijk, Baar, Tibboel, & De Wit, 2007), along with the ease of use of the tool, acceptability of the format of the tool for clinicians and patients, and length of administration time (Law et al., 1990). Smart (2006) proposed clinical utility to be multi-dimensional and to encompass the following categories: appropriateness, accessibility, practicability, and

acceptability. The clinical utility of an outcome measurement tool can influence the clinicians' selection of a tool over psychometric evidence for the tool or adherence to best practice (Stevens, Beurskens, Koke et al, 2013).

Setting patient-centred goals remains challenging within neurorehabilitation (Plant et al., 2016; Prescott, Fleming, & Doig, 2015; Sugavanam, Mead, Bulley, Donaghy, & van Wijck, 2013). Evidence suggests the process of goal setting is often dominated by clinicians leaving ABI patients disempowered and unsatisfied with the defined goals (Plant et al., 2016; Rosewilliam et al., 2011; Sugavanam et al., 2013). Goal Attainment Scaling (GAS) (Kiresuk & Sherman, 1968), as presented in Chapter 1 and 6, is one tool that provides a structured approach to define and measure individualised patient and/or program based goal attainment (Malec, 1999). GAS has demonstrated reliability (Bovend'Eerdt et al., 2011; Joyce et al., 1994), concurrent validity (Brock et al., 2009) and responsiveness (Khan et al., 2008; Lannin, 2003) when used in neurorehabilitation. GAS supports adherence to evidence-based goal setting practices, and as such facilitates the development of neurorehabilitation programs focussed on outcomes important to patients, which also evaluates the level of goal attainment and thus permits measurement of intervention outcomes (Turner-Stokes, 2009).

A systematic review investigating the practical use of patient-specific measurement instruments for goal setting by clinicians and the patient, which included a summary of the amount of time required for administration, the instructions and the necessity of training and costs, concluded that GAS is considered useful in the goal setting process for both patients and professionals (Stevens et al., 2013). Furthermore, GAS was reported to foster teamwork and to facilitate a patient-centred approach (Stevens et al., 2013). This review also identified limitations in GAS use, particularly for elderly people and people with cognitive, communication and emotional problems, for whom, they report, identifying performance problems can be difficult, and consumed a large amount of time (Stevens et al., 2013). This review was not specific to people with ABI; evidence drawn from a neurorehabilitation subsample supports the ability of GAS to capture improvement in relevant activities more than general measures; and it can facilitate collaborative goal setting (Grant & Ponsford, 2014; Turner-Stokes, 2009). The review

also concluded however, that GAS has reduced clinical utility due to the time required to set and score (Grant & Ponsford, 2014; Plant et al., 2016; Turner-Stokes, 2009).

In response to identified limitations, Turner-Stokes (2009) developed the GAS-Light tool. GAS-Light applies a six-point verbal rating scale to determine the level of goal attainment, removing the time consuming step of setting individual levels. GAS-Light has an emerging evidence base including some published psychometric data demonstrating correlations with measures of function, impairment and carer report (Turner-Stokes, Fheodoroff, Jacinto, & Maisonobe, 2013). Findings reported in Chapter 6 showed that GAS-Light has a high level of agreement, reliability and sensitivity to change when compared to the original GAS in a neurorehabilitation population. GAS-Light is proposed to hold increased clinical utility due to being briefer with preliminary evaluation (unpublished) suggesting time savings when compared against a pre-prepared follow up guide (Turner-Stokes, 2009). As yet, the clinical utility, ABI survivor acceptance of GAS-Light, and the level of engagement and satisfaction with the goal setting process when using GAS-Light is currently unknown.

7.2. Study Aim

The primary aim of this study was to determine the perceived clinical utility of the two goal attainment scaling methods, GAS and GAS-Light, when used to determine upper limb activity goals in neurorehabilitation from clinician and ABI survivor perspectives. The secondary aim was to understand the ABI survivor's experience using GAS-Light, and to explore their acceptance of the goal setting process.

7.3. Research Questions

1. What is the clinical utility of GAS-Light when compared to GAS when used by clinicians working with ABI survivor’s undergoing neurorehabilitation for upper limb motor impairment?

2. What is the level of ABI survivor’s engagement in the goal setting process when using GAS-Light to set and evaluate upper limb neurorehabilitation goals?

3. How satisfied are ABI survivors with the goals that are set?

4. Are ABI survivors accepting of the GAS-Light method in upper limb neurorehabilitation?

7.4. Method

7.4.1. Study Design

Institutional ethical approval was obtained from the University of Wollongong/Illawarra Shoalhaven Local Health District Medical Human Research Ethics Committee (approval number HE12/077), La Trobe University Faculty Human Ethics Committee (FHEC12/152) and The Alfred Human Research Ethics Committee (approval number 38/13). Local site-specific approval was obtained from the Murrumbidgee Local Health District (MLHD) (SSA/12/MLHD/107), see Appendix B. Informed written consent was provided by all participants (ABI survivors) or their nominated person responsible.

This study was nested within a multi-site repeated measure study comparing the GAS-Light tool to the original GAS tool. The comparative study is presented in detail in Chapter six and included adults who presented to either the community rehabilitation service or metropolitan spasticity clinic. Patients presenting to these sites with an upper limb motor impairment limiting engagement in daily activities or, in the absence of disruption of functional use, moderate spasticity (as indicated by a score of 3 or greater on the Modified Ashworth Scale), and were naïve to GAS (that is, had not previously set goals using the GAS method) were invited to participate in the comparative GAS study. Patients were included if upper limb motor impairment and/or spasticity occurred as a result of an ABI due to stroke, traumatic brain injury, multiple sclerosis, or cerebral palsy. Persons under 18 years of age, highly dependent on medical care or pregnant were excluded. Following informed consent, recruited patient participants received assessment and identification of neurorehabilitation goals. The treating clinician facilitated this session and completed the GAS-Light tool, while an independent assessor completed the GAS tool. Patient participants then received usual care, which included participation in a tailored, goal-directed neurorehabilitation program. At the patient participants pre-determined follow up appointment, the treating clinician directed the session, including repeating any baseline standardised assessments and/or observing functional task performance as required, to complete the GAS-Light tool. The

independent assessor attended this appointment and completed the GAS tool independently with neither result compared nor discussed between assessors.

To investigate the clinical utility of both GAS methods, and patient acceptance of the GAS-Light method, data were collected at baseline (immediately following the goal setting session) and follow-up (at the clinically pre-determined date for goal evaluation). At baseline, patient participants completed the patient satisfaction in goal setting scale (Turner-Stokes, Rose, Ashford, & Singer, 2015) and the treating clinician and independent assessor collaboratively completed the patient engagement in goal setting scale (Turner-Stokes et al., 2015). At the follow-up session, the patient participant completed the patient acceptance scale (developed by the thesis author). Then, clinicians at each site were invited to complete GAS and GAS-Light clinical utility surveys. The surveys were anonymous and completed in hard copy and returned to the author. All responses from patient participants and clinician participants were collated in an Excel™ spreadsheet and imported into SPSS Version 25 to analyse descriptive statistics with percentages, means and standard deviations reported as relevant. There were no compulsory questions in any of the scales or surveys. Missing items are noted in the text or tables.

7.4.2. Outcome measures

Clinical utility

Clinical utility was measured through thesis author designed paper-based surveys (Clinician Clinical Utility Survey: GAS and Clinician Clinical Utility Survey: GAS-Light), see Appendix H. The surveys included categorical and open-ended questions with some rated on a 5-point Likert scale. The surveys included information regarding clinicians' years of clinical practice and experience using GAS and GAS-Light tools. The survey also enquired about the clinical utility of both tools independent of one another and in comparison to one another drawing on factors identified in the literature. Sub questions included those related to accessibility, ease of understanding, administration time, value add to clinician decision making, cost-effectiveness and responsiveness to clinically meaningful change. Time to complete the GAS and GAS light, including time taken to

develop the goals and follow up guide, but excluding the time required to administer standardised assessments or observe task performance was collected.

Patient Engagement in the Goal Setting Process

The Patient Engagement Scale (Turner-Stokes et al., 2015) was used to rate the patient participants level of engagement in the goal setting process. This scale considers cognitive, communication and behavioural factors related to goal setting and the level of support required to determine goals. This scale was completed by the treating clinician in collaboration with the independent assessor and rated engagement on a scale from 0-5, see Figure 7.1 and Appendix H. This scale has previously been used with patients with a neurological diagnosis participating in neurorehabilitation, to examine the relationship between the level of patient engagement in the goal setting process, as rated by the clinicians, with the achievement of neurorehabilitation outcomes including goal attainment (Turner-Stokes et al. 2015).

Patient Satisfaction with the Goal Setting Process

The Patient Satisfaction Screen (Turner-Stokes et al., 2015) was used to rate the patient participants level of satisfaction with the goal setting process. This scale considers how well the identified goals matched their priorities for neurorehabilitation, the extent to which they agreed with the goals, the extent of choice in goal areas and the extent to which they felt in control of the goal setting process. Patient participants self-rated their satisfaction on a scale from 0-5, see Figure 7.2 and Appendix H. This scale has previously been used with patients with a neurological diagnosis participating in neurorehabilitation, to examine the relationship between self and/or family / carer rated satisfaction with the goal planning process and achieved neurorehabilitation outcomes and goal attainment (Turner-Stokes et al. 2015).

Patient Acceptance Survey

The thesis author designed a patient acceptance survey and invited patient participants to rate their experience using GAS-Light on a five-point Likert scale. The survey included

five questions and was designed to elicit information from the patient participant regarding their comfort, the time cost, the perceived worth in assisting to determine goal achievement, and their willingness to complete GAS-Light again in the future (Appendix H).

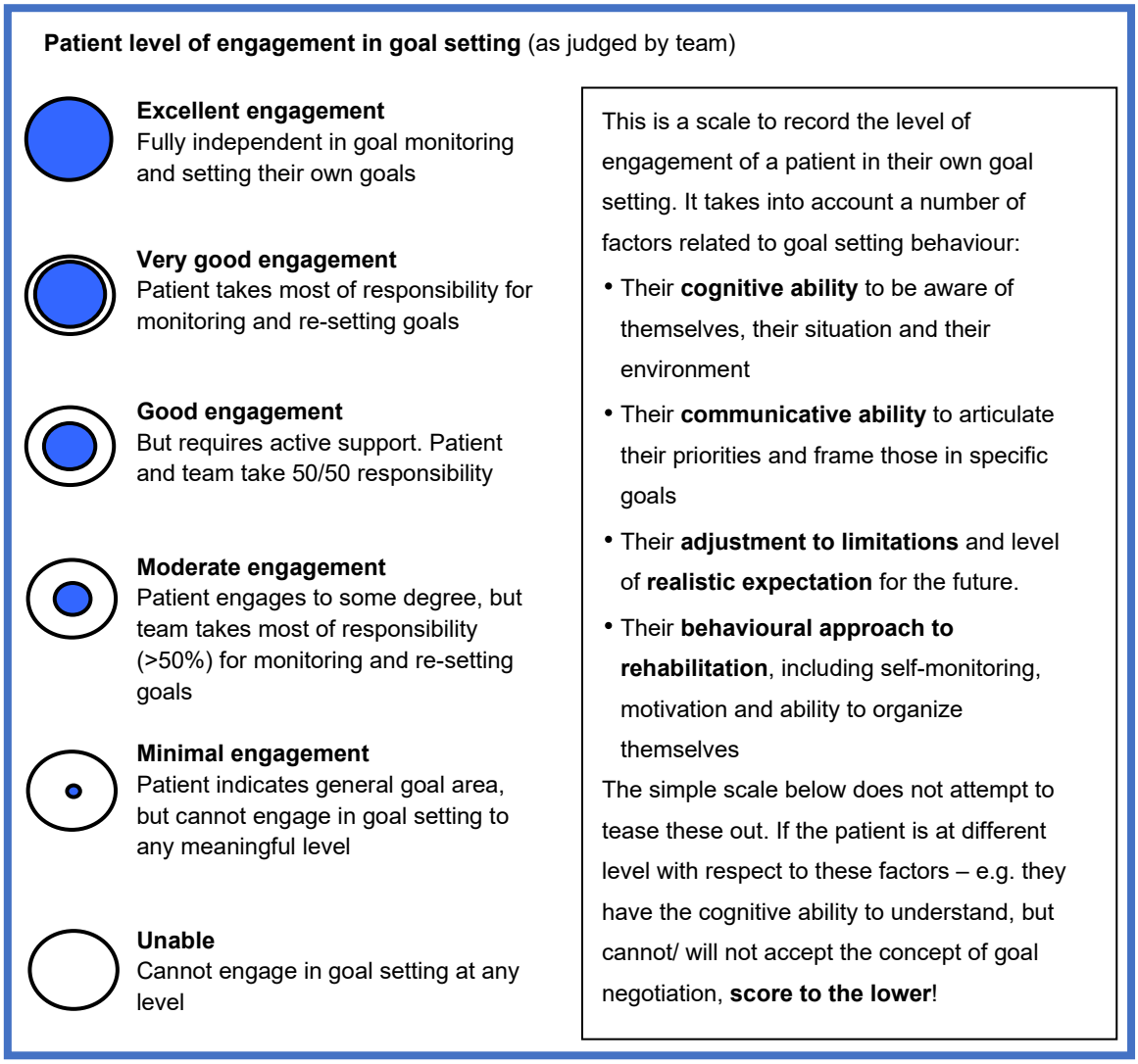
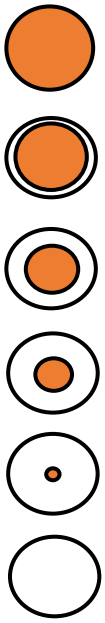


Figure 7.1. *Patient level of engagement with the goal setting process scale (Turner-Stokes et al., 2015)*

Patient satisfaction with the goal setting process (as judged by the patient/family)



Excellent
My goals matched all my key priorities for rehab and were entirely my own choice

Very good
My goals matched my main priorities for rehab and I was pretty happy with my agreed goal-set

Good
My goals met most of my priorities for rehab and I agreed with some of them

Moderate
My goals met some of my priorities for rehab and I agreed with some of them

Poor
My goals were largely irrelevant to me and I disagreed with most of them

None
My goals were completely irrelevant and I did not agree with any of them
Or
What goals??

This is a scale to record the patient's satisfaction with goal setting. It takes into account a number of factors related to goals:

- How well the goals matched their priorities for rehab
- The extent to which they agreed with the goals
- The extent of choice in goal areas
- The extent to which they felt involved with / in charge of the goal setting process

The simple scale does not attempt to tease these out. If the patient is at different level with respect to these factors – e.g. they had a wide choice of goals but did not agree with any of them, **score to the lower!**

Figure 7.2. *Patient satisfaction with the goal setting process scale (Turner-Stokes et al., 2015)*

7.5. Results

Sixty patient participants aged between 18 – 88 years (mean 59, SD 15.89) participated in the study. From these 60 patient participants, there was a response rate of 97% for the patient satisfaction screen and 87% for the patient acceptance survey. There were 31 (58%) male patient participants and 53 (88%) were stroke survivors. Nineteen clinician participants completed the clinical utility surveys; they were all the clinicians working at the hospital site. Clinician participants included occupational therapists and physiotherapists with 2-19 years of clinical experience, and ranged in their years of

experience using GAS in clinical practice. Five clinician participants had no prior experience using GAS in clinical practice, and 14 clinician participants had used the original GAS tool (range from 1 to 10 years). Eight clinician participants reported prior GAS-Light use for 1 to 2 years only. Table 7.1 presents further details regarding patient participant and clinician participant characteristics.

Table 7.1. *Patient participant and clinician participant characteristics*

Participant characteristics	Total (n=60)	Diagnosis			
		Stroke (n=53)	TBI (n=3)	CP (n=3)	MS plus Stroke (n=1)
Age *(yr)*, mean (SD)	59.17 (15.89) [18-88]	61 (13.5)	45 (12)	23 (7)	65 (0)
Sex, number male (%)	35 (58)	31 (58)	2 (67)	1 (33)	1 (100)
MMSE, number <19, (%)	8 (13)	5 (9)	0 (0)	2 (67)	1 (100)
MMSE≥19, mean (SD)	27 (3.3) [19-30]	27 (3.3)	26 (4.6)	30 (0)	0 (0)
UL spasticity, n (%)	24 (40)	17 (32)	3 (100)	3 (100)	1 (100)
Affected UL					
n right (%)	27 (45)	23 (43)	3 (100)	0 (0)	1 (100)
n left (%)	30 (50)	29 (55)	0 (0)	1 (33)	0 (0)
n both (%)	3 (5)	1 (2)	0 (0)	2 (67)	0 (0)
Baseline UL Function n (%)					
No weakness	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Mild weakness	30 (50)	27 (51)	2 (67)	1 (33)	0 (0)
Significant weakness	23 (38)	19 (36)	1 (33)	2 (67)	1 (100)
Total paralysis	7 (12)	7 (13)	0 (0)	0 (0)	0 (0)

Clinician characteristics	Total (n=19)	Occupational therapist (n=15)	Physiotherapist (n=4)
Clinical experience *(yr),* mean (SD) [range]	7.8 (4.9) [2-19]	7.8 (4.3) [2-16]	8 (7.5) [2-19]

7.5.1. Clinical Utility

The average reported time required to complete the GAS tool at baseline was 33 minutes (SD 18, median 30, range 10 to 60 minutes). The average reported time to

complete at follow up was 17 minutes (SD 17, median 10, range 5-60). The average reported time required to complete the GAS-Light tool at baseline was 22 minutes (SD 19, median 15, range 5 to 60 minutes) and 12 minutes (SD 14, median 5, range 2-40) at follow up. Clinical utility responses regarding GAS are presented within Figure 7.3 and GAS-Light within 7.4.

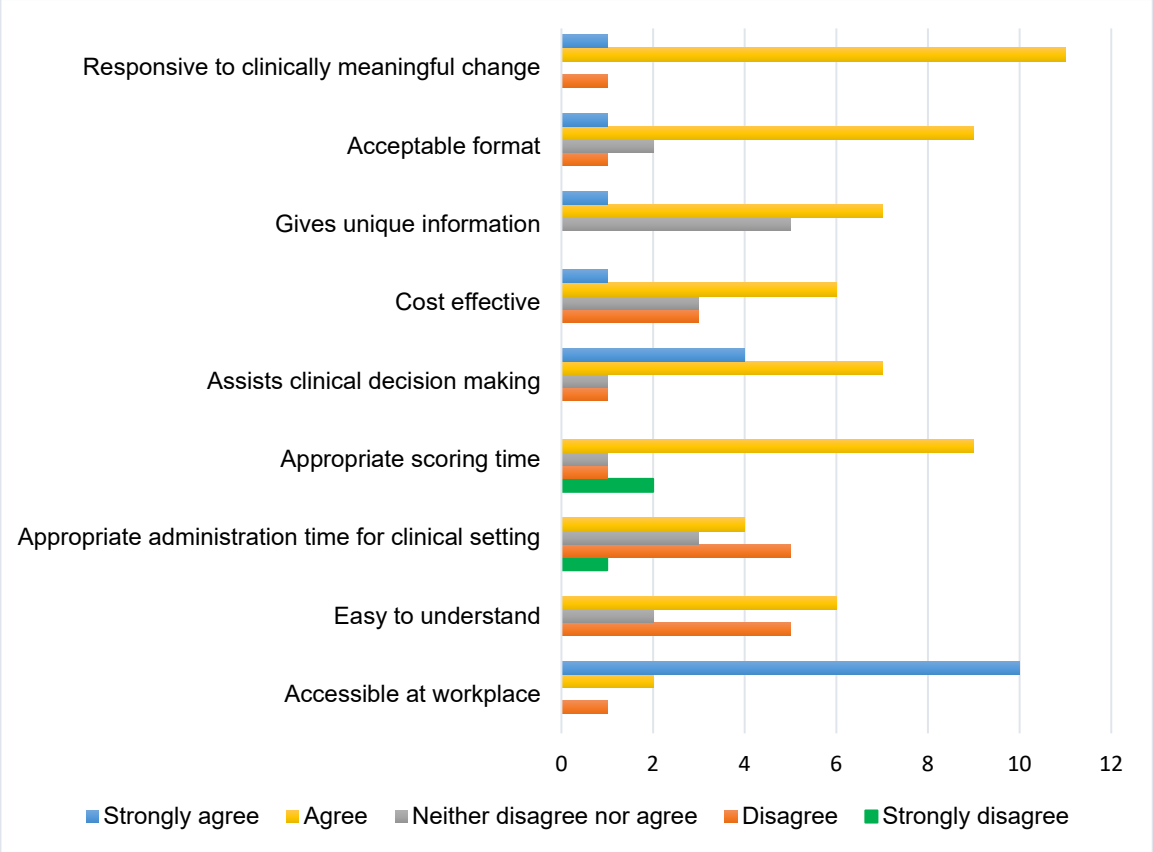


Figure 7.3. *Clinician perception of the clinical utility of GAS*

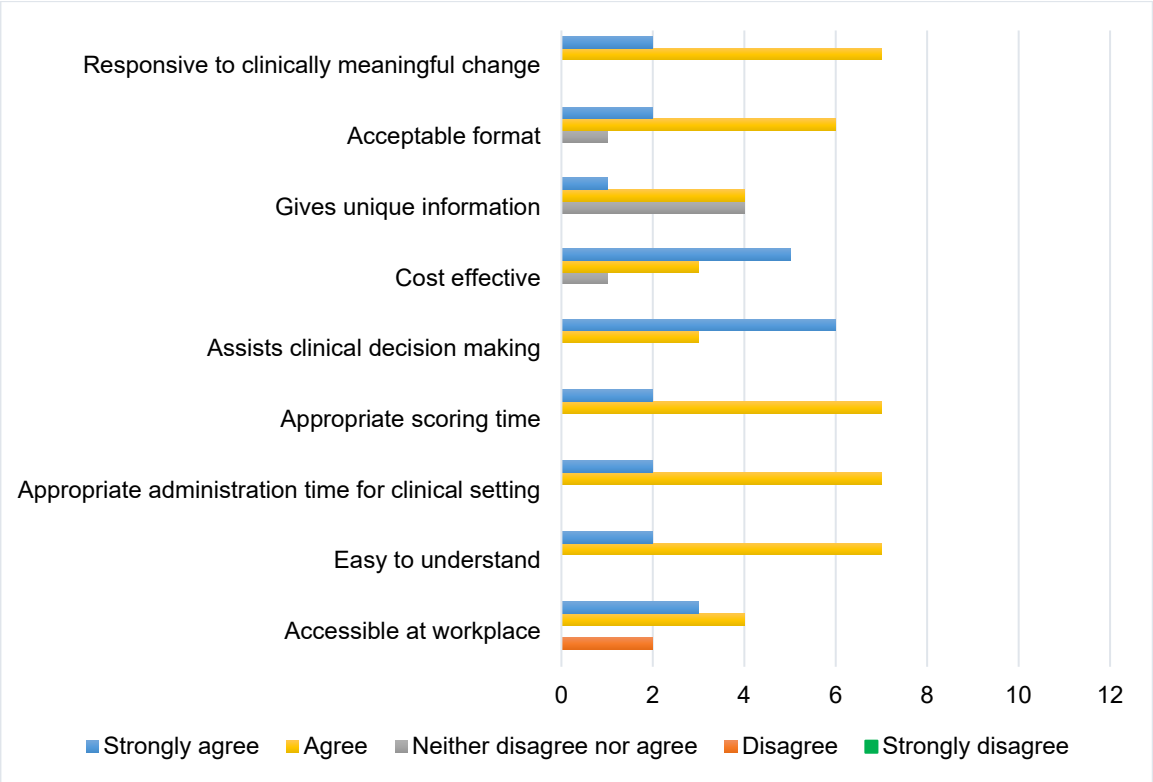


Figure 7.4. *Clinician perception of the clinical utility of GAS-Light*

Clinician participants with experience in both goal attainment scaling methods were asked to consider the clinical utility of GAS-Light compared to GAS. Data from thirteen clinician participants are presented in Figure 7.5. Clinician participants rated GAS-Light to hold greater clinical utility across three of five categories. Three (23%) clinician participants 'strongly agreed' and six (46%) 'agreed' that GAS-Light was easier to understand than GAS. Nine (69%) clinician participants 'strongly agreed' and three (23%) 'agreed' GAS-Light administration time was more appropriate and eight (62%) 'strongly agreed' and five (39%) 'agreed' the time required for scoring was more appropriate for GAS-Light. Clinician participants rated GAS and GAS-Light equivalent in assisting with clinical decision making with seven (54%) 'strongly agreeing' or 'agreeing' and six (46%) 'neither agreed nor disagreed' or 'disagreed'. Furthermore, eight (62%) clinician participants 'strongly agreed' or 'agreed' that GAS-Light was more useful and five (39%) 'neither agreed nor disagreed'.

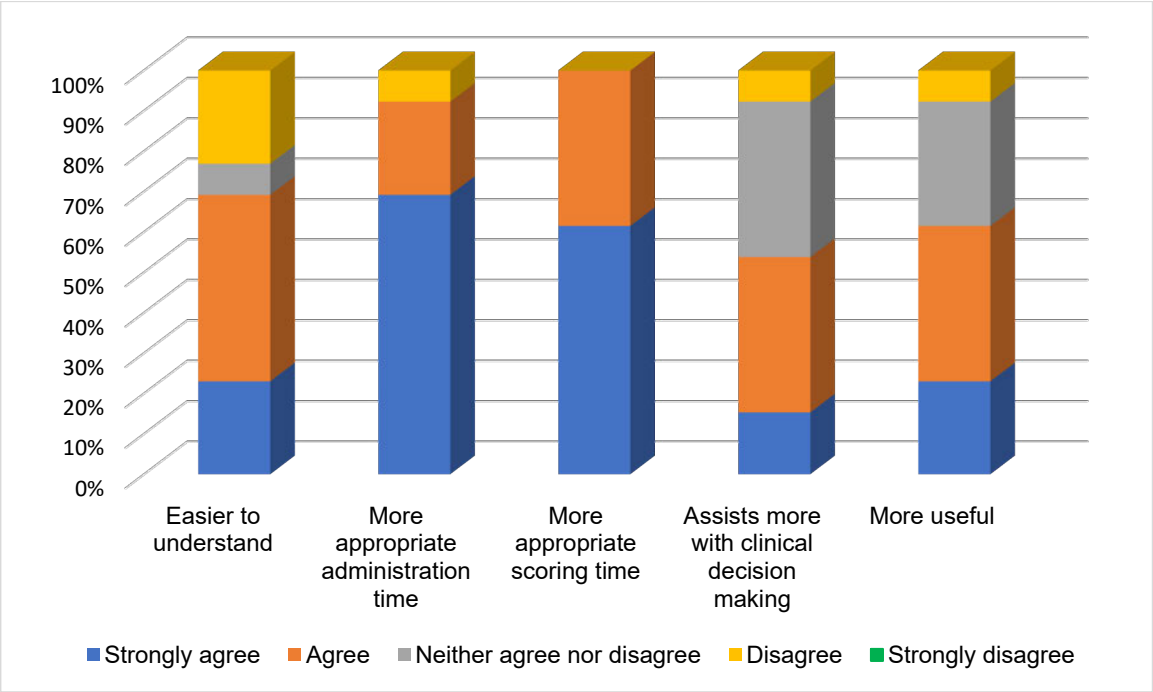


Figure 7.5. *Clinical utility of GAS-Light compared to GAS*

7.5.2. Patient Engagement in the Goal Setting Process

The level of engagement in the goal setting process was rated by the treating clinician in collaboration with the independent assessor for all 60 patient participants and was primarily 'very good' (n=19, 32%) or 'good' (n=19, 32%), see Figure 7.6. Patient participant characteristics including age, sex, level of schooling and cognition were considered for any emerging patterns on the varying levels of engagement. All patient participants who were rated to have 'minimal engagement' in the sample had completed less than high school level of schooling, see Table 7.2. No other patterns were identified through descriptive analysis.

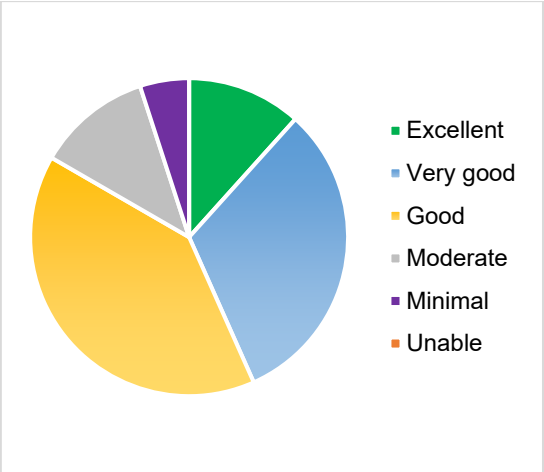


Figure 7.6. *Patient engagement in the goal setting process*

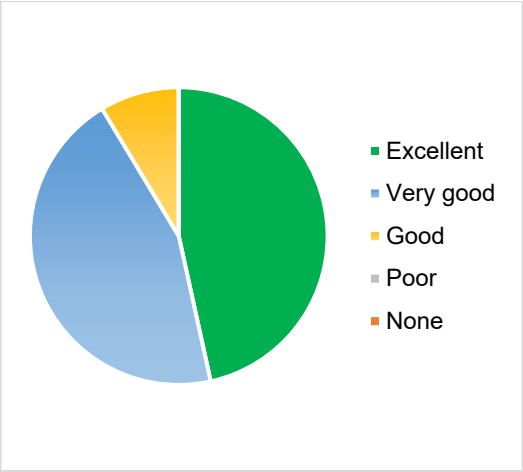


Figure 7.7*. Patient satisfaction with the setting process*

7.5.3. Patient Satisfaction in the Goal Setting Process

Data from fifty-eight patient participants who completed the screen were analysed, two patient participants had missing data. Results indicated a high level of self-rated satisfaction with the goal setting process, see Figure 7.7 Fifty-four (93%) of patient participants rated their satisfaction as 'excellent' (n=27) or 'very good' (n=27).

Table 7.2. *Goal setting engagement and satisfaction per education and cognition*

	Patient level of engagement in goal setting						Patient satisfaction with goal setting process						
	Excellent	Very good	Good	Moderate	Minimal	Unable	Excellent	Very good	Good	Moderate	Poor	None	Not completed
Number	7	19	24	7	3	0	27	26	5	0	0	0	2
Sex, male n	2	14	12	5	2	0	13	17	4	0	0	0	1
Level of schooling													
Less than high school	4	12	9	5	3	0	14	15	3	0	0	0	1
Finished high school	2	2	6	0	0	0	6	4	0	0	0	0	0
TAFE	1	4	6	2	0	0	5	6	2	0	0	0	0
University or higher	0	1	3	0	0	0	2	1	0	0	0	0	1
Cognition													
MMSE<19	2	2	2	2	0	0	4	3	1	0	0	0	0
MMSE≥27	5	14	13	1	1	0	20	12	1	0	0	0	1

MMSE: Mini Mental State Examination

7.5.4. Patient Acceptance of GAS-Light

Fifty-two patient participants completed the patient acceptance survey. Overall, patient participants reported a high level of acceptance of the GAS-Light tool, presented in Figure 7.8. Thirty-five (67%) patient participants 'strongly agreed' and 15 (28%) 'agreed' that completion of the GAS-Light tool was not an uncomfortable process. Thirty-five (67%) 'strongly agreed' and 16 (31%) 'agreed' they were comfortable to answer all questions. Patient participants perceived the GAS-Light to be a worthwhile time cost; 26 (50%) 'strongly agreed' and 25 (48%) 'agreed'. Patient participants perceived GAS-Light to provide helpful information to determine progression towards goal achievement; 23 (%) 'strongly agree', 25 (48%) 'agree', three (6%) 'neither agree nor disagree' and one (2%) 'strongly disagreed' noting that it was the doctors who make the decision. All patient participants would be happy to complete GAS-Light again if needed; 28 (54%) 'strongly agree' and 24 (46%) 'agree'.

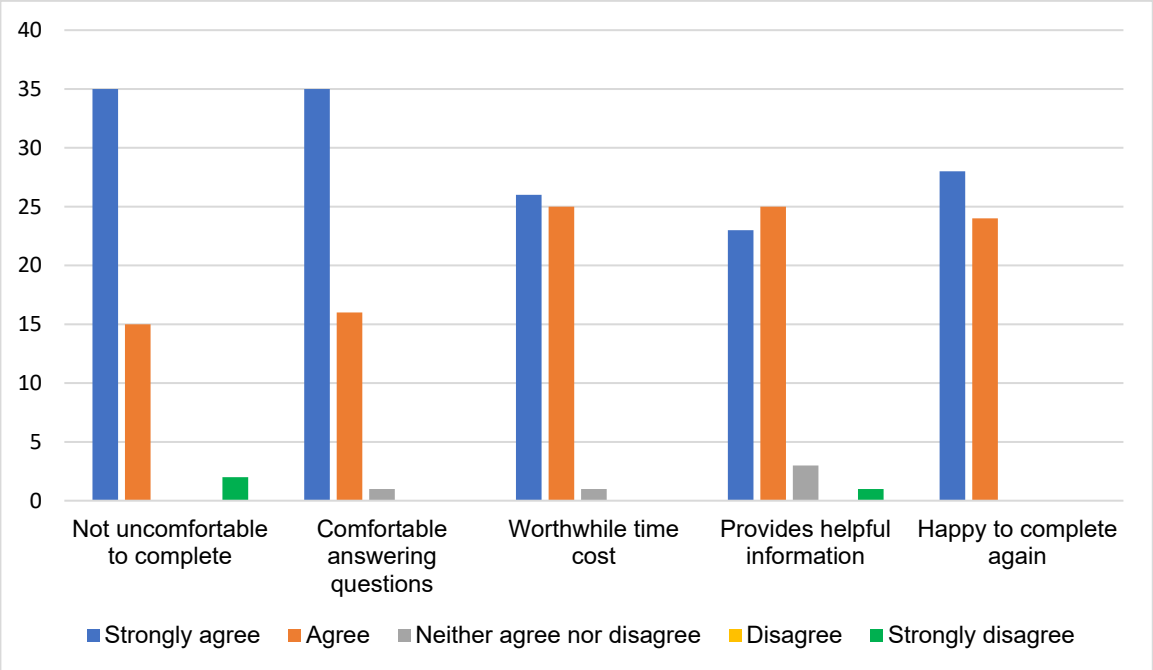


Figure 7.8. *GAS-Light patient acceptance survey results*

7.6. Discussion

The results of this study reveal that occupational therapists and physiotherapists perceive GAS-Light to possess stronger clinical utility properties than the original GAS

method, particularly with respect to the properties of time for administration and scoring, and ease of understanding. Furthermore, this study found that ABI survivors undergoing upper limb neurorehabilitation perceived the GAS-Light as “acceptable”, “comfortable”, “worthwhile” and “helpful”, with results indicating a high level of engagement and satisfaction with the goal setting process.

The results of this study support existing evidence for the original GAS (Khan et al., 2008; Lannin, 2003; Malec, 1999; Stevens et al., 2013; Turner-Stokes, 2009) and emerging evidence for GAS-Light (Turner-Stokes, 2009), suggesting both tools provide useful information that assists in clinical decision making. The time cost and perceived burden, a recurrent reported theme with GAS use (Grant & Ponsford, 2014; Plant et al., 2016; Turner-Stokes, 2009), was also reflected within the results of this study. The time required to complete the GAS-Light compared to GAS was found to be faster at both baseline (mean 12 minutes, median 15 minutes) and at follow up (mean 5 minutes, median 5 minutes). Consistent with findings reported by the GAS-Light developers (Turner-Stokes, 2009), clinicians in this study supported the proposition that GAS-Light requires less time for both administration, and scoring. The ability to derive information that can assist with clinical decision-making and determine goal attainment more efficiently, without sacrificing robustness or detracting the survivor's engagement in the process, is critical within the often time-poor clinical setting.

The benefits, challenges, barriers and facilitators to goal setting and the influence of patient-centred goal setting on achieved outcomes has previously been investigated (Plant et al., 2016; Rosewilliam et al., 2011; Sugavanam et al., 2013). Engaging patients in the process of goal setting is a frequently cited challenge, with patients often reporting goals do not match their priorities. The ‘very good’ to ‘good’ level of engagement in goal setting, and ‘excellent’ to ‘very good’ level of self-rated satisfaction in this study suggests that identified goals match patient key priorities for participants in this study. Engagement and satisfaction ratings of goal setting in a mixed rehabilitation sample has been reported to be correlated with patients’ level of cognitive function (Turner-Stokes et al., 2015). This relationship was not apparent in this study however this study sample

did not have sufficient representation of adults with impaired cognitive function. This assertion requires further testing in a deliberatively selected sample.

This study recruited a small sample and participant were primarily adults post-stroke, therefore caution should be applied when drawing conclusions about relationships, or generalising the findings outside of upper limb neurorehabilitation post-stroke. Further research to better understand the implications of a cognitive impairment, in addition to the clinical utility and acceptance of GAS-Light outside of upper limb neurorehabilitation is still required. Also, this study collected data via surveys only. Future research could employ qualitative and semi-structured interview methods to explore the clinician and ABI survivors experience using GAS and GAS-Light which would extend the evidence generated from this study. Furthermore, exploring factors that inform and behaviours that influence clinical decision making regarding the selection of either GAS method to measure upper limb outcomes in neurorehabilitation is recommended.

7.7. Conclusions

GAS and GAS-Light were found to be valued tools considered by clinicians to provide useful information to assist in clinical decision making, particularly regarding individualised outcomes. Findings suggest GAS-Light is accepted by ABI survivors and requires less time for administration and scoring than GAS. The enhanced clinical utility of GAS-Light make it a suitable tool for use within clinical practice and research to set, document and evaluate the level of attainment of upper limb goals in neurorehabilitation. Further research is required with a larger sample including wider representation of diagnoses causing upper limb impairment and varying cognitive abilities. Furthermore whilst this study gathered useful information, limited data was collected and future qualitative enquiry may provide further insights.

7.8. Chapter Synopsis

This study explored clinicians and ABI survivors experience of using GAS and GAS-Light within upper limb neurorehabilitation. This Chapter has provided evidence for the clinical utility of GAS and GAS-Light, the experience and acceptance of this tool by ABI survivors. Consideration of the evidence presented in this Chapter complimented by that

provided in Chapter 6 contributes to building the developing evidence base for GAS-Light and extend existing knowledge for GAS to inform neurorehabilitation practice and research. Chapter 8 will draw together the five studies presented in this thesis and provide recommendations for clinical practice, education and future research.

Chapter 8:

Discussion

8.0. Introduction

This PhD program of studies was undertaken to examine two key requisites of best-practice neurorehabilitation: precise measurement of upper limb activity, and setting participation goals. The qualitative study presented in Chapter 2a sought to understand first, the impact of upper limb motor impairment on everyday activity from the perspective of adults living with spasticity after ABI. The thematic analysis identified the lived experience of post-stroke upper limb spasticity was one of continual adaptation and adjustment, using distinct processes in a variety of contexts. Chapter 2b explored stroke survivors' perspectives as linked to the ICF to reveal the scope of impact of upper limb spasticity on body function, activity and participation. The ability of the ICF Comprehensive and Brief Core Sets for stroke to capture the breadth of impact of upper limb spasticity was explored. The precision of outcome measurement tools used in the measurement of function in everyday for ABI populations was undertaken next; Chapters 3, 4 and 5 present the systematic review method and findings from the synthesis of existing evidence. This systematic review highlighted the need to investigate the measurement properties of available tools specific to upper limb performance following ABI. GAS-Light and original GAS were then compared and GAS-Light was shown to be equivalent to the original GAS with respect to reliability, validity and sensitivity as described in Chapter 6. Furthermore GAS-Light was reported to be acceptable for use by patients, and when clinical utility properties of GAS-Light were compared to the original GAS, clinicians reported GAS-Light to require less time for administration and scoring and was easier to understand as described in Chapter 7. This final Chapter 8 aims to incorporate and discuss the key findings from each of the studies completed as part of this PhD thesis, and discuss the implications for clinical practice, education and future research. The contribution of this series of studies to extending knowledge in the area of upper limb outcome measurement in neurorehabilitation is also discussed, along with the strengths and limitations of the program of research, together this will afford readers the opportunity to independently interpret the findings in context.

8.1. Key Findings

This program of studies has identified the individual (Chapter 2a) and broad impact (Chapter 2b) of acquiring an upper limb motor impairment following ABI on activity

outcomes and participation more broadly. While there are commonalities, there are unique differences supporting the value of individualised outcome measurement, and the view that outcome measurement cannot be solely focussed on impairments such as range of movement or strength. The interviewed survivors reminded clinicians they must also address activities that are important to people with upper limb motor impairment following ABI, and that activity and participation goals are frequent and can be measured in individualised outcomes. The study findings suggest neurorehabilitation teams gain an understanding of the person's participation goals, such as using the arm in domestic roles or roles such as being a friend through goal setting processes such as GAS and GAS-Light. The applicability of the ICF as a framework capable of capturing and describing the acquired disability arising from upper limb spasticity post-ABI is supported by the original evidence generated within this research program. Further, the synthesis of existing psychometric evidence, which formed a substantial part of this PhD program, and the generation of new psychometric evidence together provide information that can be used by clinicians in their clinical reasoning and tool selection decisions. Together, findings show that further research on the existing outcome measurement tools is still needed. This program of studies contributes novel insights into the experience of living with post-stroke upper limb spasticity, novel evidence for the psychometric properties of GAS-Light and synthesised evidence for the psychometric properties of tools to measure upper limb activity, thus providing clinicians with information they can use to guide their measurement practices in upper limb neurorehabilitation.

8.1.1. Major Findings and their Significance

At the commencement of this program of studies, exploring the experience of adults living with upper limb spasticity after their ABI was a recognized knowledge gap (Esquenazi, 2011; Esquenazi, Jacinto, & Lysandropoulos, 2020; Kerstens et al., 2020). Study 1 (Chapter 2a), a qualitative enquiry into the experience of living with upper limb post-stroke spasticity, was designed to address this identified gap. Findings from this first study revealed that stroke survivors' experience was a continuous process of adaptation and adjustment framed by two organising concepts – lived experience in *context* and *processes*. Stroke survivors described continuous adaptation to new contexts of their body, in time, their life situation and accessing health, community and

home care services using processes of expecting, learning, practicing, evaluating, listing and committing. These dynamic experiences, fuelled by hope, complimented current evidence regarding the lived experience post-stroke with spasticity (Kerstens et al., 2020) without spasticity (Arntzen et al., 2015; Barker & Brauer, 2005; Faircloth et al., 2004; Lloyd et al., 2018; Lou et al., 2017; Luker et al., 2015; Manning et al., 2019; Pallesen, 2014; Purton et al., 2020; Shipley et al., 2018), and with spasticity from any cause (Barnes et al., 2017; Bhimani et al., 2012). Whilst similarities with these past studies were identified, experiences that were unique to this sample also emerged.

The experience of living with upper limb spasticity was then considered through the lens of the ICF. The applicability of the Brief, Comprehensive and Extended Core Sets for stroke were explored, their ability to capture reported areas of impact and therefore, indicate the scope of potential impact of post-stroke spasticity on daily life that can inform outcome measurement. Study 2 (Chapter 2b) found that most adults highlighted the importance of the Body Functions domains of energy and drive, and the Activity and Participation domain of hand and arm use. Findings provided novel evidence to support the applicability of the ICF Comprehensive Core Set for stroke and showed its capacity to capture the experience of living with chronic post-stroke upper limb spasticity and impact on everyday life. Findings indicated an additional eight categories outside of Comprehensive Core Set categories were required to capture the reported experience fully. The Brief Core Set for stroke, however, failed to adequately capture the breadth of impact and experience of upper limb spasticity for this sample. Such findings echoed that of Tavener et al. (2015) who also found the Brief Core Set to be inadequate when describing the everyday experience for a different post-stroke cohort. Furthermore, the extended version did not offer further advantage than the Comprehensive Core Set.

The systematic review (presented across Chapters 3, 4, 5) drew together the available psychometric evidence for outcome measurement tools commonly used within upper limb neurorehabilitation. While novel, as it applied the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) method to synthesise the psychometric properties of these tools for the first time, overall conclusions of this review reflected findings of previous systematic reviews (Alt Murphy et al., 2015;

Ashford et al., 2008; Ashford & Turner-Stokes, 2013). Similar to these earlier reviews, the systematic review presented in Chapters 3, 4, 5 was unable to recommend one single outcome measurement tool to measure upper limb activity and participation without evidence restrictions; all tools had evidence gaps and no tool met all Consensus Based COSMIN criteria for "good measurement" properties. The review also identified part tool use, use of tools modified from their validated form, and use of non-validated language translations, all of which make the selection and use of tools challenging for clinicians. Despite these limitations, the findings provide clinicians with an evidence base previously lacking for outcome measurement tool selection. The FIM™ had the highest quality evidence for all tools thus supporting its validity and reliability. However, as a global measure of activity, the FIM suffered from both floor and ceiling effects. When selecting a tool to specifically target the evaluation of upper limb activity, the systematic review recommends the use of the ArmA and the DAS, patient-reported measures of upper limb activity, and the UL-MAS and ARAT, performance-based measures of upper limb activity. The SIS is recommended when seeking to capture upper limb performance and activity within global evaluation of health-related quality of life. If upper limb spasticity is present, the EQ-5D and SA-SIP should then be considered as they were the only two global health-related quality of life measures with evidence for construct validity for participants with upper limb spasticity.

In neurorehabilitation, client-centred goal setting is important (Hebert et al., 2016; Levack, Dean, McPherson, & Siegert, 2014; Stroke Foundation, 2020), although as discussed in Chapter 1, there is a lack of consensus on which goal setting approach to use (Levack et al., 2015; Playford et al., 2009; Prescott et al., 2015; Stevens et al., 2013). Furthermore, clinicians report challenges in routinely adopting and measuring achievement of goals. One tool to overcome these challenges is Goal Attainment Scaling (GAS) (Turner-Stokes, 2009). Study 4 (Chapter 6) and Study 5 (Chapter 7) has compared two goal attainment scaling methods to provide clinicians with new psychometric and clinical utility evidence to help inform their practice choices. GAS-Light, was compared to the original GAS within the practise context of upper limb goals for neurorehabilitation. Findings suggest that GAS-Light was equally valid and as sensitive to detecting meaningful change as the original GAS when determining upper limb goals. The high

level of agreement between GAS and GAS-Light permits interchanging tools in clinical practice such that one could compare levels of goal attainment across services using either tool. It is recommended that only one tool be used consistently for research due to the determined limits of agreement. Clinical utility of tools also influences tool selection (Greenhalgh et al., 1998; Smart, 2006; Stevens et al., 2013), and thus Chapter 7 (Study 5) compared utility across the two tools, GAS and GAS-Light in the upper limb neurorehabilitation context. Clinicians reported that both GAS and GAS-Light were considered clinically useful tools and assisted with clinical decisions, but the GAS-Light required less time to administer and score. Patients who suffered from an ABI were also accepting of GAS-Light, reported by clinicians as being able to fully engage in the goal setting process and patients themselves were satisfied with the goals set when using the GAS-Light tool.

Together, findings from this PhD program of studies suggested that adults with upper limb spasticity and activity limitations are hopeful yet reserved in their expectations of outcomes and demonstrate an ability to set goals so that they may adapt to the dynamic experience of life after ABI. Despite the identified gaps in the available psychometric evidence, neurorehabilitation clinicians can draw on the evidence provided to select appropriate outcome measurement tools to measure impact and determine intervention effectiveness. Clinicians can consider use of either the GAS-Light or GAS to identify and measure the attainment of individualised goals as both demonstrated equivalent reliability and validity, however clinical utility properties did differ which will likely ultimately determine clinician choice.

8.2. Methodological Strengths and Limitations

A strength of this PhD program was the use of a range of research methods to address clear research questions. These research questions dictated the methodological approaches chosen. Qualitative research describes, interprets and constructs meaning through the eyes of the researcher or research participant (Gerber, 1999, p. 18). Prior to this program of research, there was an acknowledged lack of understanding of the experience of living with upper limb spasticity from the perspective of a stroke survivor (Esquenazi, 2011; Esquenazi et al., 2020; Kerstens et al., 2020). Thus a qualitative

approach offered insights into what it is like to live with upper limb spasticity. Both studies 1 (Chapter 2a) and 2 (Chapter 2b) interrogated the same transcripts but applied different conceptual approaches and data analysis methods. Delimitations in the design of these studies included completing only one interview with each participant, and not member checking with participants on the findings. Given the brevity of some of the transcripts and the post-program context of interviews, it was accepted that construction of a rich and contextualised description of experience would not be possible. Therefore, analysis using content analysis independent of any framework aimed to identify themes characteristic of participant experience rather than trying to construe relationships between concept categories and meaning units. Preliminary results for Study 1 (Chapter 2a) were presented to the qualitative interviewer to gain their perspective and level of agreement to enhance trustworthiness and authenticity of these inductive findings. The ICF was identified as an appropriate framework to commence conceptual prospective deductive coding in Study 2 (Chapter 2b). Application of established linking rules (Cieza et al., 2019; Cieza et al., 2005) guided the interrogation of the transcripts and increased transparency.

The systematic review method applied in Study 3 (Chapter 3, 4, 5), including protocol development, registration and publication, and provided an opportunity to gain an in-depth overview of the psychometric properties of available outcome measurement tools to measure the constructs highlighted in the qualitative research. Applying the “careful preparation, planning, organisation and critical evaluation” (Axford, Minichiello, Coulson, & O'Brien, 1999, p. 4) inherent within the systematic review method enhanced the research rigour underpinning the conclusions about each measurement tool. Adhering to guidelines for registered protocols (Stewart, Moher, & Shekelle, 2012), including capturing and documenting reasoning for changes made to the study design post-registration, created an open and accessible audit trail. The systematic review adopted a sensitive, validated search strategy (Mokkink, Prinsen, Bouter, de Vet, & Terwee, 2016) to address known challenges to identifying relevant studies (Terwee et al., 2009) highlighted in Chapter 3. The observed lack of critical appraisal of the measurement tools to date was addressed within Study 3. Failure to critically appraise

the methodological quality of the individual studies included in systematic reviews was evident in other reviews (Alt Murphy et al., 2015) and is thus a strength of this research.

While methods chosen for Studies 1-4 were clear strengths, the use of a survey method to collect data in Study 5 is an acknowledged limitation. The method was chosen because a survey design allowed insights and exploration of clinical utility factors across the two GAS tools to be equally performed. For patient participants their perspective of acceptance of tools could be efficiently collected through a self report survey, reducing participant burden, and clinician ratings of patient participants engagement could be efficiently collected. Whilst surveys were disseminated to clinician participants for independent completion and collected in bulk to increase anonymity, there was a risk of researcher influence since the PhD investigator was present at one of the sites. This may have led to clinician responder bias. Self-report of time required to complete either GAS tool was used since comparative brevity is one of the claims of GAS-Light developers, a stronger data collection method could have been used such as a time and motion study design recording actual tool use; this approach may trace variations to reported timeframes (Baldwin, 2000). Notwithstanding these data collection limitations in Study 5, nuanced data was collected that reveals perceptions across the samples and topics worthy of further qualitative investigation.

Small sample sizes are common in neurorehabilitation studies (Stinear, Ackerley, & Byblow, 2013; Verbeek et al., 2014). This is also a limitation across this PhD series of studies, in particular Study 3 and 4. It is plausible that the challenge faced in recruiting larger samples may reveal suffering spasticity alongside a motor impairment is not as common as clinically perceived. It is also plausible, however, that recruitment challenges are a feature of these patients being already overwhelmed in dealing with their neurorehabilitation commitments, and thus, declining research participation. This was evident in Study 3, where individual study sample sizes were commonly small (less than n=100 in over half of the included studies) and has been highlighted by other reviews of functional measurement tools (Dobson et al., 2012; Wales, Clemson, Lannin, & Cameron, 2016). Recommendations made by Dobson et al. (2012) to account for small sample

sizes and to avoid reduced levels of evidence quality due to individual study sample size were adopted to overcome this limitation.

Recruitment challenges were evident in recruiting the sample for Study 4 (Chapter 6), the psychometric study comparing the GAS-Light to the GAS. An *a*priori power calculation suggested a sample size of 70 participants would be required to detect meaningful effects. Despite a lengthy recruitment period of five years and expanding the inclusion criteria and number of recruiting sites to increase recruitment rates, only n=60 adults were recruited. Despite this slow recruitment, a strength of this study was that there were no dropouts (i.e. nil attrition), and statistically significant results were found – i.e. results were not likely to occur by chance but instead likely to be attributable to the experiment. Given that *a*priori sample size calculations were based on estimates, an observed power analysis may provide assurance as to whether our recruited sample size was sufficient to detect effects reported in Chapter 6 (Study 4). The observed power for Study 4 (n=58) was 0.999 suggesting the smaller sample size provided sufficient power to undertake the test statistics (see Appendix G). Data collected within Study 4 (Chapter 6) was embedded within routine clinical practice. Numerous factors were considered and examined with field experts during the design phase to eliminate bias and contamination of the independent rating of goal attainment, and omitted or varied information provided via patient self-report to inform goal setting, importance ratings and goal attainment. Furthermore, consideration of the effect of repeat administration of performance-based measures that may form part of the assessment process could yield differing levels of performances due to fatigue or practice benefit. The final decision reflected in the protocol for both researcher and clinician to be present at the same goal setting and evaluation session was elected to reduce the identified risk with protocol adherence minimising the opportunity for biased results.

8.3. Recommendations and Implications

Recommendations and implications of the findings from this program of research are now discussed. First, recommendations for clinical practice are presented. Second, implications for education, and finally implications for future research will be proposed so as to move the field forwards as a result of this program of studies

8.3.1. Implications for Clinical Practice and Policy

Psychometric evaluation of measurement tools included in the systematic review provided relevant information for neurorehabilitation clinicians and adds to the evidence base for upper limb functional outcome measurement tools for patients undergoing neurorehabilitation. When seeking to measure the personalised experience of people with upper limb impairments after ABI and understand the impact of these, the activity-level measurement tool the ArmA has the greatest potential followed by the ARAT, DAS and UL-MAS. That said, while those tools demonstrated acceptable reliability and validity, a barrier remains in that the DAS and UL-MAS have not undergone testing for responsiveness or measurement error. Furthermore, as the testing of the UL-MAS did not include participants with post-ABI upper limb spasticity, evidence for use with that specific cohort of adults is unknown.

This series of studies provide clinicians with novel information to inform their goal setting practices and measurement of individualised upper limb goals. GAS-Light and GAS demonstrated equivalence in determining goal achievement within clinical practice. Deciding which tool to use will therefore rely on clinician or service preference. When presented with options, tool selection is likely to be influenced by the clinical utility of the tools (Smart, 2006; Stevens et al., 2013). Both GAS-Light and GAS provided clinically useful information but varied in the time required to administer and score goal achievement. Clinicians can consider this variation in the time required when selecting a goal attainment scaling tool.

Implementing evidence to achieve practice change is challenging (McCluskey & O'Connor, 2017). This program of studies provides psychometric evidence and includes a table of tools allowing clinicians to choose a tool that meets their clinical purpose. The study findings thus reduce the time taken to translate evidence into practice, however this thesis alone is unlikely to be sufficient. Strategies to translate evidence into practice and increase the routine use of outcome measurement tools outside of this study have used an educational approach, audit and feedback (Colquhoun et al., 2017; Van Peppen, Schuurmans, Stutterheim, Lindeman, & Van Meeteren, 2009), and clinical practice guidelines (Alt Murphy, Björkdahl, Forsberg-Wärleby, & Persson, 2021). Strategies

adopted to translate findings from this program of studies included an educational approach via conference presentations, publishing in national and international journals and contributing to consensus-based recommendations (Kwakkel et al., 2017). Published interventions to increase outcome measurement tool use have targeted the individual clinician only (Colquhoun et al., 2017) despite evidence that poor tool use is multifactorial and occurs at an individual, team and organisational level (Duncan & Murray, 2012). Following completion of Study 3 (Chapter 3, 4, 5), Study 4 (Chapter 6) and Study 5 (Chapter 7), local participating sites translated findings into outcome measurement and goal attainment scaling practices across the acute, subacute and community-based neurorehabilitation continuum. Education was provided and measurement processes created at a clinician and service level. This involved the introduction of consistent psychometrically sound outcome measurement practices post-stroke, repeat administration of a selected number of tools increasing objective measurement of change to inform the neurorehabilitation program. Secondly, GAS-Light was introduced to replace the GAS as the routinely used tool to set, document and evaluate goal achievement.

8.3.2. Implications for Education

This program of research reminds both clinicians and academics of the importance of listening to people living with the effects of an ABI. Educational opportunities offered should therefore include lived experiences and perspectives, such as the experience of living with upper limb motor impairment and participation in outcome measurement. Those educational insights should also be delivered via the voice of the survivor to retain authenticity, given the identified mismatch in language used between stroke survivors and clinicians discussed in Chapter 2a (Study 1).

Entry level programs for neurorehabilitation clinicians focus on the use of theories to guide practice. Findings outlined in Chapter 2b support the theoretical value and application of the ICF within neurorehabilitation and should therefore be shared during education. Both the complete framework and Comprehensive Core Set demonstrated the ability to characterise the scope of the impact of upper limb spasticity and other motor impairments post-stroke on everyday life for the survivor. Knowledge of the ICF

theoretical foundations and structure, and how it can be applied to recognise the dynamic interplay between Body Functions, Body Structures, Activity and Participation, Environmental and Personal Factors may enhance holistic clinical approaches.

Understanding the interplay between ICF domains may also facilitate the selection of the most appropriate outcome measurement tool or combination of tools to meet the stroke survivors' personal goals. Checking back on which ICF domain the chosen outcome measurement tool targets may assist in achieving an appropriate range of measurement. For example impairments in muscle strength, which impact on upper limb performance (Ada et al., 2006; Richard W Bohannon et al., 1991; Boissy et al., 1999; Burridge et al., 2009; Harris & Eng, 2007), and individualised activity and participation goals (Dewey, Sherry, & Collier, 2007; Royal College of Physicians et al., 2018) identified by the survivor.

It is acknowledged that there is a difference in the value placed on standardised outcome measurement and knowledge of psychometric properties (Duncan & Murray, 2012). Gaps in psychometric evidence, low methodological quality of many psychometric studies, and evidence of poor adherence to the standardised use of outcome measurement tools were identified during conduct of the systematic review. The systematic review provides current best available evidence for use, however, the call for additional validation of existing outcome measurement tools (Alt Murphy et al., 2015) will most likely see available evidence evolve and new tools released. A focus on psychometric properties should continue with undergraduate students and clinicians developing evidence based practice competencies to locate and critically appraise literature to integrate the best available evidence with their clinical experience, patient values and professional practice context (Hoffmann, Bennett, & Del Mar, 2017). Skills in translating evidence may inform measurement decisions including ceasing inappropriate or invalid measurement practices and introducing new measurement practices. Furthermore, use of psychometric evidence may support clinicians to overcome reported barriers (Duncan & Murray, 2012) and advocate for resources and managerial and organisational support to acquire outcome measurement tools and training in tool use.

8.3.3. Implications for Future Research

The research that has been undertaken for this PhD program has highlighted a number of topics on which further research would be beneficial. Several areas where information is lacking were highlighted in both Chapter 1 and in the systematic review (Chapters 3, 4, 5). Whilst some of these were addressed by the research in this thesis, others remain. In particular the lack of available psychometric evidence for tools that claim to measure upper limb outcomes and goal attainment. Further, whether adults living with an acquired upper limb motor impairment perceive tools to measure areas of most concern to them. In addition, further research is required regarding factors and processes in the clinician's tool selection decisions.

The systematic review identified a lack of published studies providing evidence for the psychometric properties of established outcome measurement tools. Of those published, significant variation in the methodological quality and reporting of studies was observed. There are several areas for further development, and applications for, the work undertaken in this PhD Program. In particular, the approaches refined in conducting the systematic review of measurement tools could be usefully applied in the evaluation of those tools beyond functional assessment measures in neurorehabilitation. Ensuring that measurement tools used in clinical practice have been selected on the basis of the systematic review of psychometric studies (rated using the COSMIN method) would provide a level of evidence not currently expected from researchers or tool developers. Furthermore research quality will likely improve due to the clear analysis and reporting guidelines offered by COSMIN (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2012; Terwee et al., 2018; Zaki et al., 2013). To ensure dissemination of evidence, pooling and comparison of results, scientific journals and outcome tool publishers require such reporting standards to publish such research.

The perspectives of adults living with acquired upper limb motor impairment is missing from the evidence base. This program of studies has contributed to filling this gap, revealing how adults with post-stroke upper limb spasticity experience life, participate in goal setting, and their views on the GAS-Light. An understanding of how adults with acquired upper limb motor impairment experience the process of measuring motor

impairment and activity outcomes, their perspectives on both the tools used and whether their findings are helpful and can inform goal-directed neurorehabilitation is still lacking. Further exploration of outcome measurement from the perspective of the ABI survivor is recommended. Such knowledge may extend outcome measurement practices and enhance individualised measurement within upper limb neurorehabilitation.

Neurorehabilitation suffers from poor uptake and low implementation of existing outcome measurement evidence into clinical practice which leads to variations in practice. It is not possible to effectively address this challenge without understanding the behaviours (outcome measurement selection) in the context in which they occur (i.e. from within clinical settings) (Atkins et al., 2017). Barriers and facilitators to outcome measurement use have been identified (Burton et al., 2013; Colquhoun et al., 2017; Duncan & Murray, 2012). One is clinical utility, and this study has contributed evidence on the clinical utility of goal attainment scaling tools explored in this program of studies. However, the factors influencing clinician's selection of measures of upper limb activity and goal attainment in neurorehabilitation based on a theoretical framework, to the knowledge of the author, is yet to be explored and is thus recommended for future research.

8.4. Conclusions

This program of studies sought to understand the measurement of upper limb activity and goal attainment within neurorehabilitation. Specifically, through a series of five studies, research produced key data on the ability to achieve individualized and precise measurement of upper limb activity.

The measurement of outcomes and patient-centred goal setting are recognized as integral components of best practice neurorehabilitation. How to best adhere to those recommendations are a challenge for clinicians when evidence to guide practice is lacking. This program of studies targeted four areas within neurorehabilitation for upper limb motor impairments.

This program of studies first took a step back to listen to stroke survivors and learn, from their perspective, what it is like to live with upper limb spasticity (Study 1). This initial study was of importance due to the surprising lack of evidence available from the perspective of the survivor living with the impairment. Limited knowledge of how survivors identify spasticity and other motor impairments to impact on the ability to use their upper limb in everyday activity may likely restrict the individualised and precise nature of current and future outcome measurement approaches.

Chapter 1 discussed the value of using the ICF and the tailored ICF Core Sets to consider and understand the interplay between upper limb motor impairment and the ability to perform an activity or to participate in a life situation. The potential for the complete ICF framework and the Core sets to suggest scope for outcome measurement was also discussed in Chapter 1. However, as previously stated it became apparent that it was unknown whether any of the ICF Core Sets for stroke were applicable to and of sufficient breadth to capture and understand the impact of post-stroke upper limb spasticity and thus potentially guide measurement. Study 2 (Chapter 2b) answered this research question demonstrating the Comprehensive Core Set for stroke to hold appropriate capability. Conversely, whilst the brevity of the Brief Core Set is appealing, it failed to adequately capture the breadth and depth illuminated and the Extended version offered no further advantage.

Study 2 also identified that survivors reported upper limb spasticity to primarily impact the ability to control voluntary movement and to use the arm and hand in activity. This knowledge of the lived experience coupled with the insights into the applicability of the ICF framework led to investigating whether precise measurement of the identified areas of concern, upper limb activity, was possible. This was completed within Study 3 which determined the psychometric properties of measures of upper limb activity performance and goal attainment scaling for adults with and without spasticity. The systematic review findings whilst limited in the conclusions able to be drawn due to evidence gaps, provide clinicians with a selection of outcome measurement tools with supporting evidence to guide measurement practice. To attend to the second component of best practice neurorehabilitation, goal setting, and to further consider personalised and precise

outcome measurement, the psychometric properties and clinical utility of GAS and GAS-Light were investigated. Study 4 (Chapter 6) and Study 5 (Chapter 7) results demonstrated both goal attainment scaling tools provided equally important clinical information and equal determination of goal attainment. Variations detected in the time required to administer and score provide clinicians with novel information to inform their clinical practice. In this cohort of ABI survivors, GAS-Light requires less time whilst still resulting in good engagement and satisfaction with the goals set by ABI survivors.

The findings from this program of studies have been integrated into clinical recommendations on the use of the GAS-Light, education recommendations to ensure clinicians gain an understanding of the impact of living with upper limb movement impairments coupled with spasticity, and research recommendations for further research to determine the psychometrics of activity measurement tools prior to their use.

8.5. Chapter Synopsis

This chapter has presented the research questions and results of the five studies conducted in this program of research to answer those questions. The contribution of this series of studies to extending knowledge in the area of upper limb outcome measurement is also discussed, along with implications for clinical practice, education and research. Findings from this program of studies support the ability to achieve precise and individualized measurement of upper limb activity following an ABI. Exploration of the lived experience and the measurement of outcomes is imperative to continue to build the evidence base for the intervention effectiveness and most importantly to target outcomes that are meaningful to those living with an upper limb impairment following ABI.

References

Ada, L., O'Dwyer, N., & O'Neill, E. (2006). Relation between spasticity, weakness and contracture of the elbow flexors and upper limb activity after stroke: An observational study. *Disability and Rehabilitation, 28*(13-14), 891-897. doi:10.1080/09638280500535165

Adams, S. A., Pickering, R. M., Ashburn, A., & Lincoln, N. B. (1997). The scalability of the Rivermead Motor Assessment in nonacute stroke patients. *Clinical Rehabilitation, 11*, 52-59. doi:10.1177/026921559701100108

Adams, S. A., Pickering, R. M., & Taylor, D. (1997). The scalability of the Rivermead Motor Assessment in acute stroke patients. *Clinical Rehabilitation, 11*, 42-51. doi:10.1177/026921559701100107

Alderman, N., Dawson, K., Rutterford, N. A., & Reynolds, P. J. (2001). A comparison of the validity of self-report measures amongst people with acquired brain injury: A preliminary study of the usefulness of EuroQol-5D. *Neuropsychological Rehabilitation, 11*(5), 529-537. doi:10.1080/09602010042000231

Algurén, B., Lundgren-Nilsson, Å., & Sunnerhagen, K. S. (2010). Functioning of stroke survivors - A validation of the ICF core set for stroke in Sweden. *Disability and Rehabilitation, 32*(7), 551-559. doi:10.3109/09638280903186335

Ali, M., Fulton, R., Quinn, T., Brady, M., & VISTA. (2013). How well do standard stroke outcome measures reflect quality of life? A retrospective analysis of clinical trial data. *Stroke, 44*, 3161-3165. doi:10.1161/STROKEAHA.113.001126

Alt Murphy, M., Björkdahl, A., Forsberg-Wärleby, G., & Persson, C. U. (2021). Implementation of evidence-based assessment of the upper extremities in stroke rehabilitation: From evidence to clinical practice. *Journal of Rehabilitation Medicine*. *53*(1). doi:10.2340/16501977-2790

Alt Murphy, M., Resteghini, C., Feys, P., & Lamers, I. (2015). An overview of systematic reviews on upper extremity outcome measures after stroke. *BMC Neurology, 15*(1). doi:10.1186/s12883-015-0292-6

Alusi, S. H., Worthington, J., Glickman, S., Findley, L. J., & Bain, P. G. (2000). Evaluation of three different ways of assessing tremor in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry, 68*, 756-760. doi:10.1136/jnnp.68.6.756

Anderson, C., Laubscher, S. D., & Burns, R. (1996). Validation of the Short Form 36 (SF-36) Health Survey Questionnaire Among Stroke Patients. *Stroke A Journal of Cerebral Circulation, 27*(10), 1812-1816. doi:10.1161/01.STR.27.10.1812

Andringa, A., van de Port, I., van Wegen, E., Ket, J., Meskers, C., & Kwakkel, G. (2019). Effectiveness of Botulinum Toxin Treatment for Upper Limb Spasticity Poststroke Over Different ICF Domains: A Systematic Review and Meta-Analysis. *Archives of Physical Medicine and Rehabilitation, 100*(9), 1703-1725. doi:10.1016/j.apmr.2019.01.016

Arntzen, C., Borg, T., & Hamran, T. (2015). Long-term recovery trajectory after stroke: an ongoing negotiation between body, participation and self. *Disability and Rehabilitation, 37*(18), 1626-1634. doi:10.3109/09638288.2014.972590

Ashford, S., Jackson, D., & Turner-Stokes, L. (2015). Goal setting, using goal attainment scaling, as a method to identify patient selected items for measuring arm function. *Physiotherapy (United Kingdom), 101*(1), 88-94. doi:10.1016/j.physio.2014.04.001

Ashford, S., Siegert, R. J., & Alexandrescu, R. (2016). Rasch measurement: the Arm Activity measure (ArmA) passive function sub-scale. *Disability and Rehabilitation, 38*(4), 384-390. doi:10.3109/09638288.2015.1041613

Ashford, S., Slade, M., Malaprade, F., & Turner-Stokes, L. (2008). Evaluation of functional outcome measures for the hemiparetic upper limb: A systematic review. *Journal of Rehabilitation Medicine, 40*(10), 787-795. doi:10.2340/16501977-0276

Ashford, S., Slade, M., Nair, A., & Turner-Stokes, L. (2014). Arm Activity measure (ArmA) application for recording functional gain following focal spasticity treatment.

International Journal of Therapy & Rehabilitation, 21(1), 10-17. doi:10.12968/ijtr.2014.21.1.10

Ashford, S., Slade, M., & Turner-Stokes, L. (2013). Conceptualisation and development of the arm activity measure (ArmA) for assessment of activity in the hemiparetic arm. *Disability and Rehabilitation, 35*(18), 1513-1518. doi:10.3109/09638288.2012.743602

Ashford, S., & Turner-Stokes, L. (2013). Systematic review of upper-limb function measurement methods in botulinum toxin intervention for focal spasticity. *Physiotherapy Research International, 18*(3), 178-189. doi:10.1002/pri.1554

Ashford, S., Turner-Stokes, L., Siegert, R., & Slade, M. (2013). Initial psychometric evaluation of the Arm Activity Measure (ArmA): a measure of activity in the hemiparetic arm. *Clinical Rehabilitation, 27*(8), 728-740. doi:10.1177/0269215512474942

Ashford, S., & Turner-Stokes, L. (2006). Goal attainment for spasticity management using botulinum toxin. *Physiotherapy Research International, 11*(1), 24-34. doi:10.1002/pri.36

Ashford, S., & Turner-Stokes, L. (2013). Systematic review of upper-limb function measurement methods in botulinum toxin intervention for focal spasticity. *Physiotherapy Research International, 18*(3), 178-189. doi:10.1002/pri.1554

Ashford, S., Williams, H., Nair, A., Orridge, S., & Turner-Stokes, L. (2019). Categorisation of goals set using Goal Attainment Scaling for treatment of leg spasticity: a multicentre analysis. *Disability and Rehabilitation, 41*(16), 1925-1930. doi:10.1080/09638288.2018.1451927

Atkins, L., Francis, J., Islam, R., O'Connor, D., Patey, A., Ivers, N., . . . Michie, S. (2017). A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implementation Science, 12*(1). doi:10.1186/s13012-017-0605-9

Australasian Rehabilitation Outcomes Centre. (2020a). *Anywhere Hospital AROC Ambulatory Report, July 2019-June 2020*. Retrieved from https://www.uow.edu.au/ahsri/aroc/dataset/reports-benchmarks/

Australasian Rehabilitation Outcomes Centre. (2020b). *The AROC Annual Report: The state of Rehabilitation in Australia in 2019*. Retrieved from https://www.uow.edu.au/ahsri/aroc/dataset/reports-benchmarks/

Australian Institute of Health and Welfare. (2007). *Disability in Australia: acquired brain injury. Bulletin no. 55. Cat no. AUS 96*. Canberra: AIHW

Australian Institute of Health and Welfare. (2020). *Cardiovascular disease. Cat. no. CVD 83*. Canberra: AIHW Retrieved from https://www.aihw.gov.au/reports/heart-stroke-vascular-diseases/cardiovascular-health-compendium

Axford, R., Minichiello, V., Coulson, I., & O'Brien, A. (1999). Research in health: an overview. In V. Minichiello, G. Sullivan, K. Greenwood, & R. Axford (Eds.), *Handbook for research methods in health sciences*. Sydney Australia: Pearson Education Australia.

Bakheit, A. M., Zakine, B., Maisonobe, P., Aymard, C., Fhedoroff, K., Hefter, H., . . . Wissel, J. (2010). The profile of patients and current practice of treatment of upper limb muscle spasticity with botulinum toxin type A: an international survey. *International Journal of Rehabilitation Research, 33*(3), 199-204. doi:10.1097/MRR.0b013e328332f5e0

Baldwin, W. (2000). Information no one else knows: the value of self-report. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: implication for research and practice*. Taylor & Francis Group.

Bamford, J. M., Sandercock, P. A., Warlow, C. P., & Slattery, J. (1989). Interobserver agreement for the assessment of handicap in stroke patients. *Stroke, 20*(6), 828.

Barden, H. L., Baguley, I. J., Nott, M. T., & Chapparo, C. (2014). Dynamic computerised hand dynamometry: Measuring outcomes following upper limb botulinum toxin-A injections in adults with acquired brain injury. *Journal of Rehabilitation Medicine, 46*, 314-320. doi:10.2340/16501977-1781

Barden, H. L., Nott, M. T., Heard, R., Chapparo, C., & Baguley, I. J. (2012). Clinical assessment of hand motor performance after acquired brain injury with dynamic computerised hand dynamometry: Construct, concurrent and predictive validity. *Archives of Physical*

Medicine and Rehabilitation, 93, 2257-2263. doi:10.1016/j.apmr.2012.06.028
Barden, H. L. H., Baguley, I. J., Nott, M. T., & Chapparo, C. (2014). Measuring spasticity and fine motor control (Pinch) change in the hand after botulinum toxin-A injection using dynamic computerized hand dynamometry. *Archives of Physical Medicine and Rehabilitation, 95*, 2402-2409. doi:10.1016/j.apmr.2014.05.017
Barer, D. H., & Murphy, J. J. (1993). Scaling the Barthel: a 10-point hierarchical version of the activities of daily living index for use with stroke patients. *Clinical Rehabilitation, 7*(4), 271-277. doi:10.1177/026921559300700401
Barker, R. N., & Brauer, S. G. (2005). Upper limb recovery after stroke: The stroke survivors' perspective. *Disability and Rehabilitation, 27*(20), 1213-1223. doi:10.1080/09638280500075717
Barnes, M., Kocer, S., Murie Fernandez, M., Balcaitiene, J., & Fheodoroff, K. (2017). An international survey of patients living with spasticity. *Disability and Rehabilitation, 39*(14), 1428-1434. doi:10.1080/09638288.2016.1198432
Barnes, M. P. (2003). Principles of neurological rehabilitation. *Journal of Neurology, Neurosurgery and Psychiatry, 74*, 3-7. doi:10.1136/jnnp.74.suppl_4.iv3
Barnes, M. P. (2008). An overview of the clinical management of spasticity. In M. Barnes & G. Johnson (Eds.), *Upper motor neurone syndrome and spasticity: Clinical management and neurophysiology*. Cambridge: Cambridge University Press.
Barreca, S. R., Stratford, P. W., Lambert, C. L., Masters, L. M., & Streiner, D. L. (2005). Test-retest reliability, validity and sensitivity of the Chedoke Arm and Hand Activity Inventory: A new measure of upper-limb function for survivors of stroke. *Archives of Physical Medicine and Rehabilitation, 86*, 1616-1622. doi:10.1016/j.apmr.2005.03.017
Barreca, S. R., Stratford, P. W., Masters, L. M., Lambert, C. L., & Griffiths, J. (2006). Comparing two versions of the Chedoke Arm and Hand Activity Inventory with the Action Research Arm Test. *Physical Therapy, 86*, 245-253. doi:10.1093/ptj/86.2.245
Barreca, S. R., Stratford, P. W., Masters, L. M., Lambert, C. L., Griffiths, J., & McBay, C. (2006). Validation of three shortened versions of the Chedoke Arm and Hand Activity Inventory. *Physiotherapy Canada, 58*, 148-156. doi:10.3138/ptc.58.2.148
Barton, B., & Peat, J. (2014). *Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal* (2nd ed.): Somerset: Wiley.
Barton, G. R., Sach, T. H., Avery, A. J., Jenkinson, C., Doherty, M., Whynes, D. K., & Muir, K. R. (2008). A comparison of the performance of the EQ-5D and SF-6D for individuals aged ≥ 45 years. *Health Economics, 17*, 815-832. doi:10.1002/hec.1298
Barton, G. R., Sach, T. H., Doherty, M., Avery, A. J., Jenkinson, C., & Muir, K. R. (2008). An assessment of the discriminative ability of the EQ-$5D_{index,}$ SF-6D, and EQ VAS, using sociodemographic factors and clinical conditions. *European Journal of Health Economics, 9*, 237-249. doi:10.1007/s10198-007-0068-z
Becker, G. (1993). Continuity after a stroke? implications of life-course disruption in old age. *The Gerontologist, 33*(2), 148-158. doi:10.1093/geront/33.2.148
Beebe, A. J., & Lang, E. C. (2009). Relationships and responsiveness of six upper extremity function tests during the first six months of recovery after stroke. *Journal of Neurologic Physical Therapy, 33*(2), 96-103. doi:10.1097/NPT.0b013e3181a33638
Benedict, H. B., Holtzer, R., Motl, R. W., Foley, F. W., Kaur, S., Hojnacki, D., & Weinstock-Guttman, B. (2011). Upper and lower extremity motor function and cognitive impairment in multiple sclerosis. *Journal of the International Neuropsychological Society, 17*, 643-653. doi:10.1017/S1355617711000403
Bhakta, B. B., Cozens, J. A., Chamberlain, M. A., & Bamford, J. M. (2000). Impact of botulinum toxin type A on disability and carer burden due to arm spasticity after stroke: a randomsed double blind placebo controlled trial. *Journal of Neurology, Neurosurgery & Psychiatry, 69*, 217-221. doi:10.1136/jnnp.69.2.217
Bhakta, B. B., O'Connor, R. J., & Cozens, A. J. (2008). Associated reactions after stroke: A randomized controlled trial of the effect of botulinum toxin type A. *Journal of*

Rehabilitation Medicine, 40(1), 36-41. doi:10.2340/16501977-0120
Bhimani, R., & Anderson, L. (2014). Clinical understanding of spasticity: Implications for practice. *Rehabilitation Research and Practice, 2014, 279175*. doi:10.1155/2014/279175
Bhimani, R. H., McAlpine, C. P., & Henly, S. J. (2012). Understanding spasticity from patients' perspectives over time. *Journal of Advanced Nursing, 68*(11), 2504-2514. doi:10.1111/j.1365-2648.2012.05949.x
Blennerhassett, J. M., Avery, R. M., & Carey, L. M. (2010). The test-retest and responsiveness to change of the Hand Function Survey during stroke rehabilitation. *Australian Occupational Therapy Journal, 57*, 431-438. doi:10.1111/j.1440-1630.2010.00884.x
Bohannon, R. W. (1999). Motricity index scores are valid indicators of paretic upper extremity strength following stroke. *Journal of Physical Therapy Science, 11*, 59-61. doi:10.1589/jpts.11.59
Bohannon, R. W., Warren, M. E., & Cogman, K. A. (1991). Motor variables correlated with the hand-to-mouth maneuver in stroke patients. *Archives of Physical Medicine and Rehabilitation, 72*(9), 682-684. doi:10.5555/uri:pii:0003999391902155
Boissy, P., Bourbonnais, D., Carlotti, M. M., Gravel, D., & Arsenault, B. A. (1999). Maximal grip force in chronic stroke subjects and its relationship to global upper extremity function. *Clinical Rehabilitation, 13*(4), 354-362. doi:10.1191/026921599676433080
Bosma, M. S., Nijboer, T. C. W., Caljouw, M. A. A., & Achterberg, W. P. (2020). Impact of visuospatial neglect post-stroke on daily activities, participation and informal caregiver burden: A systematic review. *Annals of Physical and Rehabilitation Medicine, 63*(4), 344-358. doi:10.1016/j.rehab.2019.05.006
Bouwens, S. F. M., van Heugten, C. M., & Verhey, F. R. J. (2008). Review of Goal Attainment Scaling as a useful outcome measure in psychogeriatric patients with cognitive disorders. *Dementia and Geriatric Cognitive Disorders, 26*(6), 528-540. doi:10.1159/000178757
Bovend'Eerdt, T. J., Botell, R. E., & Wade, D. T. (2009). Writing SMART rehabilitation goals and achieving goal attainment scaling: a practical guide. *Clinical Rehabilitation, 23*(4), 352-361. doi:10.1177/0269215508101741
Bovend'Eerdt, T. J. H., Dawes, H., Izadi, H., & Wade, D. T. (2011). Agreement between two different scoring procedures for goal attainment scaling is low. *Journal of Rehabilitation Medicine, 43*(1), 46-49. doi:10.2340/16501977-0624
Brashear, A., Gordon, M. F., Eliovic, E., Kassicieh, V. D., Marciniak, C., Do, M., . . . Turkel, C. (2002). Intramuscular injection of Botulinum Toxin for the treatment of wrist and finger spasticity after a stroke. *The New England Journal of Medicine, 347*, 395-400. doi:10.1056/NEJMoa011892
Brashear, A., Zafonte, R., Corcoran, M., Galvez-Jimenez, N., Gracies, J. M., Gordon, M. F., . . . Turkel, C. (2002). Inter- and Intrarater reliability of the Ashworth scale and the Disability Assessment Scale in patients with upper-limb poststroke spasticity *Archives of Physical Medicine and Rehabilitation, 83*, 1349-1354. doi:10.1053/apmr.2002.35474
Brock, K., Black, S., Cotton, S., Kennedy, G., Wilson, S., & Sutton, E. (2009). Goal achievement in the six months after inpatient rehabilitation after stroke. *Disability and Rehabilitation, 31*(11), 880-886. doi:10.1080/09638280802356179
Brown, A. W., Therneau, T. M., Schultz, B. A., Niewczyk, P. M., & Granger, C. V. (2015). Measure of functional independence dominates discharge outcome prediction after inpatient rehabilitation for stroke. *Stroke, 46*, 1038-1044. doi:10.1161/STROKEAHA.114.007392.
Burnes, D., Connolly, M.-T., Hamilton, R., & Lachs, M. S. (2018). The feasibility of goal attainment scaling to measure case resolution in elder abuse and neglect adult protective services intervention. *Journal of Elder Abuse and Neglect, 30*(3), 209-222. doi:10.1080/08946566.2018.1454864
Burridge, J., Alt Murphy, M., Buurke, J., Feys, P., Keller, T., Klamroth-Marganska, V., . . . Hughes, A.-M. (2019). A systematic review of international clinical guidelines for rehabilitation of people with neurological conditions: what recommendations are made for upper limb assessment? *Frontiers in Neurology, 10*, 567-567. doi:10.3389/fneur.2019.00567

Burridge, J. H., Turk, R., Notley, S. V., Pickering, R. M., & Simpson, D. M. (2009). The relationship between upper limb activity and impairment in post-stroke hemiplegia. *Disability and Rehabilitation, 31*(2), 109-117. doi:10.1080/09638280701824699

Burton, L.-J., Tyson, S., & McGovern, A. (2013). Staff perceptions of using outcome measures in stroke rehabilitation. *Disability and Rehabilitation, 35*(10), 828-834. doi:10.3109/09638288.2012.709305

Cadilhac, D. A., Purvis, T., Reyneke, M., Dalli, L., Kim, J., & Kilkenny, M. F. (2019). *Evaluation of the National Stroke Audit Program: 20-year report.* Retrieved from https://informme.org.au/en/stroke-data/About-stroke-audits

Calvert, M., Kyte, D., Price, G., Valderas, J. M., & Hjollund, N. H. (2019). Maximising the impact of patient reported outcome assessment for patients and society. *BMJ, 364*, k5267-k5267. doi:10.1136/bmj.k5267

Canning, C. G., Ada, L., Adams, R., & O'Dwyer, N. J. (2004). Loss of strength contributes more to physical disability after stroke than loss of dexterity. *Clinical Rehabilitation, 18*(3), 300-308. doi:10.1191/0269215504cr715oa

Carey, L. M. (1995). Somatosensory loss after stroke. *Critical Reviews in Physical and Rehabilitation Medicine, 7*(1), 51-91. doi:10.1615/CritRevPhysRehabilMed.v7.i1.40

Carey, L. M. (2012). Touch and body sensations. In L. M. Carey (Ed.), *Stroke Rehabilitation: Insights from Neuroscience and Imaging*. New York, United States: Oxford University Press, Incorporated.

Carr, J. H., Shephard, R. B., Nordholm, L., & Lynne, D. (1985). Investigation of a new motor assessment scale for stroke patients. *Physical Therapy, 65*, 175-180. doi:10.1093/ptj/65.2.175

Carroll, D. (1965). A quantitative test of upper extremity function. *Journal of Chronic Disability, 18*, 479-491.

Cattaneo, D., Lamers, I., Bertoni, R., Feys, P., & Jonsdottir, J. (2017). Participation restriction in people with multiple sclerosis: prevalence and correlations with cognitive, walking, balance, and upper limb impairments. *Archives of Physical Medicine and Rehabilitation, 98*(7), 1308-1315. doi:10.1016/j.apmr.2017.02.015

Cawood, J., Visagie, S., & Mji, G. (2016). Impact of post-stroke impairments on activities and participation as experienced by stroke survivors in a Western Cape setting. *South African Journal of Occupational Therapy, 46*(2). doi:10.17159/2310-3833/2016/v46n2a3

Celik, O., O'Malley, M. K., Boake, C., Levin, H. S., Yozbatiran, N., & Reistetter, T. A. (2010). Normalized movement quality measures for therapeutic robots strongly correlate with clinical motor impairment measures. *IEEE Transactions on Neural Systems and Rehabilitation Engineering., 18*(4), 433-444. doi:10.1109/TNSRE.2010.2047600

Chang, C.-L., Munin, M. C., Skidmore, E. R., Niyonkuru, C., Huber, L. M., & Weber, D. J. (2009). Effect of baseline spastic hemiparesis on recovery of upper-Limb function following botulinum toxin type A injections and postinjection therapy. *Archives of Physical Medicine and Rehabilitation, 90*(9), 1462-1468. doi:10.1016/j.apmr.2009.03.008

Chen, H. F., Lin, K. C., Wu, C. Y., & Chen, C. L. (2012). Rasch validation and predictive validity of the action research arm test in patients receiving stroke rehabilitation. *Archives of Physical Medicine and Rehabilitation, 93*, 1039-1045. doi:10.1016/j.apmr.2011.11.033

Chen, S., Wolf, S. L., Zhang, Q., Thompson, P. A., & Winstein, C. J. (2012). Minimal detectable change of the actual amount of use test and the motor activity log: The EXCITE trial. *Neurorehabilitation and Neural Repair, 26*(5), 507-514. doi:10.1177/1545968311425048

Cieza, A., Fayed, N., Bickenbach, J., & Prodinger, B. (2019). Refinements of the ICF Linking Rules to strengthen their potential for establishing comparability of health information. *Disability and Rehabilitation, 41*(5), 574-583. doi:10.3109/09638288.2016.1145258

Cieza, A., Geyh, S., Chatterji, S., Kostanjsek, N., Ustun, B., & Stucki, G. (2005). ICF linking rules: an update based on lessons learned. *Journal of Rehabilitation Medicine, 37*(37), 212-218. doi:10.1080/16501970510040263

Clarke, P. (2009). Understanding the experience of stroke: A mixed-method research agenda.

The Gerontologist, 49(3), 293-302. doi:10.1093/geront/gnp047
Collin, C., & Wade, D. (1990). Assessing motor impairment after stroke: a pilot reliability study. *Journal of Neurology, Neurosurgery, and Psychiatry, 53*(7), 576-579. doi:10.1136/jnnp.53.7.576
Collin, C., Wade, D. T., Davies, S., & Horne, V. (1988). The Barthel ADL Index: A reliability study. *International Disability Studies, 10*(2), 61-63. doi:10.3109/09638288809164103
Colquhoun, H. L., Lamontagne, M.-E., Duncan, E. A., Fiander, M., Champagne, C., & Grimshaw, J. M. (2017). A systematic review of interventions to increase the use of standardized outcome measures by rehabilitation professionals. *Clinical Rehabilitation, 31*(3), 299-309. doi:10.1177/0269215516644309
Connell, L. A., & Tyson, S. F. (2012). Clinical reality of measuring upper-limb ability in neurologic conditions: a systematic review. *Archives of Physical Medicine and Rehabilitation, 93*(2), 221-228. doi:10.1016/j.apmr.2011.09.015
Corrigan, J. D., Smith-Knapp, K., & Granger, C. V. (1997). Validity of the functional independence measure for persons with traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 78*(8), 828-834. doi:10.1016/S0003-9993(97)90195-7
Costelloe, L., O'Rourke, K., McGuigan, C., Walsh, C., Tubridy, N., & Hutchinson, M. (2008). The longitudinal relationship between the patient-reported Multiple Sclerosis Impact Scale and the clinician-assessed Multiple Sclerosis Functional Composite. *Multiple Sclerosis, 14*, 255-258. doi:10.1177/1352458507081274
Cullen, N., Krakowski, A., & Taggart, C. (2014). Functional independence measure at rehabilitation admission as a predictor of return to driving after traumatic brain injury. *Brain Injury, 28*(2), 189-195. doi:10.3109/02699052.2013.862738
Cusick, A., Lannin, N., & Kinnear, B. Z. (2015). Upper limb spasticity management for patients who have received Botulinum Toxin A injection: Australian therapy practice. *Australian Occupational Therapy Journal, 62*(1), 27-40. doi:10.1111/1440-1630.12142
Cuthbert, J. P., Harrison-Felixx, C., Corrigan, J. D., Bell, J. M., Haarbauer-Krupa, J. K., & Miller, C. (2015). Unemployment in the United States after traumatic brain injury for working-age individuals: prevalence and associated factors 2 years postinjury. *Journal of Head Trauma Rehabilitation, 30*(3), 160-174. doi:10.1097/HTR.0000000000000090
Dang, M., Ramsaran, K. D., Street, M. E., Syed, S. N., Barclay-Goddard, R., Stratford, P. W., & Miller, P. A. (2011). Estimating the accuracy of the Chedoke-McMaster Stroke Assessment predictive equations for stroke rehabilitation. *Physiotherapy Canada, 63*(3), 334-341. doi:10.3138/ptc.2010.17
De Souza, L. H., Langton-Hewer, R., & Miller, S. (1980). Assessment of recovery of arm control in hemiplegic stroke patients. 1. Arm function tests. *International Rehabilitation Medicine, 2*(1), 3-9. doi:10.3109/09638288009163947
De Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide*. New York: Cambridge University Press.
De Weerdt, W. J. G., & Harrison, M. A. (1985). Measuring recovery of arm-hand function in stroke patients: A comparison of the Brunnstrom-Fugl-Meyer test and the Action Research Arm test. *Physiotherapy Canada, 37*(2), 65-70. doi:10.3138/ptc.37.2.065
Dekker, J., de Groot, V., ter Steeg, A. M., Vloothuis, J., Holla, J., Collette, E., . . . Littooij, E. (2019). Setting meaningful goals in rehabilitation: rationale and practical tool. *Clinical Rehabilitation, 34*(1), 3-12. doi:10.1177/0269215519876299
Deloitte Access Economics. (2017). *Stroke in Australia - No postcodes untouched* Retrieved from https://strokefoundation.org.au
Demetrios, M., Khan, F., Turner-Stokes, L., Brand, C., & McSweeney, S. (2013). Multidisciplinary rehabilitation following botulinum toxin and other focal intramuscular treatment for post-stroke spasticity. *Cochrane Database of Systematic Reviews*(6). doi:10.1002/14651858.CD009689.pub2
Demeurisse, G., Demol, O., & Robaye, E. (1980). Motor evaluation in vascular hemiplegia. *European Neurology, 19*, 382-389. doi:10.1159/000115178

Dennis, M., O'Rourke, S., Lewis, S., Sharpe, M., & Warlow, C. (2000). Emotional outcomes after stroke: factors associated with poor outcome. *Journal of Neurology, Neurosurgery & Psychiatry, 68*, 47-52. doi:10.1136/jnnp.68.1.47

Dewey, H. M., Sherry, L. J., & Collier, J. M. (2007). Stroke Rehabilitation 2007: What Should it Be? *International Journal of Stroke, 2*(3), 191-200. doi:10.1111/j.1747-4949.2007.00146.x

Doan, Q. V., Brashear, A., Gillard, P. J., Varon, S. F., Vandenburgh, A. M., Turkel, C. C., & Elovic, E. P. (2012). Relationship between disability and health-related quality of life and caregiver burden in patients with upper limb poststroke spasticity. *Physical Medicine & Rehabilitation, 4*(1), 4-10. doi:10.1016/j.pmrj.2011.10.001

Dobson, F., Hinman, R. S., Hall, M., Terwee, C. B., Roos, E. M., & Bennell, K. L. (2012). Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. *Osteoarthritis and Cartilage, 20*(12), 1548-1562. doi:10.1016/j.joca.2012.08.015

Doig, E., Fleming, J., Kuipers, P., & Cornwell, P. L. (2010). Clinical utility of the combined use of the Canadian Occupational Performance Measurer and Goal Attainment Scaling. *American Journal of Occupational Therapy, 64*(6), 904-914. doi:10.5014/ajot.2010.08156

Donkor, E. S. (2018). Stroke in the 21st Century: A snapshot of the burden, epidemiology, and quality of life. *Stroke Research and Treatment, 2018*, 1-10. doi:10.1155/2018/3238165

Donnelly, C., & Carswell, A. (2002). Individualized outcome measures: A review of the literature. *Canadian Journal of Occupational Therapy, 69*(2), 84-94. doi:10.1177/000841740206900204

Donovan, K., Lord, S. E., McNaughton, H. K., & Weatherall, M. (2008). Mobility beyond the clinic: the effect of environment on gait and its measurement in community-ambulant stroke survivors. *Clinical Rehabilitation, 22*, 556-563. doi:10.1177/0269215507085378

Dorman, P. J., Dennis, M., & Sandercock, P. (1999). How do scores on the EuroQol relate to scores on the SF-36 after stroke? *Stroke, 30*, 2146-2151. doi:10.1161/01.STR.30.10.2146

Dorman, P. J., Slattery, J., Farrell, B., Dennis, M., & Sandercock, P. (1998). Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 Questionnaires after stroke. *Stroke, 29*, 63-68. doi:10.1161/01.STR.29.1.63

Dorman, P. J., Waddell, F., Slattery, J., Dennis, M., & Sandercock, P. (1997). Is the EuroQol a valid measure of health related quality of life after stroke? *Stroke, 28*, 1876-1882. doi:10.1161/01.STR.28.10.1876

Doyle, S., Bennett, S., Fasoli, S. E., & McKenna, K. T. (2010). Interventions for sensory impairment in the upper limb after stroke. *The Cochrane Database of Systematic Reviews, 2010* (6), CD006331. doi:10.1002/14651858.CD006331.pub2

Dromerick, A. W., Lang, C. E., Birkenmeier, R., Hahn, M. G., Sahrmann, S. A., & Edwards, D. F. (2006). Relationships between upper-limb functional limitation and self-reported disability 3 months after stroke. *Journal of Rehabilitation and Development, 43*(3), 401-408. doi:10.1682/JRRD.2005.04.0075

Duncan, E., & Murray, J. (2012). The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. *BMC Health Services Research, 12*(1), 96. doi:10.1186/1472-6963-12-96

Duncan Millar, J., van Wijck, F., Pollock, A., & Ali, M. (2019). Outcome measures in post-stroke arm rehabilitation trials: do existing measures capture outcomes that are important to stroke survivors, carers, and clinicians? *Clinical Rehabilitation, 33*(4), 737-749. doi:10.1177/0269215518823248

Duncan, P. W., Bode, R. K., Min Lai, S., Perera, S., & Investigators, G. A. i. N. A. (2003). Rasch analysis of a new stroke-specific outcome scale: The Stroke Impact Scale. *Archives of Physical Medicine and Rehabilitation, 84*, 950-963. doi:10.1016/S0003-9993(03)00035-2

Duncan, P. W., Min Lai, S., Tyler, D., Perera, S., Reker, D. M., & Studenski, S. (2002). Evaluation of proxy responses to the Stroke Impact Scale. *Stroke, 33*, 2593-2599. doi:10.1161/01.STR.0000034395.06874.3E

Duncan, P. W., Reker, D., Kwon, S., Lai, S.-M., Studenski, S., Perera, S., . . . Marquez, J. (2005).

Measuring stroke impact with the Stroke Impact Scale: telephone versus mail administration in veterans with stroke. *Medical Care, 43*(5), 507-515. doi:10.1097/01.mlr.0000160421.42858.de

Duncan, P. W., Samsa, G. P., Weinberger, M., Goldtein, L. B., Bonito, A., Witter, D. M., . . . Matchar, D. (1997). Health status of individuals with mild stroke. *Stroke, 28*, 740-745. doi:10.1161/01.STR.28.4.740

Duncan, P. W., Wallace, D., Min Lai, S., Johnson, D., Embretson, S., & Laster, L. J. (1999). The Stroke Impact Scale Version 2.0: Evaluation of reliability, validity, and sensitivity to change. *Stroke, 30*(10), 2131-2140. doi:10.1161/str.30.10.2131

Dunn, G. (1989). *Design and analysis of reliability studies: the statistical evaluation of measurement errors*. New York : London: Oxford University Press.

Edwards, D. F., Hahn, M., Baum, C., & Dromerick, A. W. (2006). The impact of mild stroke on meaningful activity and life satisfaction. *Journal of Stroke and Cerebrovascular Diseases, 15*(4), 151-157. doi:10.106/j.jstrokecerebrovasdis.2006.04.001

Edwards, D. F., Lang, C. E., Wagner, J. M., Birkenmeier, R., & Dromerick, A. W. (2012). An evaluation of the Wolf Motor Function Test in motor trials early after stroke. *Archives of Physical Medicine and Rehabilitation, 93*, 660-668. doi:10.1016/j.apmr.2011.10.005

Eftekhar, P., Mochizuki, G., Dutta, T., Richardson, D., & Brooks, D. (2016). Goal Attainment Scaling in individuals with upper limb spasticity post stroke. *Occupational Therapy International, 23*(4), 379-389. doi:10.1002/oti.1440

Egan, M., Davis, C. G., Dubouloz, C.-J., Kessler, D., & Kubina, L.-A. (2014). Participation and well-being poststroke: Evidence of reciprocal effects. *Archives of Physical Medicine and Rehabilitation, 95*, 262-268. doi:10.1016/j.apmr.2013.08.013

Elia, A. E., Filippini, G., Calandrella, D., & Albanese, A. (2009). Botulinum neurotoxins for post-stroke spasticity in adults: A systematic review. *Movement Disorders, 24*(6), 801-812. doi:10.1002/mds.22452

Elovic, P. E., Simone, K. L., & Zafonte, K. R. (2004). Outcome assessment for spasticity management in the patient with traumatic brain injury: The state of the art. *Journal of Head Trauma Rehabilitation, 19*(2), 155-177. doi:10.1097/00001199-200403000-00007

Eriksson, G., Baum, M. C., Wolf, T. J., & Connor, L. T. (2013). Perceived participation after stroke: The influence of activity retention, reintegration, and perceived recovery. *The American Journal of Occupational Therapy, 67*, e13-e18. doi:10.5014/ajot.2013.008292

Ertzgaard, P., Ward, A. B., Wissel, J., & Borg, J. (2011). Practical considerations for goal attainment scaling during rehabilitation following acquired brain injury. *Journal of Rehabilitationl Medicine, 43*(1), 8-14. doi:10.2340/16501977-0664

Esquenazi, A. (2011). The human and economic burden of poststroke spasticity and muscle overactivity. *Clinical Outcomes Management, 18*(1), 607-614.

Esquenazi, A., Jacinto, J., & Lysandropoulos, A. (2020). Patient perspectives on the impact of adult spasticity on daily life and its management with botulinum toxin type A (1639). *Neurology*, 94(Supplement 15).

Ewert, T., Grill, E., Bartholomeyczik, S., Finger, M., Mokrusch, T., Kostanjsek, N., & Stucki, G. (2005). ICF Core Set for patients with neurological conditions in the acute hospital. *Disability and Rehabilitation, 27*(7-8), 367-373. doi:10.1080/09638280400014014

Faircloth, C. A., Boylstein, C., Rittman, M., Young, M. E., & Gubrium, J. (2004). Sudden illness and biographical flow in narratives of stroke recovery. *Sociology of Health and Illness, 26*(2), 242-261. doi:10.1111/j.1467-9566.2004.00388.x

Faria-Fortini, I., Michaelsen, S. M., Cassiano, J. G., & Teixeira-Salmela, L. F. (2011). Upper extremity function in stroke subjects: Relationships between the International Classification of Functioning, Disability, and Health domains. *Journal of Hand Therapy, 24*(3), 257-265. doi:10.1016/j.jht.2011.01.002

Fheodoroff, K., Jacinto, J., Geurts, A., Molteni, F., Franco, J. H., Santiago, T., . . . Gracies, J.-M. (2016). How can we improve current practice in spastic paresis. *Eur Neurol Rev, 11*(2), 79-86. doi:10.17925/ENR.2016.11.02.79

Filiatrault, J., Bertrand Arsenault, A., Dutil, E., & Bourbonnais, D. (1991). Motor function and activities of daily living assessments: a study of three tests for persons with hemiplegia. *The American Journal of Occupational Therapy, 45*(9), 806-810. doi:10.5014/ajot.45.9.806

Findler, M., Cantor, J., Haddad, L., Gordon, W., & Ashman, T. (2001). The reliability and validity of the SF-36 health survey questionnaire for use with individuals with traumatic brain injury. *Brain Injury, 15*(8), 715-723. doi:10.1080/02699050010013941

Fisk, J. D., Brown, M. G., Sketris, I. S., Metz, L. M., Murray, T. J., & Stadnyk, K. J. (2005). A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *Journal of Neurology, Neurosurgery and Psychiatry, 76*, 58-63. doi:10.1136/jnnp.2003.017897

Fleming, M. K., Newham, D. J., Roberts-Lewis, S. F., & Sorinola, I. O. (2014). Self-perceived utilization of the paretic arm in chronic stroke requires high upper limb functional mobility. *Archives of Physical Medicine and Rehabilitation, 95*, 918-924. doi:10.1016/j.apmr.2014.01.009

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198. doi:10.1016/0022-3956(75)90026-6

Francisco, E. G., & McGuire, R. J. (2012). Poststroke spasticity management. *Stroke, 43*(11), 3132-3136. doi:10.1161/STROKEAHA.111.639831

Freeman, J. A., Hobart, J. C., Langdon, D. W., & Thompson, A. J. (2000). Clinical appropriateness: a key factor in outcome measure selection: the 36 item short form health survey in multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry, 68*, 150-156. doi:10.1136/jnnp.68.2.150

Freeman, J. A., Langdon, D. W., Hobart, J. C., & Thompson, A. J. (1996). Health-related quality of life in people with Multiple Sclerosis undergoing inpatient rehabilitation. *Journal of Neurological Rehabilitation, 10*(3), 185-194. doi:10.1177/154596839601000305

Fugl-Meyer, A. R., Jaakso, L., Leyman, I., Olsson, S., & Steglind, S. (1975). The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. *Scandinavian Journal of Rehabilitation Medicine, 7*, 13-31.

Gadidi, V., Katz-Leurer, M., Carmeli, E., & Bornstein, N. M. (2011). Long-term outcome poststroke: Predictors of activity limitation and participation restriction. *Archives of Physical Medicine and Rehabilitation, 92*(11), 1802-1808. doi:10.1016/j.apmr.2011.06.014

Ganapathy, V., Graham, G., Dibonaventura, M. D., Gillard, P., Goren, A., & Zorowitz, R. D. (2015). Caregiver burden, productivity loss, and indirect costs associated with caring for patients with poststroke spasticity. *Clinical Interventions in Ageing, 10*, 1793-1802. doi:org/10.2147/CIA.S91123

Gerber, R. (1999). The role of theory in social and health research. In V. Minichiello, G. Sullivan, K. Greenwood, & R. Axford (Eds.), *Handbook for Research Methods in Health Sciences*. Sydney Australia: Pearson Education Australia.

Geyh, S., Cieza, A., Schouten, J., Dickson, H., Frommelt, P., Omar, Z., . . . Stucki, G. (2004). ICF Core Sets for stroke. *Journal of Rehabilitation Medicine* (44 Suppl), 135. doi:10.1080/16501960410016776

Gillard, P. J., Sucharew, H., Kleindorfer, D., Belagaje, S., Varon, S., Alwell, K., . . . Kissela, B. (2015). The negative impact of spasticity on the health-related quality of life of stroke survivors: a longitudinal cohort study. *Health and Quality of Life Outcomes, 13*(1). doi:10.1186/s12955-015-0340-3

Glässel, A., Coenen, M., Kollerits, B., & Cieza, A. (2012). Validation of the extended ICF core set for stroke from the patient perspective using focus groups. *Disability and Rehabilitation, 34*(2), 157-166. doi:10.3109/09638288.2011.593680

Glässel, A., Coenen, M., Kollerits, B., & Cieza, A. (2014). Content validation of the international classification of functioning, disability and health core set for stroke from gender

perspective using a qualitative approach. *European Journal of Physical and Rehabilitation Medicine, 50*(3), 285-299.
Glässel, A., Kirchberger, I., Kollerits, B., Amann, E., & Cieza, A. (2011). Content validity of the Extended ICF Core Set for stroke: an international Delphi survey of physical therapists. *Physical Therapy, 91*(8), 1211-1222. doi:10.2522/ptj.20100262
Glässel, A., Kirchberger, I., Linseisen, E., Stamm, T., Cieza, A., & Stucki, G. (2010). Content Validation of the International Classification of Functioning, Disability and Health (ICF) Core Set for Stroke: The Perspective of Occupational Therapists. *Canadian Journal of Occupational Therapy, 77*(5), 289-302. doi:10.2182/cjot.2010.77.5.5
Goodkin, D. E., Hertsgaard, D., & Seminary, J. (1988). Upper extremity function in multiple sclerosis: Improving assessment sensitivity with Box-and-Block and Nine-Hole Peg Tests. *Archives of Physical Medicine and Rehabilitation, 69*, 850-854.
Gowland, A. C. (1990). Staging Motor Impairment After Stroke. *Stroke, 21*(9 Supp 2).
Gowland, C., Stratford, P., Ward, M., Moreland, J., Torresin, W., Van Hullenaar, S., . . . Plews, N. (1993). Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment. *Stroke, 24*(1), 58-63. doi:10.1161/str.24.1.8418551
Granger, C. V., Hamilton, B. B., Linacre, J. M., Heinemann, A. W., & Wright, B. D. (1993). Performance profiles for the functional independence measure. *American Journal of Physical Rehabilitation, 72*(2), 84-89. doi:10.1097/00002060-199304000-00005
Grant, C., Goldsmith, C. H., & Anton, H. A. (2014). Inpatient stroke rehabilitation lengths of stay in Canada derived from the national rehabilitation reporting system, 2008 and 2009. *Archives of Physical Medicine and Rehabilitation, 95*, 74-78. doi:10.1016/j.apmr.2013.08.014
Grant, M., & Ponsford, J. (2014). Goal Attainment Scaling in brain injury rehabilitation: Strengths, limitations and recommendations for future applications. *Neuropsychological Rehabilitation, 24*(5), 661-677. doi:10.1080/09602011.2014.901228
Green, J., Forster, A., & Young, J. (2001). A test-retest reliability study of the Barthel Index, the Rivermead Mobility Index, the Nottingham extended Activities of Daily Living Scale and the Frenchay Activities Index in stroke patients. *Disability and Rehabilitation, 23*(15), 670-676. doi:10.1080/09638280110045382
Greenhalgh, J., Long, A. F., Brettle, A. J., & Grant, M. J. (1998). Reviewing and selecting outcome measures for use in routine practice. *Journal of Evaluation in Clinical Practice, 4*(4), 339-350. doi:10.1111/j.1365-2753.1998.tb00097.x
Grenville, J., & Lyne, P. (1995). Patient-centred evaluation and rehabilitative care. *Journal of Advanced Nursing, 22*(5), 965-972. doi:10.1111/j.1365-2648.1995.tb02649.x
Grill, E., Strobl, R., Müller, M., Quittan, M., Kostanjsek, N., & Stucki, G. (2011). ICF Core Sets for early post-acute rehabilitation facilities. *Journal of Rehabilitation Medicine, 43*(2), 131. doi:10.2340/16501977-0641
Grill, E., & Stucki, G. (2011). Criteria for validating comprehensive ICF Core Sets and developing brief ICF Core Set versions. *Journal of Rehabilitation Medicine, 43*(2), 87. doi:10.2340/16501977-0616
Guilfoyle, M. R., Seeley, H. M., Corteen, E., Harkin, C., Richards, H., Menon, D. K., & Hutchinson, P. J. (2010a). Assessing quality of life after traumatic brain in jury: examination of the Short Form 36 Health Survey. *Journal of Neurotrauma, 27*, 2173-2181. doi:10.1089/neu.2010.1353
Hagen, S., Bugge, C., & Alexander, H. (2003). Psychometric properties of the SF-36 in the elderly post-stroke phase. *Journal of Advanced Nursing, 44*(5), 461-468. doi:10.1046/j.0309-2402.2003.02829.x
Hall, K. M., Hamilton, B. B., Gordon, W. A., & Zasler, N. D. (1993). Characteristics and comparisons of functional assessment indices: Disability rating scale, functional independence measure and functional assessment measure. *Journal of Head Trauma Rehabilitation, 8*(2), 60-74. doi:10.1097/00001199-199308020-00008
Hamilton, B. B., & Granger, C. V. (1994). Disability outcomes following inpatient rehabilitation for

stroke. *Physical Therapy, 74*(5), 494-503. doi:10.1093/ptj/74.5.494
Harpster, K., Sheehan, A., Foster, E. A., Leffler, E., Schwab, S. M., & Angeli, J. M. (2019). The methodological application of goal attainment scaling in pediatric rehabilitation research: a systematic review. *Disability and Rehabilitation, 41*(24), 2855-2864. doi:10.1080/09638288.2018.1474952
Harris, J. E., & Eng, J. J. (2007). Paretic upper-limb strength best explains arm activity in people with stroke. *Physical Therapy, 87*(1), 88-97. doi:10.2522/ptj.20060065
Harris, J. E., & Eng, J. J. (2010). Strength training improves upper-limb function in Individuals with stroke: A meta-analysis. *Stroke, 41*(1), 136-140. doi:10.1161/STROKEAHA.109.567438
Hawthorne, G., Gruen, R. L., & Kaye, A. H. (2009). Traumatic brain injury and long-term quality of life: Findings from an Australian study. *Journal of Neurotrauma, 26*, 1623-1633. doi:10.1089/neu.2008.0735
Hawthorne, G., Richardson, J., & Osborne, R. (1999). The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of Health-Related Quality of Life. *Quality of Life Research, 8*(3), 209-224. doi:10.1023/A:1008815005736
Hebert, D., Lindsay, M. P., McIntyre, A., Kirton, A., Rumney, P. G., Bagg, S., . . . Stack, B. (2016). Canadian stroke best practice recommendations: Stroke rehabilitation practice guidelines, update 2015. *International Journal of Stroke, 11*(4), 459-484. doi:10.1177/1747493016643553
Heinemann, A. W., Kirk, P., Hastie, B. A., Semik, P., Hamilton, B. B., Linacre, J. M., . . . Granger, C. (1997). Relationships between disability measures and nursing effort during medical rehabilitation for patients with traumatic brain and spinal cord injury. *Archives of Physical Medicine and Rehabilitation, 78*, 143-149. doi:10.1016/S0003-9993(97)90255-0
Heinemann, A. W., Linacre, J. M., Wright, B. D., Hamilton, B. B., & Granger, C. (1993). Relationships between impairment and physical disability as measured by the Functional Independence Measure. *Archives of physical medicine and rehabilitation, 74*, 566-573.
Heinemann, A. W., Linacre, J. M., Wright, B. D., Hamilton, B. B., & Granger, C. (1994). Measurement characteristics of the Functional Independence Measure. *Topics in Stroke Rehabilitation, 1*(3), 1-15. doi:10.1080/10749357.1994.11754030
Heller, A., Wade, D. T., Wood, V. A., Sunderland, A., Hewer, R. L., & Ward, E. (1987). Arm function after stroke: measurement and recovery over the first three months. *Journal of Neurology, Neurosurgery, and Psychiatry, 50*(6), 714-719. doi:10.1136/jnnp.50.6.714
Herrmann, B. P., Vickrey, B., Hays, R. D., Cramer, J., Devinsky, O., Meador, K., . . . Ellison, G. W. (1996). A comparison of health-related quality of life in patients with epilepsy, diabetes and multiple sclerosis. *Epilepsy Research, 25*(2), 113-118. doi:10.1016/0920-1211(96)00024-1
Hinderer, S. R., & Gupta, S. (1996). Functional outcome measures to assess interventions for spasticity. *Archives of Physical Medicine and Rehabilitation, 77*(10), 1083-1089. doi:10.1016/S0003-9993(96)90073-8
Hirsch, T., Barthel, M., Aarts, P., Chen, Y-A., Freivogel, S., Johnson, M. J., . . . Heise, K-F. (2021). A first step towards the operationalization of the Learned Non-Use phenomenon: A Delphi study. *Neurorehabilitation and Neural Repair,* 35(5), 383-392. doi:10.117/1545968321999064
Hirsch, M. A., Williams, K., Norton, H. J., & Hammond, F. (2014). Reliability of the timed 10-metre walk test during inpatient rehabilitation in ambulatory adults with traumatic brain injury. *Brain Injury, 28*(8), 1115-1120. doi:10.3109/02699052.2014.910701
Hobart, J. C., Williams, L. S., Moran, K., & Thompson, A. J. (2002). Quality of life measurement after stroke: Uses and abuses of the SF-36. *Stroke, 33*, 1348-1356. doi:10.1161/01.STR.0000015030.59594.B3
Hoffmann, T., Bennett, S., & Del Mar, C. (2017). Introduction to evidence-based practice. In T. Hoffmann, S. Bennett, & C. Del Mar (Eds.), *Evidence-Based Practice Across the Health Professions* (3rd ed.). Australia: Elsevier.
Hough, M. (1999). Correlation and regression. In V. Minichiello, G. Sullivan, K. Greenwood, & R.

Axford (Eds.), *Handbook for Research Methods in Health Sciences*. South Melbourne: Addison Wesley.

Houlden, H., Edwards, M., McNeil, J., & Greenwood, R. (2006). Use of the Barthel Index and Functional Independence Measure during early inpatient rehabilitation after single incident brain injury. *Clinical Rehabilitation, 20*, 153-159. doi:10.1191/0269215506cr917oa

Hsueh, I. P., & Hsieh, C. L. (2002). Responsiveness of two upper extremity function instruments for stroke inpatients receiving rehabilitation. *Clinical Rehabilitation, 16*(6), 617-624. doi:10.1191/0269215502cr530oa

Hsueh, I. P., Lee, M. M., & Hsieh, C. L. (2002). The Action Research Arm Test: is it necessary for patients being tested to sit at a standardised table? *Clinical Rehabilitation, 16*(4), 382-388. doi:10.1191/0269215502cr509oa

Hurn, J., Kneebone, I., & Cropley, M. (2006). Goal setting as an outcome measure: a systematic review. *Clinical Rehabilitation, 20*(9), 756-772. doi:10.1177/0269215506070793

Ibuki, A., & Bernhardt, J. (2007). What is spasticity? The discussion continues. *International Journal of Therapy and Rehabilitation, 14*(9), 391-395. doi:10.12968/ijtr.2007.14.9.24579

Intercollegiate Stroke Working Party (Ed.) (2008). *National Clinical Guideline for Stroke: Third Edition*. London: Royal College of Physicians

Intiso, D., Simone, V., Di Rienzo, F., Santamato, A., Russo, M., Tolfa, M., & Basciani, M. (2013). Does spasticity reduction by botulinum toxin type A improve upper limb functionality in adult post-stroke patients? A systematic review of relevant studies. *Journal of Neurology and Neurophysiology, 4*(4), 1-13. doi:10.4172/2155-9562.1000167

Jacob-Lloyd, H. A., Dunn, O. M., Brain, N. D., & Lamb, S. E. (2005). Effective measurement of the functional progress of stroke clients. *British Journal of Occupational Therapy, 68*(6), 253-259. doi:10.1177/030802260506800603

Jenkinson, C., Fitzpatrick, R., Crocker, H., & Peters, M. (2013). The Stroke Impact Scale: Validation in a UK setting and development of a SIS Short Form and SIS Index. *Stroke, 44*, 2532-2535. doi:10.1161/STROKEAHA.113.001847

Joanna Briggs Institute. (2017). JBI Critical Appraisal Checklist for Qualitative Research. Retrieved from https://joannabriggs.org/sites/default/files/2020-08/Checklist_for_Qualitative_Research.pdf

Johnson, G., & Pandyan, A. (2008). The measurement of spasticity. In M. Barnes & G. Johnson (Eds.), *Upper Motor Neurone Syndrome and Spasticity: Clinical Management and Neurophysiology*. Cambridge: Cambridge University Press.

Johnson, L., & Selfe, J. (2004). Measurement of mobility following stroke: a comparison of the Modified Rivermead Mobility Index and the Motor Assessment Scale. *Physiotherapy, 90*(3), 132-138. doi:10.1016/j.physio.2004.01.004

Jones, F. (1998). The accuracy of predicting functional recovery in patients following a stroke by physiotherapists and patients. *Physiotherapy Research International, 3*(4), 244 - 256. doi:10.1002/pri.148

Joyce, B. M., Rockwood, K. J., & Mate-Kole, C. C. (1994). Use of goal attainment scaling in brain injury in a rehabilitation hospital *American Journal of Physical Medicine and Rehabilitation, 73*(1), 10-14. doi:10.1097/00002060-199473010-00003

Kavanagh, A. M., Krnjacki, L., Aitken, Z., Lamontagne, A. D., Beer, A., Baker, E., & Bentley, R. (2015). Intersections between disability, type of impairment, gender and socio-economic disadvantage in a nationally representative sample of 33,101 working-aged Australians. *Disability and Health Journal, 8*(2), 191-199. doi:10.1016/j.dhjo.2014.08.008

Keith, R. A., Granger, C. V., Hamilton, B. B., & Sherwin, F. S. (1987). The functional independence measure: a new tool for rehabilitation. *Advanced Clinical Rehabilitation, 1*, 6-18.

Kellor, M., Frost, J., Silberberg, N., Iversen, I., & Cummings, R. (1971). Norms for clinical use: Hand strength and dexterity. *The American Journal of Occupational Therapy, 25*(2), 77-83.

Kennedy, C. A., Beaton, D. E., Smith, P., Van eerd, D., Tang, K., Inrig, T., . . . Couban, R. (2013).

Measurement properties of the QuickDASH (Disabilities of the Arm, Shoulder and Hand) outcome measure and cross-cultural adaptations of the QuickDASH: a systematic review. *Quality of Life Research, 22*, 2509-2547. doi:10.1007/s11136-013-0362-4

Kerstens, H. C. J. W., Satink, T., Nijkrake, M. J., De Swart, B. J. M., Nijhuis-van Der Sanden, M. W. G., Van Der Wees, P. J., & Geurts, A. C. H. (2020). Experienced consequences of spasticity and effects of botulinum toxin injections: a qualitative study amongst patients with disabling spasticity after stroke. *Disability and Rehabilitation* (ahead-of-print), 1-8. doi:10.1080/09638288.2020.1746843

Khan, A., Chien, C. W., & Brauer, S. G. (2013). Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. *Journal of Clinical Epidemiology, 66*, 681-687. doi:10.1016/j.jclinepi.2012.12.014

Khan, F., Pallant, J., & Turner-Stokes, L. (2008). Use of goal attainment scaling in inpatient rehabilitation for persons with multiple sclerosis. *Archives of Physical Medicine and Rehabilitation, 89*(4), 652-659. doi:10.1016/j.apmr.2007.09.049

Kilbreath, S. L., & Heard, R. C. (2005). Frequency of hand use in healthy older persons. *Australian Journal of Physiotherapy, 51*(2), 119-122. doi:10.1016/S0004-9514(05)70040-4

Kinnear, B. Z., Lannin, N. A., Cusick, A., Harvey, L. A., & Rawicki, B. (2014). Rehabilitation therapies after botulinum toxin-A injection to manage limb spasticity: a systematic review. *Physical Therapy, 94*(11), 1569. doi:10.2522/ptj.20130408

Kiresuk, T., & Sherman, R. (1968). Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community Mental Health Journal, 4*(6), 443-453. doi:10.1007/BF01530764

Kirkevold, M. (2002). The unfolding illness trajectory of stroke. *Disability and Rehabilitation, 24*(17), 887-898. doi:10.1080/09638280210142239

Kitt, C., Wang, V., Harvey-Fitzgerald, L., Kayes, N., & Saywell, N. (2016). Gaining perspectives of people with stroke, to inform development of a group exercise programme: A qualitative study. *New Zealand Journal of Physiotherapy, 44*(1), 58-64. doi:10.15619/NZJP/44.1.07

Klein, R. M., & Bell, B. (1982). Self-care skills: Behavioural measurement with the Klein-Bell ADL Scale. *Archives of Physical Medicine and Rehabilitation, 63*, 335-338.

Koh, C.-L., Hsueh, I.-P., Wang, W.-C., Sheu, C.-F., Yu, T.-Y., Wang, C.-H., & Hsieh, C.-L. (2006). Validation of the Action Research Arm Test using item response theory in patients after stroke. *Journal of Rehabilitation Medicine, 38*, 375-380. doi:10.1080/16501970600803252

Kohn, C. G., Sidovar, M. F., Kaur, K., Zhu, Y., & Coleman, C. I. (2014). Estimating a minimal clinically important difference for the EuroQol 5-dimension health status index in persons with multiple sclerosis. *Health and Quality of Life Outcomes, 12*(66). doi:10.1186/1477-7525-12-66

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155-163. doi:10.1016/j.jcm.2016.02.012

Kuspinar, A., Finch, L. E., Pickard, S., & Mayo, N. E. (2014). Using existing data to identify candidate items for a health state classification system in multiple sclerosis. *Quality of Life Research, 23*, 1445-1457. doi:10.1007/s11136-013-0604-5

Kuspinar, A., & Mayo, N. E. (2013). Do generic utility measures capture what is important to the quality of life of people with multiple sclerosis? *Health and Quality of Life Outcomes, 11*(71). doi:10.1186/1477-7525-11-71

Kuys, S. S., Bew, P. G., Lynch, M. R., Morrison, G., & Brauer, S. G. (2009). Measures of activity limitation on admission to rehabilitation after stroke predict walking speed at discharge: an observational study. *Australian Journal of Physiotherapy, 55*(4), 265-268. doi:10.1016/S0004-9514(09)70006-6

Kwah, L. K., Harvey, L. A., Diong, J., & Herbert, R. D. (2013). Models containing age and NIHSS predict recovery of ambulation and upper limb function six months after stroke: an observational study. *Journal of Physiotherapy, 59*(3), 189-197. doi:10.1016/S1836-

9553(13)70183-8
Kwakkel, G., Lannin, N. A., Borschmann, K., English, C., Ali, M., Churilov, L., . . . Bernhardt, J. (2017). Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *International Journal of Stroke, 12*(5), 451-461. doi:10.1177/1747493017711813
Kwakkel, G., van Wegen, E. E. H., Burridge, J. H., Winstein, C. J., van Dokkum, L. E. H., Alt Murphy, M., . . . van Wijck, F. (2019). Standardized Measurement of Quality of Upper Limb Movement After Stroke: Consensus-Based Core Recommendations From the Second Stroke Recovery and Rehabilitation Roundtable. *Neurorehabilitation and Neural Repair, 33*(11), 951-958. doi:10.1177/1545968319886477
Kwon, S., Duncan, P., Studenski, S., Perera, S., Lai, S. M., & Reker, D. (2006). Measuring stroke impact with SIS: Construct validity of SIS telephone administration. *Quality of Life Research, 15*(3), 367-376. doi:10.1007/s11136-005-2292-2
Kwon, S., Hartzema, A. G., Duncan, P. W., & Min Lai, S. (2004). Disability measures in stroke: Relationship among the Barthel Index, the Functional Independence Measure, and the Modified Rankin Scale. *Stroke, 35*(4), 918-923. doi:10.1161/01.STR.0000119385.56094.32
Lai, S. M., Studenski, S., Duncan, P. W., & Perera, S. (2002). Persisting consequences of stroke measured by the Stroke Impact Scale. *Stroke, 33*(7), 1840-1844. doi:10.1161/01.STR.0000019289.15440.F2
Lamers, I., Kelchtermans, S., Baert, I., & Feys, P. (2014). Upper Limb Assessment in Multiple Sclerosis: A Systematic Review of Outcome Measures and their Psychometric Properties. *Archives of Physical Medicine and Rehabilitation, 95*(6), 1184-1200. doi:10.1016/j.apmr.2014.02.023
Lance, J. W. (1980). Symposium synopsis. In Feldman RG, Young RR, & K. WP (Eds.), *Spasticity: Disordered Motor Control*. Miami: Symposia Specialists.
Lang, C. E., Edwards, D. F., Birkenmeier, R. L., & Dromerick, A. W. (2008a). Estimating minimal clinically important differences of upper-extremity measures early after stroke. *Archives of Physical Medicine and Rehabilitation, 89*(9), 1693-1700. doi:10.1016/j.apmr.2008.02.022
Lang, C. E., Edwards, D. F., Birkenmeier, R. L., & Dromerick, A. W. (2008b). Estimating minimal clinically important differences of upper-extremity measures early after stroke. *Archives of Physical Medicine and Rehabilitation, 89*, 1693-1700.
Lang, C. E., Wagner, J. M., Dromerick, A. W., & Edwards, D. F. (2006). Measurement of upper-extremity function early after stroke: Properties of the Action Research Arm Test. *Archives of Physical Medicine and Rehabilitation, 87*(12), 1605-1610. doi:10.1016/j.apmr.2006.09.003
Lannin, N. (2003). Goal attainment scaling allows program evaluation of a home-based occupational therapy program. *Occupational Therapy in Health Care, 17*(1), 43-54. doi:10.1080/J003v17n01_04
Lannin, N. A. (2004). Reliability, validity and factor structure of the upper limb subscale of the Motor Assessment Scale (UL-MAS) in adults following stroke. *Disability and Rehabilitation, 26*(2), 109-115. doi:10.1080/0963828032000157970
Lannin, N. A., Ada, L., English, C., Ratcliffe, J., & Crotty, M. (2018). Effect of adding upper limb rehabilitation to botulinum toxin-A on upper limb activity after stroke: Protocol for the InTENSE trial. *International Journal of Stroke, 13*(6), 648-653. doi:10.1177/1747493018765228
Lannin, N. A., Ada, L., English, C., Ratcliffe, J., Faux, S. G., Palit, M., . . . Coulter. (2020). Effect of Additional Rehabilitation After Botulinum Toxin-A on Upper Limb Activity in Chronic Stroke: The InTENSE Trial. *Stroke, 51*(2), 556-562. doi:10.1161/STROKEAHA.119.027602
Law, M., Baptiste, S., McColl, M., Opzoomer, A., Polatajko, H., & Pollock, N. (1990). The Canadian Occupational Performance Measure: An Outcome Measure for Occupational Therapy. *Canadian Journal of Occupational Therapy, 57*(2), 82-87.

doi:10.1177/000841749005700207
Lawrence, S. E., Coshall, G. C., Dundas, D. A. R., Stewart, D. A. J., Rudd, D. A. A., Howard, D. A. R., & Wolfe, D. A. C. (2001). Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. *Stroke: Journal of the American Heart Association, 32*(6), 1279-1284. doi:10.1161/01.STR.32.6.1279
Lees, K. R. (2009). Evidence-Based Stroke Rehabilitation: An Expanded Guidance Document from the European Stroke Organisation (ESO) Guidelines for Management of Ischaemic Stroke and Transient Ischaemic Attack 2008. *Journal of Rehabilitation Medicine, 41*(2), 99-111. doi:10.2340/16501977-0301
Lemberg, I., Kirchberger, I., Stucki, G., & Cieza, A. (2010). The ICF Core Set for stroke from the perspective of physicians: a worldwide validation study using the Delphi technique. *European Journal of Physical and Rehabilitation Medicine, 46*(3), 377-388.
Lesko, L. J., Zineh, I., & Huang, S. M. (2010). What Is Clinical Utility and Why Should We Care? *Clinical Pharmacology and Therapeutics, 88*(6), 729. doi:10.1038/clpt.2010.229
Levack, W., Weatherall, M., Hay-Smith, E., Dean, S. G., McPherson, K., & Siegert, R. (2015). Goal setting and strategies to enhance goal pursuit for adults with acquired disability participating in rehabilitation. *Cochrane Database Syst Rev., 2015*(7). doi:10.1002/14651858.CD009727.pub2
Levack, W. M. M., Dean, S. G., McPherson, K. M., & Siegert, R. J. (2014). Evidence-Based goal setting: Cultivating the science of rehabilitation. In R. J. Siegert & W. M. M. Levack (Eds.), *Rehabilitation Goal Setting: Theory, Practice and Evidence*: CRC Press.
Levack, W. M. M., & Siegert, R. J. (2014). Challenges in theory, practice and evidence. In R. J. Siegert & W. M. M. Levack (Eds.), *Rehabilitation Goal Setting: Theory, Practice and Evidence*: CRC Press.
Levack, W. M. M., Taylor, K., Siegert, R. J., Dean, S. G., McPherson, K. M., & Weatherall, M. (2006). Is goal planning in rehabilitation effective? A systematic review. *Clinical Rehabilitation, 20*(9), 739-755. doi:10.1177/0269215506070791
Levy, T., J. Christie, L., Killington, M., Laver, K., Crotty, M., & Lannin, N. A. (2021). "Just that four letter word, hope": stroke survivors' perspectives of participation in an intensive upper limb exercise program; a qualitative exploration. *Physiotherapy Theory and Practice*, 1-15. doi:10.1080/09593985.2021.1875525
Li, S. (2017). Spasticity, motor recovery, and neural plasticity after stroke. *Frontiers in Neurology,* 8, 1-8. doi: 10.3389/fneur.2017.00120
Limsakul, C., Noten, S., Selb, M., Stam, H. J., van Der Slot, W. M. A., & Roebroeck, M. E. (2020). Developing an ICF Core Set for adults with cerebral palsy: A Global expert survey of relevant functions and contextual factors. *Journal of Rehabilitation Medicine, 52*(4), jrm00049-jrm00049. doi:10.2340/16501977-2669
Lin, J.-H., Hsu, M.-J., Sheu, C.-F., Wu, T.-S., Lin, R.-T., Chen, C.-H., & Hsieh, C.-L. (2009). Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. *Physical Therapy, 89*(8), 840-850. doi:10.2522/ptj.20080285
Lin, K.-C., Chuang, L.-L., Wu, C.-Y., Hsieh, Y.-W., & Chang, W.-Y. (2010). Responsiveness and validity of three dexterous function measures in stroke rehabilitation. *Journal of Rehabilitation Research & Development, 47*(6), 563-572. doi:10.1682/JRRD.2009.09.0155
Lincoln, N., & Leadbitter, D. (1979). Assessment of Motor Function in Stroke Patients. *Physiotherapy, 65*(2), 48-51.
Lloyd, A., Bannigan, K., Sugavanam, T., & Freeman, J. (2018). Experiences of stroke survivors, their families and unpaid carers in goal setting within stroke rehabilitation: a systematic review of qualitative evidence. *JBI Database of Systematic Reviews and Implementation Reports, 16*(6), 1418-1453. doi:10.11124/JBISRIR-2017-003499
Loewen, S. C., & Anderson, B. A. (1988). Reliability of the Modified Motor Assessment Scale and the Barthel Index. *Physical Therapy, 68*(7), 1077-1081.
Loewen, S. C., & Anderson, B. A. (1990). Predictors of stroke outcome using objective measurement scales. *Stroke, 21*(1), 78-81. doi:10.1161/str.21.1.2300994

Lohmann, S., Decker, J., Müller, M., Strobl, R., & Grill, E. (2011). The ICF Forms a Useful Framework for Classifying Individual Patient Goals in Post-acute Rehabilitation. *Journal of Rehabilitation Medicine, 43*(2), 151-155. doi:10.2340/16501977-0657

Lou, S., Carstensen, K., Jørgensen, C. R., & Nielsen, C. P. (2017). Stroke patients' and informal carers' experiences with life after stroke: an overview of qualitative systematic reviews. *Disability and Rehabilitation, 39*(3), 301-313. doi:10.3109/09638288.2016.1140836

Luker, J., Lynch, E., Bernhardsson, S., Bennett, L., & Bernhardt, J. (2015). Stroke survivors' experiences of physical rehabilitation: a systematic review of qualitative studies. *Archives of Physical Medicine and Rehabilitation, 96*(9), 1698-1708.e1610. doi:10.1016/j.apmr.2015.03.017

Lund, A., Mangset, M., Wyller, T. B., & Sveen, U. (2015). Occupational transaction after stroke constructed as threat and balance. *Journal of Occupational Science: Special Issue on Occupation, Health and Well-being, 22*(2), 146-159. doi:10.1080/14427591.2013.770363

Lyden, P., Lu, M., Jackson, C., Marler, J., Kothari, R., Brott, T., & Zivin, J. (1999). Underlying structure of the National Institutes of Health Stroke Scale: results of a factor analysis. *Stroke, 30*(11), 2347-2354. doi:10.1161/01.STR.30.11.2347

Lyle, R. C. (1981). A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *International Journal of Rehabilitation Research, 4*(4), 483-492. doi:10.1097/00004356-198112000-00001

Mackenzie, E. J., McCarthy, M. L., Ditunno, J. F., Forrester-Staz, C. S., Gruen, G. W., Marion, D. C., & Schwab, W. C. (2002). Using the SF-36 for characterizing outcome after multiple trauma Involving head injury. *The Journal of Trauma: Injury, Infection, and Critical Care, 52*(3), 527-534. doi:10.1097/00005373-200203000-00018

Madden, S., Hopman, W. M., Bagg, S., Verner, J., & O'Callaghan, C. J. (2006). Functional status and health-related quality of life during inpatient stroke rehabilitation. *American Journal of Physical Medicine and Rehabilitation, 85*(10), 831-838. doi:10.1097/01.phm.0000240666.24142.f7

Mahoney, F. I., & Barthel, D. W. (1965). Functional evaluation: The Barthel Index. *Maryland State Medical Journal, 14*, 61-65.

Malec, J. F. (1999). Goal attainment scaling in rehabilitation. *Neuropsychological Rehabilitation, 9*(3/4), 253-275. doi:10.1080/096020199389365

Malec, J. F., Smigielski, J. S., & DePompolo, R. W. (1991). Goal attainment scaling and outcome measurement in postacute brain injury rehabilitation. *Archives of Physical medicine and Rehabilitation, 72*(2), 138-143. doi:10.5555/uri:pii:000399939190122Y

Malhotra, S., Cousins, E., Ward, A., Day, C., Jones, P., Roffe, C., & Pandyan, A. (2008). An investigation into the agreement between clinical, biomechanical and neurophysiological measures of spasticity. *Clinical Rehabilitation, 22*(12), 1105-1115. doi:10.1177/0269215508095089

Malhotra, S., Pandyan, A., Day, C., Jones, P., & Hermens, H. (2009). Spasticity, an impairment that is poorly defined and poorly measured. *Clinical Rehabilitation, 23*(7), 651-658. doi:10.1177/0269215508101747

Manning, M., MacFarlane, A., Hickey, A., & Franklin, S. (2019). Perspectives of people with aphasia post-stroke towards personal recovery and living successfully: A systematic review and thematic synthesis. *PLoS One, 14*(3), e0214200. doi:10.1371/journal.pone.0214200

Mathiowetz, V., Volland, G., Kashman, N., & Weber, K. (1985). Adult norms for the Box and Block Test of manual dexterity. *The American Journal of Occupational Therapy, 39*(6), 386-391.

Mathiowetz, V., Weber, K., Kashman, N., & Volland, G. (1985). Adult norms for the nine hole peg test of finger dexterity. *The Occupational Therapy Journal of Research, 5*(1), 24-38. doi:10.1177/153944928500500102

McCluskey, A., & O'Connor, D. (2017). Implementing evidence: closing the research-practice gap. In T. Hoffmann, S. Bennett, & C. Del Mar (Eds.), *Evidence-Based Practice Across the Health Professions* (3rd ed.). Australia: Elsevier.

McCulloch, K. L., Cook, E. W. I., Fleming, W. C., Novack, T. A., & Taub, E. (1988). A reliable test of upper extremity ADL function. *Archives of Physical Medicine and Rehabilitation, 69*, 755.

McDonnell, M. N., Hillier, S. L., Ridding, M. C., & Miles, T. S. (2006). Impairments in precision grip correlate with functional measures in adult hemiplegia. *Clinical Neurophysiology, 117*, 1474-1480. doi:10.1016/j.clinph.2006.02.027

Meyer, S., Karttunen, A. H., Thijs, V., Feys, H., & Verheyden, G. (2014). How do somatosensory deficits in the arm and hand relate to upper limb impairments in activity, and participation problems after stroke? A systematic review. *Physical Therapy, 94*(9), 1220-1231. doi:10.2522/ptj.20130271

Miller, K. J., Slade, A. L., Pallant, J. F., & Galea, M. P. (2010). Evaluation of the psychometric properties of the upper limb subscales of the motor assessment scale using a rasch analysis model. *Journal of Rehabilitation Medicine, 42*(4), 315-322. doi:10.2340/16501977-0519

Miller, K. K., Combs, S. A., van Puymbroeck, M., Altenburger, P. A., Kean, J., Dierks, T. A., & Schmid, A. A. (2013). Fatigue and Pain: Relationships with physical performance and patient beliefs after stroke. *Topics in Stroke Rehabilitation, 20*(4), 347-355. doi:10.1310/tsr2004-347

Mokkink, L., Prinsen, C. A. C., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy, 20*(2), 105. doi:10.1590/bjpt-rbf.2014.0143

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias Checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Outcomes, 27*, 1171-1179. doi:10.1007/211136-017-1765-4

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., . . . de Vet, H. C. W. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology, 10*, 22-22. doi:10.1186/1471-2288-10-22

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. W. (2010). International consensus on taxonomy, terminology, and definitions of measurement properties: results of the COSMIN study. *Journal of Clinical Epidemiology, 63*, 737-745. doi:10.1016/j.jclinepi.2010.02.006

Moore, F., Wolfson, C., Alexandrov, L., & Lapierre, Y. (2004). Do general and multiple sclerosis-specific quality of life instruments differ? *Canadian Journal of Neurological Sciences, 31*(1), 64-71. doi:10.1017/S0317167100002857

Moreland, J., Gowland, C., Van Hullenaar, S., & Huijbregts, M. (1993). Theoretical basis of the Chedoke-McMaster Stroke Assessment. *Physiotherapy Canada, 45*(4), 231-238.

Morris, J. H., van Wijck, F., Joice, S., & Donaghy, M. (2013). Predicting health related quality of life 6 months after stroke: the role of anxiety and upper limb dysfunction. *Disability and Rehabilitation, 35*(4), 291-299. doi:10.3109/09638288.2012.691942

Mudge, S., & Stott, N. S. (2009). Timed walking tests correlate with daily step activity in persons with stroke. *Archives of Physical Medicine and Rehabilitation, 90*(2), 296 - 301. doi:0.1016/j.apmr.2008.07.025

Murrell, R. C., Kenealy, P. M., Graham Beaumont, J., & Lintern, T. C. (1999). Assessing quality of life in persons with severe neurological disability associated with multiple sclerosis: The psychometric evaluation of two quality of life. measures. *British Journal of Health Psychology, 4*(4), 349-362. doi:10.1348/135910799168696

Nakayama, H., Jørgensen, H. S., Raaschou, H. O., & Olsen, T. S. (1994). Recovery of upper extremity function in stroke patients: the Copenhagen Stroke Study. *Archives of Physical Medicine and Rehabilitation, 75*(4), 394-398. doi:10.1016/0003-9993(94)90161-9

National Clinical Guideline Centre (NICE). (2013). *Stroke Rehabilitation: Long-Term Rehabilitation After Stroke.* London: NICE, Clinical Guideline no. 162

Nguyen, R., Fiest, K. M., McChesney, J., Kwon, C.-S., Jette, N., Frolkis, A. D., . . . Gallagher, C. (2016). The International Incidence of Traumatic Brain Injury: A Systematic Review and Meta-Analysis. *Can. J. Neurol. Sci, 43*(6), 774-785. doi:10.1017/cjn.2016.290

Nicholl, C. R., Lincoln, N. B., Francis, V. M., & Stephan, T. F. (2001). Assessing quality of life in people with multiple sclerosis. *Disability and Rehabilitation, 23*(14), 597-603. doi:10.1080/09638280110043933

Notley, S. V., Turk, R., Pickering, R. M., Simpson, D. M., & Burridge, J. H. (2007). Analysis of the quality of wrist movement during a simple tracking task. *Physiological Measurement, 28*, 881-895. doi:10.1088/0967-3334/28/8/011

Nott, M., Barden, H., & Baguley, I. (2014). Goal attainment following upper-limb botulinum toxin-A injections: Are we facilitating achievement of client-centred goals? *Journal of Rehabilitation Medicine, 46*(9), 864-868. doi:10.2340/16501977-1853

O'Dell, M. W., Kim, G., Rivera, L., Fieo, R., Christos, P., Polistena, C., . . . Gorga, D. (2013). A psychometric evaluation of the Arm Motor Mobility Test. *Journal of Rehabilitation Medicine, 45*, 519-527. doi:10.2340/16501977-1138

O'Mahony, P. G., Rodgers, H., Thomson, R. G., Dobson, R., & James, O. F. W. (1998). Is the SF-36 suitable for assessing health status of older stroke patients? *Age and Ageing, 27*(1), 19-22. doi:10.1093/ageing/27.1.19

Oczkowski, W., & Barreca, S. (1993). The functional independence measure: Its use to identify rehabilitation needs in stroke survivors. *Archives of Physical Medicine and Rehabilitation, 74*(12), 1291-1294. doi:10.1016/0003-9993(93)90081-K

Opheim, C. A., Danielsson, S. A., Alt Murphy, S. M., Persson, S. H., & Sunnerhagen, S. K. (2014). Upper-Limb Spasticity During the First Year After Stroke: Stroke Arm Longitudinal Study at the University of Gothenburg. *American Journal of Physical Medicine and Rehabilitation, 93*(10), 884-896. doi:10.1097/PHM.0000000000000157

Ottawa Panel. (2006). Ottawa Panel Evidence-Based Clinical Practice Guidelines for Post-Stroke Rehabilitation. *Topics in Stroke Rehabilitation, 13*(2), 1-269. doi:10.1310/3TKX-7XEC-2DTG-XQKH

Ouellette, D. S., Timple, C., Kaplan, S. E., Rosenberg, S. S., & Rosario, E. R. (2015). Predicting discharge destination with admission outcome scores in stroke patients. *NeuroRehabilitation, 37*(2), 173-179. doi:10.3233/NRE-151250

Paanalahti, M., Alt Murphy, M., Lundgren-Nilsson, Å., & Sunnerhagen, K. S. (2014). Validation of the Comprehensive ICF Core Set for stroke by exploring the patient's perspective on functioning in everyday life: a qualitative study. *International Journal of Rehabilitation Research, 37*(4), 302-310. doi:10.1097/MRR.0000000000000070

Paanalahti, M., Lundgren-Nilsson, A., Arndt, A., & Sunnerhagen, K. S. (2013). Applying the Comprehensive International Classification of Functioning, Disability and Health Core Sets for stroke framework to stroke survivors living in the community. *Journal of Rehabilitation Medicine, 45*(4), 331-340. doi:10.2340/16501977-1110

Page, S. J., Hade, E., & Persch, A. (2015). Psychometrics of the wrist stability and hand mobility subscales of the Fugl-Meyer assessment in moderately impaired stroke. *Physical Therapy, 95*(1), 103-108. doi:10.2522/ptj.20130235

Page, S. J., Levine, P., & Hade, E. (2012). Psychometric properties and administration of the Wrist/Hand Subscales of the Fugl-Meyer assessment in minimally impaired upper extremity hemiparesis in stroke. *Archives of Physical Medicine and Rehabilitation, 93*, 2373-2376. doi:10.1016/j.apmr.2012.06.017

Pallesen, H. (2014). Body, coping and self-identity. A qualitative 5-year follow-up study of stroke. *Disability and Rehabilitation, 36*(3), 232-241. doi:10.3109/09638288.2013.788217

Pandyan, A., Gregoric, M., Barnes, M., Wood, D., van Wijck, F., Burridge, J., . . . Johnson, G. (2005). Spasticity: Clinical perceptions, neurological realities and meaningful measurement. *Disability and Rehabilitation, 27*(1-2), 2-6. doi:10.1080/09638280400014576

Pandyan, A., Hermens, H., Conway, B., & Johnson, G. (2018). Definition and measurement of

spasticity and contracture. In A. D. Pandyan, H. J. Hermens, & B. A. Conway (Eds.), *Neurological Rehabilitation: Spasticity and Contracture in Clinical Practice and Research* Boca Raton: CRC Press.

Patel, A. T., Wein, T., Bahroo, L. B., Wilczynski, O., Rios, C. D., & Murie-Fernández, M. (2020). Perspective of an International Online Patient and Caregiver Community on the Burden of Spasticity and Impact of Botulinum Neurotoxin Therapy: Survey Study. *JMIR Public Health and Surveillance, 6*(4), e17928-e17928. doi:10.2196/17928

Patrick, E., & Ada, L. (2006). The Tardieu Scale differentiates contracture from spasticity whereas the Ashworth Scale is confounded by it. *Clinical Rehabilitation, 20*(2), 173-182. doi:10.1191/0269215506cr922oa

Peoples, H., Satink, T., & Steultjens, E. (2011). Stroke survivors' experiences of rehabilitation: A systematic review of qualitative studies. *Scandinavian Journal of Occupational Therapy, 18*(3), 163-171. doi:10.3109/11038128.2010.509887

Peters, M., Crocker, H., Dummett, S., Jenkinson, C., Doll, H., & Fitzpatrick, R. (2014). Change in health status in long-term conditions over a one year period: a cohort survey using patient-reported outcome measures. *Quality of Life Outcomes, 12*(1). doi:10.1186/s12955-014-0123-2

Pfeiffer, E. (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society, 23*(10), 433-441.

Pickard, A. S., Johnson, J. A., & Feeny, D. H. (2005). Responsiveness of generic health-related quality of life measures in stroke. *Quality of Life Research, 14*(1), 207-219. doi:10.1007/s11136-004-3928-3

Pickering, R. L., Hubbard, I. J., Baker, K. G., & Parsons, M. W. (2010). Assessment of the upper limb in acute stroke: The validity of hierarchical scoring for the Motor Assessment Scale. *Australian Occupational Therapy Journal, 57*(3), 174-182. doi:10.1111/j.1440-1630.2009.00810.x

Pike, S., Lannin, N. A., Cusick, A., Wales, K., Turner-Stokes, L., & Ashford, S. (2015). A systematic review protocol to evaluate the psychometric properties of measures of function within adult neuro-rehabilitation. *Systematic Reviews, 4*(1), 86. doi:10.1186/s13643-015-0076-5

Pittock, S. J., Mayr, W. T., McClelland, R. L., Jorgensen, N. W., Weigand, S. D., Noseworthy, J. H., & Rodriguez, M. (2004). Quality of life is favorable for most patients with multiple sclerosis. *Archives of Neurology, 61*(5), 679-686. doi:10.1001/archneur.61.5.679

Plant, S. E., Tyson, S. F., Kirk, S., & Parsons, J. (2016). What are the barriers and facilitators to goal-setting during rehabilitation for stroke and other acquired brain injuries? A systematic review and meta-synthesis. *Clinical Rehabilitation, 30*(9), 921-930. doi:10.1177/0269215516655856

Platz, T., Eickhof, C., Nuyens, G., & Vuadens, P. (2005). Clinical scales for the assessment of spasticity, associated phenomena, and function: a systematic review of the literature. *Disability and Rehabilitation, 27*(1-2), 7-18. doi:10.1080/09638280400014634

Platz, T., Pinkowski, C., van Wijck, F., & Johnson, G. (2005). *Arm - Arm Rehabilitation Measurement: Manual for Performance and Scoring of Fugl-Meyer test, Action Research Arm Test and Box and Block test*. Baden-Baden: Deutscher Wissenschafts-Verlag.

Platz, T., Pinkowski, C., van Wijck, F., Kim, I. H., di Bella, P., & Johnson, G. (2005). Reliability and validity of arm function assessment with standardised guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. *Clinical Rehabilitation, 19*(4), 404-411. doi:10.1191/0269215505cr832oa

Playford, E. D. (2020). The International Classification of Functioning, Disability and Health. In V. Dietz & N. S. Ward (Eds.), *Oxford Textbook of Neurorehabilitation* (2 ed.): Oxford University Press.

Playford, E. D., Siegert, R., Levack, W., & Freeman, J. (2009). Areas of consensus and controversy about goal setting in rehabilitation: a conference report. *Clinical Rehabilitation, 23*(4), 334-344. doi:10.1177/0269215509103506

Ponsford, J. (1995). Mechanisms, recovery, and sequelae of traumatic brain injury: a foundation for the REAL approach In J. Ponsford, S. Sloan, & P. Snow (Eds.), *Traumatic brain injury Rehabilitation for Everyday Adaptive Living*. New York: Psychology Press.

Ponsford, J. L., Downing, M. G., Olver, J., Ponsford, M., Acher, R., Carty, M., & Spitz, G. (2014). Longitudinal Follow-Up of Patients with Traumatic Brain Injury: Outcome at Two, Five, and Ten Years Post-Injury. *Journal of Neurotrauma, 31*(1), 64-77. doi:10.1089/neu.2013.2997

Poole, J. L., Nakamoto, T. N., McNulty, T., Montoya, J. R., Weill, D., Dieruf, K., & Skipper, B. (2010). Dexterity, visual perception, and activities of daily living in persons with multiple sclerosis. *Occupational Therapy In Health Care, 24*(2), 159- 170. doi:10.3109/07380571003681202

Pozzato, I., Tate, R. L., Rosenkoetter, U., & Cameron, I. D. (2019). Epidemiology of hospitalised traumatic brain injury in the state of New South Wales, Australia: a population-based study. *Australian & New Zealand Journal of Public Health, 43*(4), 382-388. doi:10.1111/1753-6405.12878

Prescott, S., Fleming, J., & Doig, E. (2015). Goal setting approaches and principles used in rehabilitation for people with acquired brain injury: A systematic scoping review. *Brain Injury, 29*(13-14), 1515-1529. doi:10.3109/02699052.2015.1075152

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147-1157. doi:10.1007/s11136-018-1798-3

Purton, J., Sim, J., & Hunter, S. M. (2020). The experience of upper-limb dysfunction after stroke: a phenomenological study. *Disability and Rehabilitation*, 1-10. doi:10.1080/09638288.2020.1743775

QSR International. (2018). NVivo qualitative data analysis software (Version 12).

Rabadi, M. H., & Rabadi, F. M. (2006a). Comparison of the Action Research Arm Test and the Fugl-Meyer Assessment as measures of upper-extremity motor weakness after stroke *Archives of Physical Medicine and Rehabilitation, 87*(7), 962-966. doi:10.1016/j.apmr.2006.02.036

Rabadi, M. H., & Rabadi, F. M. (2006b). Comparison of the Action Research Arm Test and the Fugl-Meyer Assessment as measures of upper-extremity motor weakness after stroke *Archives of Physical Medicine and Rehabilitation, 87*, 962-966.

Rabadi, M. H., & Vincent, A. S. (2013). Comparison of the Kurtkze Expanded Disability Status Scale and the Functional Independence Status Scale and the Functional Independence Measure: measures of multiple sclerosis-related disability. *Disability and Rehabilitation, 35*(22), 1877-1884. doi:10.3109/09638288.2013.766269

Rand, D., & Eng, J. J. (2015). Predicting daily use of the affected upper extremity 1 year after stroke. *Journal of Stroke and Cerebrovascular Diseases, 24*(2), 274-283. doi:10.1016/j.strokecerebrovasdis.2014.07.039

Riazi, A., Hobart, J. C., Lamping, D. L., Fitzpatrick, R., Freeman, J. A., Jenkinson, C., . . . Thompson, A. J. (2003). Using the SF-36 measure to compare the health impact of multiple sclerosis and Parkinson's disease with normal population health profiles. *Journal of Neurology, Neurosurgery & Psychiatry, 74*(6), 710-714. doi:10.1136/jnnp.74.6.710

Riberto, M., Lopes, K. A. T., Chiappetta, L. M., Lourenção, M. I. P., & Battistella, L. R. (2013). The use of the comprehensive International Classification of Functioning, Disability and Health core set for stroke for chronic outpatients in three Brazilian rehabilitation facilities. *Disability and Rehabilitation, 35*(5), 367-374. doi:10.3109/09638288.2012.694573

Rigby, H., Gubitz, G., Eskes, G., Reidy, Y., Christian, C., Grover, V., & Phillips, S. (2009). Caring for stroke survivors: baseline and 1-year determinants of caregiver burden. *International Journal of Stroke, 4*(3), 152 -158. doi:10.1111/j.1747-4949.2009.00287.x

Robinson Jr, D., Zhao, N., Gathany, T., Kim, L.-L., Cella, D., & Revicki, D. (2009). Health

perceptions and clinical characteristics of relapsing-remitting multiple sclerosis patients: baseline data from an international clinical trial. *Current Medical Research and Opinion., 25*(5), 1121-1130. doi:10.1185/03007990902797675

Rockwood, K., Howlett, S., Stadnyk, K., Carver, D., Powell, C., & Stolee, P. (2003). Responsiveness of goal attainment scaling in a randomized controlled trial of comprehensive geriatric assessment. *Journal of Clinical Epidemiology, 56*(8), 736-743. doi:10.1016/S0895-4356(03)00132-X

Rose, A., Rosewilliam, S., & Soundy, A. (2017). Shared decision making within goal setting in rehabilitation settings: A systematic review. *Patient Education and Counseling, 100*(1), 65-75. doi:10.1016/j.pec.2016.07.030

Rosenkoetter, U., & Tate, R. L. (2018). Assessing Features of Psychometric Assessment Instruments: A Comparison of the COSMIN Checklist with Other Critical Appraisal Tools. *Brain Impairment, 19*(1), 103-118. doi:10.1017/BrImp.2017.29

Rosewilliam, S., Roskell, C. A., & Pandyan, A. D. (2011). A systematic review and synthesis of the quantitative and qualitative evidence behind patient-centred goal setting in stroke rehabilitation. *Clinical Rehabilitation, 25*(6), 501-514. doi:10.1177/0269215510394467

Royal College of Physicians, British Society of Rehabilitation Medicine, Chartered Society of Physiotherapy, & Association of Chartered Physiotherapists Interested in Neurology. (2009). *Spasticity in Adults: Management Using Botulinum Toxin. National guidelines*. London: Royal College of Physicians.

Royal College of Physicians, British Society of Rehabilitation Medicine, The Chartered Society of Physiotherapy, Association of Chartered Physiotherapists in Neurology, & Royal College of Occupational Therapists. (2018). *Spasticity in adults: management using botulinum toxin. National guidelines*. London: Royal College of Physicians.

Sabari, J. S., Lim, A. L., Velozo, C. A., Lehman, L., Kieran, O., & Lai, J. S. (2005). Assessing arm and hand function after stroke: a validity test of the hierarchical scoring system used in the motor assessment scale for stroke. *Archives of Physical Medicine and Rehabilitation, 86*(8), 1609-1615. doi:10.1016/j.apmr.2004.12.028

Sackley, C. M. (1990). The relationship between weight-bearing asymmetry after stroke, motor function and activities of daily living. *Physiotherapy Theory and Practice, 6*(4), 179-185. doi:10.3109/09593989009048293

Sakzewski, L., Boyd, R., & Ziviani, J. (2007). Clinimetric properties of participation measures for 5- to 13-year-old children with cerebral palsy: a systematic review. *Developmental Medicine & Child Neurology, 49*(3), 232-240. doi:10.1111/j.1469-8749.2007.00232.x

Salbach, N. M., Mayo, N. E., Higgins, J., Ahmed, S., Finch, L. E., & Richards, C. L. (2001). Responsiveness and predictability of gait speed and other disability measures in acute stroke. *Archives of Physical Medicine and Rehabilitation, 82*(9), 1204-12012. doi:10.1053/apmr.2001.24907

Salter, K., Hellings, C., Foley, N., & Teasell, R. (2008). The experience of living with stroke: a qualitative meta-synthesis. *Journal of Rehabilitation Medicine, 40*(8), 595-602. doi:10.2340/16501977-0238

Salter, K. L., Moses, M. B., Foley, N. C., & Teasell, R. W. (2008). Health-related quality of life after stroke: what are we measuring? *International Journal of Rehabilitation Research, 31*(2), 111-117. doi:10.1097/MRR.0b013e3282fc0f33

Santisteban, L., Teremetz, M., Bleton, J. P., Baron, J., Maier, M. A., & Lindberg, P. G. (2016). Upper Limb Outcome Measures Used in Stroke Rehabilitation Studies: A Systematic Literature Review. *PLoS One, 11*(5). doi:10.1371/journal.pone.0154792

Sarker, S.-J., Rudd, A. G., Douiri, A., & Wolfe, C. D. A. (2012). Comparison of 2 extended activities of daily living scales with the Barthel Index and predictors of their outcomes: Cohort study within the South London Stroke Register (SLSR). *Stroke, 43*(5), 1362-1369. doi:10.1161/STROKEAHA.111.645234

Sarre, S., Redlich, C., Tinker, A., Sadler, E., Bhalla, A., & McKevitt, C. (2014). A systematic review of qualitative studies on adjusting after stroke: lessons for the study of resilience.

Disability and Rehabilitation, 36(9), 716-726. doi:10.3109/09638288.2013.814724
Schenkenberg, C. T., Bradford, T. D., & Ajax, T. E. (1980). Line bisection and unilateral visual neglect in patients with neurologic impairment. *Neurology, 30*(5), 509-517.
Schmid, A. A., Van Puymbroeck, M., Altenburger, P. A., Dierks, T. A., Miller, K. K., Damush, T. M., & Williams, L. S. (2012). Balance and balance self-efficacy are associated with activity and participation after stroke: A cross-sectional study in people with chronic stroke. *Archives of Physical Medicine and Rehabilitation, 93*(6), 1101-1107. doi:10.1016/j.apmr.2012.01.020
Schwid, S. R., Goodman, A. D., McDermott, M. P., Bever, C. F., & Cook, S. D. (2002). Quantitative functional measures in MS: What is a reliable change? *Neurology, 58*(8), 1294-1296. doi:10.1212/WNL.58.8.1294
Scientific Advisory Committee of the Medical Outcomes, T. (2002). Assessing Health Status and Quality-of-Life Instruments: Attributes and Review Criteria. *Quality of Life Research, 11*(3), 193-205. doi:10.1023/A:1015291021312
Scottish Intercollegiate Guidelines Network (SIGN). (2010). Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning. A national clinical guideline. Edinburgh: SIGN. Publication 118.
Scrivener, K., Schurr, K., & Sherrington, C. (2014). Responsiveness of the ten-metre walk test, Step Test and Motor Assessment Scale in inpatient care after stroke. *BMC Neurology, 14*(1), 129. doi:10.1186/1471-2377-14-129
Sharrack, B., Hughes, R. A. C., Soudain, S., & Dunn, G. (1999). The psychometric properties of clinical rating scales used in multiple sclerosis. *Brain, 122*(1), 141-159. doi:10.1093/brain/122.1.141
Shaw, C. L., Price, I. M. C., Van Wijck, M. J. F., Shackley, P. P., Steen, A. N., Barnes, A. M., . . . Rodgers, A. H. (2011). Botulinum Toxin for the Upper Limb After Stroke (BoTULS) Trial: Effect on Impairment, Activity Limitation, and Pain. *Stroke, 42*(5), 1371-1379. doi:10.1161/STROKEAHA.110.582197
Sheean, G., Lannin, N. A., Turner-Stokes, L., Rawicki, B., & Snow, B. J. (2010). Botulinum toxin assessment, intervention and after-care for upper limb hypertonicity in adults: international consensus statement. *European Journal of Neurology, 17*, 74-93. doi:10.1111/j.1468-1331.2010.03129.x
Sheean, L. G. (2001). Botulinum treatment of spasticity: why is it so difficult to show a functional benefit? *Current Opinion in Neurology, 14*(6), 771-776. doi:10.1097/00019052-200112000-00015
Shipley, J., Luker, J., Thijs, V., & Bernhardt, J. (2018). The personal and social experiences of community-dwelling younger adults after stroke in Australia: a qualitative interview study. *BMJ Open, 8*(12). doi:10.1136/bmjopen-2018-023525
Sikorskii, A., & Noble, P. C. (2013). Statistical considerations in the psychometric validation of outcome measures. *Clinical Orthopaedics and Related Research, 471*(11), 3489-3495. doi:10.1007/s11999-013-3028-1
Simon, C., Kumar, S., & Kendrick, T. (2008). Formal support of stroke survivors and their informal carers in the community: a cohort study. *Health and Social Care in the Community, 16*(6), 582-592. doi:10.1111/j.1365-2524.2008.00782.x
Simpson, D., Gracies, J. M., Graham, H., Miyasaki, J., Naumann, M., Russman, B., & Simpson, L. (2008). Assessment: Botulinum neurotoxin for the treatment of spasticity (an evidence-based review): Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology, 70*(19), 1691-1698. doi:10.1212/01.wnl.0000311391.00944.c4
Smart, A. (2006). A multi-dimensional model of clinical utility. *International Journal for Quality in Health Care, 18*(5), 377-382. doi:10.1093/intqhc/mzl034
Smith, S. J., Ellis, E., White, S., & Moore, A. P. (2000). A double-blind placebo-controlled study of botulinum toxin in upper limb spasticity after stroke or head injury. *Clinical Rehabilitation, 14*(1), 5-13. doi:10.1191/026921500666642221

Snögren, S. M., & Sunnerhagen, S. K. (2009). Description of functional disability among younger stroke patients: exploration of activity and participation and environmental factors. *International Journal of Rehabilitation Research, 32*(2), 124-131. doi:10.1097/MRR.0b013e328325a5be

Sommerfeld, K. D., Eek, U. B. E., Svensson, W. A.-K., Holmqvist, H. L., & Von Arbin, H. M. (2004). Spasticity After Stroke: Its Occurrence and Association With Motor Impairments and Activity Limitations. *Stroke, 35*(1), 134-139. doi:10.1161/01.STR.0000105386.05173.5E

Stevens, A., Beurskens, A., Köke, A., & van Der Weijden, T. (2013). The use of patient-specific measurement instruments in the process of goal-setting: a systematic review of available instruments and their feasibility. *Clinical Rehabilitation, 27*(11), 1005-1019. doi:10.1177/0269215513490178

Stewart, C. J., & Cramer, C. S. (2013). Patient-Reported Measures Provide Unique Insights Into Motor Function After Stroke. *Stroke, 44*(4), 1111-1116. doi:10.1161/STROKEAHA.111.674671

Stewart, L., Moher, D., & Shekelle, P. (2012). Why prospective registration of systematic reviews makes sense. *Systematic Reviews, 1*(1), 7-7. doi:10.1186/2046-4053-1-7

Stier-Jarmer, M., Grill, E., Ewert, T., Bartholomeyczik, S., Finger, M., Mokrusch, T., . . . Stucki, G. (2005). ICF Core Set for patients with neurological conditions in early post-acute rehabilitation facilities. *Disability and Rehabilitation, 27*(7-8), 389-395. doi:10.1080/09638280400014022

Stinear, C., Ackerley, S., & Byblow, W. (2013). Rehabilitation is Initiated Early After Stroke, but Most Motor Rehabilitation Trials Are Not: A Systematic Review. *Stroke, 44*(7), 2039-2045. doi:10.1161/STROKEAHA.113.000968

Stinear, C. M., Barber, P. A., Petoe, M., Anwar, S., & Byblow, W. D. (2012). The PREP algorithm predicts potential for upper limb recovery after stroke. *Brain, 135*, 2527-2535. doi:10.1093/brain/aws146

Stineman, M. G., Shea, J. A., Jette, A., Tassoni, C. J., Ottenbacher, K. J., Fiedler, R., & Granger, C. V. (1996). The functional independence measure: Test of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Archives of Physical Medicine and Rehabilitation, 77*(11), 1101-1108. doi:10.1016/S0003-9993(96)90130-6

Stolee, P., Rockwood, K., Fox, R. A., & Streiner, D. L. (1992). The use of goal attainment scaling in a geriatric care setting. *Journal of the American Geriatrics Society, 40*(6), 574-578. doi:10.1111/j.1532-5415.1992.tb02105.x

Stone, S. P., Patel, P., & Greenwood, R. J. (1993). Selection of acute stroke patients for treatment of visual neglect. *Journal of Neurology, Neurosurgery & Psychiatry, 56*(5), 463-466. doi:10.1136/jnnp.56.5.463

Stroke Foundation. (2020). Clinical Guidelines for Stroke Management. Retrieved from https://informme.org.au/en/Guidelines/Clinical-Guidelines-for-Stroke-Management.

Sturm, J. W., Osborne, R. H., Dewey, H. M., Donnan, G. A., Macdonell, R. A. L., & Thrift, A. G. (2002). Brief comprehensive quality of life assessment after stroke: The assessment of quality of life instrument in the north east Melbourne stroke incidence study (NEMESIS). *Stroke, 33*(12), 2888-2894. doi:10.1161/01.STR.0000040407.44712.C7

Sugavanam, T., Mead, G., Bulley, C., Donaghy, M., & van Wijck, F. (2013a). The effects and experiences of goal setting in stroke rehabilitation - a systematic review. *Disability and Rehabilitaton, 35*(3), 177-190. doi:10.3109/09638288.2012.690501

Sullivan, J. E., Crowner, B. E., Kluding, P. M., Nichols, D., Rose, D. K., Yoshida, R., & Zipp, G. P. (2013). Outcome measures for individuals with stroke:Process and recommendations from the American physical therapy association neurology section task force *Physical Therapy, 93*(1383-1396). doi:10.2522/ptj.20120492

Sunnerhagen, S. K., Olver, E. J., & Francisco, E. G. (2013). Assessing and treating functional impairment in poststroke spasticity. *Neurology, 80*(3, Supp 2), S35-S44. doi:10.1212/WNL.0b013e3182764aa2

Tate, R. L., Lane-Brown, A. T., Myles, B. M., & Cameron, I. D. (2020). A longitudinal study of

support needs after severe traumatic brain injury. *Brain Injury, 34*(8), 991-1000. doi:10.1080/02699052.2020.1764101

Tavener, M., Thijsen, A., Hubbard, I. J., Francis, J. L., Grennall, C., Levi, C., & Byles, J. (2015). Acknowledging How Older Australian Women Experience Life After Stroke: How Does the WHO 18-Item Brief ICF Core Set for Stroke Compare? *Health Care for Women International, 36*(12), 1311-1326. doi:10.1080/07399332.2015.1055747

Terwee, C., Mokkink, L., Knol, D., Ostelo, R., Bouter, L., & de Vet, H. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research, 21*(4), 651-657. doi:10.1007/s11136-011-9960-1

Terwee, C. B., Bot, S. D. M., de Boer, M., Van der Windt, D. A. W. M., Knol, D. L., Dekker, J., . . . de Vet, H. C. W. (2017). Criteria for good measurement properties Retrieved from http://www.cosmin.nl/Systematic%20reviews%20of%20measurement%20properties.html

Terwee, C. B., Bot, S. D. M., de Boer, M. R., W., v. D., A.W.M., D., Knol, D. L., . . . de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34-42. doi:10.1016/j.jclinepi.2006.03.012

Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research, 18*(8), 1115. doi:10.1007/s11136-009-9528-5

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., . . . Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research, 27*(5), 1159-1170. doi:10.1007/s11136-018-1829-0

The European Stroke Organisation Executive Committee. (2008). Guidelines for Management of Ischaemic Stroke and Transient Ischaemic Attack 2008. *Cerebrovascular Diseases, 25*(5), 457-507. doi:10.1159/000131083

Thomas, C. M., Allison, R., & Latour, J. M. (2018). Using blogs to explore the lived-experience of life after stroke: “A journey of discovery I never wanted to take”. *Journal of Advanced Nursing, 74*(3), 579-590. doi:10.1111/jan.13457

Tornbom, K., Lundalv, J., & Sunnerhagen, K. S. (2019). Long-term participation 7-8 years after stroke: Experiences of people in working-age. *PLoS One, 14*(3), e0213447. doi:10.1371/journal.pone.0213447

Torregosa, M. B., Sada, R., & Perez, I. (2018). Dealing with stroke: Perspectives from stroke survivors and stroke caregivers from an underserved Hispanic community. *Nursing & Health Sciences, 20*(3), 361-369. doi:10.1111/nhs.12414

Tse, T., Linden, T., Churilov, L., Davis, S., Donnan, G., & Carey, L. M. (2019). Longitudinal changes in activity participation in the first year post-stroke and association with depressive symptoms. *Disability and Rehabilitation, 41*(21), 2548-2555. doi:10.1080/09638288.2018.1471742

Turner-Stokes, L. (2009). Goal attainment scaling (GAS) in rehabilitation: a practical guide. *Clinical Rehabilitation, 23*(4), 362-370. doi:10.1177/0269215508101742

Turner-Stokes, L., Ashford, S., Jacinto, J., Maisonobe, P., Balcaitiene, J., & Fheodoroff, K. (2016). Impact of integrated upper limb spasticity management including botulinum toxin A on patient-centred goal attainment: rationale and protocol for an international prospective, longitudinal cohort study (ULIS-III). *BMJ Open, 6*(6). doi:10.1136/bmjopen-2016-011157

Turner-Stokes, L., Baguley, I. J., De Graaff, S., Katrak, P., Davies, L., McCrory, P., & Hughes, A. (2010). Goal attainment scaling in the evaluation of treatment of upper limb spasticity with botulinum toxin: a secondary analysis from a double-blind placebo-controlled randomized clinical trial. *Journal of Rehabilitation Medicine, 42*(1), 81-89. doi:10.2340/16501977-0474

Turner-Stokes, L., Fheodoroff, K., Jacinto, J., & Maisonobe, P. (2013). Results from the Upper Limb International Spasticity Study-II (ULISII):a large, international, prospective cohort study investigating practice and goal attainment following treatment with botulinum toxin A in real-life clinical management. *BMJ Open, 3*(6). doi:10.1136/bmjopen-2013-002771

Turner-Stokes, L., Fheodoroff, K., Jacinto, J., Maisonobe, P., & Zakine, B. (2013). Upper limb international spasticity study: rationale and protocol for a large, international, multicentre prospective cohort study investigating management and goal attainment following treatment with botulinum toxin A in real-life clinical practice. *BMJ Open, 3*(3). doi:10.1136/bmjopen-2012-002230

Turner-Stokes, L., Pick, A., Nair, A., Disler, P. B., & Wade, D. T. (2015). Multi-disciplinary rehabilitation for acquired brain injury in adults of working age. *The Cochrane Database of Systematic Reviews, 2015*(12), CD004170. doi:10.1002/14651858.CD004170.pub3

Turner-Stokes, L., Rose, H., Ashford, S., & Singer, B. (2015). Patient engagement and satisfaction with goal planning: Impact on outcome from rehabilitation. *International Journal of Therapy and Rehabilitation, 22*(5), 210-216. doi:10.12968/ijtr.2015.22.5.210

Turner-Stokes, L., & Turner-Stokes, T. (1997). The use of standardized outcome measures in rehabilitation centres in the UK. *Clinical Rehabilitation, 11*(4), 306-313. doi:10.1177/026921559701100407

Turner-Stokes, L., & Williams, H. (2010). Goal attainment scaling: a direct comparison of alternative rating methods. *Clinical Rehabilitation, 24*(1), 66-73. doi:10.1177/0269215509343846

Turner-Stokes, L., Williams, H., & Johnson, J. (2009). Goal attainment scaling: does it provide added value as a person-centred measure for evaluation of outcome in neurorehabilitation following acquired brain injury? *Journal of Rehabilitation Medicine, 41*(7), 528-535.

Urban, P. P., Wolf, T., Uebele, M., Marx, J. J., Vogt, T., Stoeter, P., . . . Wissel, J. (2010). Occurence and Clinical Predictors of Spasticity After Ischemic Stroke. *Stroke, 41*(9), 2016-2020. doi:10.1161/STROKEAHA.110.581991

Urbin, M. A., Bailey, R. R., & Lang, C. E. (2015). Validity of body-worn sensor acceleration metrics to index upper extremity function in hemiparetic stroke. *Journal of Neurological Physical Therapy, 39*(4), 111-118. doi:10.1097/NPT.0000000000000085

Uswatte, G., & Taub, E. (1999). Constraint-Induced movement therapy: new approaches to measurement in rehabilitation. In D. Stuss, G. Winocur, & I. Robertson (Eds.), *Cognitive Neurorehabilitation*. Cambridge Cambridge University Press.

Uswatte, G., & Taub, E. (2005). Implications of the learned nonuse formulation for measuring rehabilitation outcomes: Lessons from constraint-induced movement therapy. *Rehabilitation Psychology, 50*(1), 34-42. doi:10.1037/0090-5550.50.1.34

Uswatte, G., Taub, E., Morris, D., Light, K., & Thompson, P. A. (2006). The motor activity log-28: Assessing daily use of the hemiparetic arm after stroke. *Neurology, 67*(7), 1189-1194. doi:10.1212/01.wnl.0000238164.90657.c2

Valderas, J. M., Ferrer, M., Mendívil, J., Garin, O., Rajmil, L., Herdman, M., & Alonso, J. (2008). Development of EMPRO: A Tool for the Standardized Assessment of Patient-Reported Outcome Measures. *Value Health, 11*(4), 700-708. doi:10.1111/j.1524-4733.2007.00309.x

Van der Lee, H. J., Wagenaar, C. R., Lankhorst, J. G., Vogelaar, W. T., Devillé, L. W., & Bouter, M. L. (1999). Forced Use of the Upper Extremity in Chronic Stroke Patients: Results From a Single-Blind Randomized Clinical Trial. *Stroke, 30*(11), 2369-2375. doi:10.1161/01.STR.30.11.2369

Van der Lee, J. H., De Groot, V., Beckerman, H., Wagenaar, R. C., Lankhorst, G. J., & Bouter, L. M. (2001). The intra- and interrater reliability of the Action Research Arm Test: A practical test of upper extremity function in patients with stroke. *Archives of Physical Medicine and Rehabilitation, 82*(1), 14-19. doi:10.1053/apmr.2001.18668

Van der Lee, J. H., Roorda, L. D., Beckerman, H., Lankhorst, G. J., & Bouter, L. M. (2002). Improving the Action Research Arm test: a unidimensional hierarchical scale. *Clinical Rehabilitation, 16*, 646-653. doi:10.1191/0269215502cr534oa

van der Putten, J. J. M. F., Hobart, J. C., Freeman, J. A., & Thompson, A. J. (1999). Measuring change in disability after inpatient rehabilitation: comparison of thew responsiveness of the Barthel Index and the Functional Independence Measure. *Journal of Neurology, Neurosurgery & Psychiatry, 66*(4), 480-484. doi:10.1136/jnnp.66.4.480

Van Herk, P. M. R., Van Dijk, P. M. M., Baar, P. M. F., Tibboel, P. M. D., & De Wit, P. M. R. (2007). Observation Scales for Pain Assessment in Older Adults With Cognitive Impairments or Communication Difficulties. *Nursing Research, 56*(1), 34-43. doi:10.1097/00006199-200701000-00005

Van Peppen, R. P. S., Schuurmans, M. J., Stutterheim, E. C., Lindeman, E., & Van Meeteren, N. L. U. (2009). Promoting the use of outcome measures by an educational programme for physiotherapists in stroke rehabilitation: a pilot randomized controlled trial. *Clinical Rehabilitation, 23*(11), 1005-1017. doi:10.1177/0269215509338999

van Straten, A., de Haan, R. J., Limburg, M., Shuling, J., Bossuyt, P. M., & van den Bos, G. A. M. (1997). A Stroke-adapted 30-item version of the sickness impact profile to assess quality of life (SA-SIP30). *Stroke, 28*(11), 2155-2161. doi:10.1161/01.STR.28.11.2155

Verbeek, J. M., van Wegen, E., van Peppen, R., van der Wees, P. J., Heniks, E., Rietberg, M., & Kwakkel, G. (2014). What Is the Evidence for Physical Therapy Poststroke? A Systematic Review and Meta-Analysis. *PLoS One, 9*(2). doi:10.1371/journal.pone.0087987

Verma, J. P. (2015). *Repeated Measures Design for Empirical Researchers*. Hoboken, New Jersey: John Wiley & Sons, Incorporated.

Vickrey, B. G., Hays, R. D., Genovese, B. J., Myers, L. W., & Ellison, G. W. (1997). Comparison of a generic to disease-targeted health-related quality-of-life measures for multiple sclerosis. *Journal of Clinical Epidemiology, 50*(5), 557-569. doi:10.1016/S0895-4356(97)00001-2

Vickrey, B. G., Hays, R. D., Harooni, R., Myers, L. W., & Ellison, G. W. (1995). A health-related quality of life measure for multiple sclerosis. *Quality of Life Research, 4*(3), 187-206. doi:10.1007/BF02260859

Waddell, J. K., Tabak, G. R., Strube, J. M., Haire-Joshu, E. D., & Lang, E. C. (2019). Belief, Confidence, and Motivation to Use the Paretic Upper Limb in Daily Life Over the First 24 Weeks After Stroke. *Journal of Neurologic Physical Therapy, 43*(4), 197-203. doi:10.1097/NPT.0000000000000287

Wade, D. T. (2020). What is rehabilitation? An empirical investigation leading to an evidence-based description. *Clinical Rehabilitation, 34*(5), 571-583. doi:10.1177/0269215520905112

Wade, D. T., & Hewer, R. L. (1987). Functional abilities after stroke: measurement, natural history and prognosis. *Journal of Neurology, Neurosurgery & Psychiatry, 50*(2), 177-182. doi:10.1136/jnnp.50.2.177

Wagenaar, R. C., Meijer, O. G., van Wieringen, P. C., Kuik, D. J., Hazenberg, G. J., Lindeboom, J., . . . Rijswik, H. (1990). The functional recovery of stroke: a comparison between neuro-developmental teatment and the Brunnstrom method. *Scandinavian Journal of Rehabilitation Medicine, 22*(1), 1-8.

Walder, K., & Molineux, M. (2017). Re-establishing an occupational identity after stroke – a theoretical model based on survivor experience. *The British Journal of Occupational Therapy, 80*(10), 620-630. doi:10.1177/0308022617722711

Wales, K., Clemson, L., Lannin, N. A., & Cameron, I. (2016). Functional assessments used by occupational therapists with older adults at risk of activity and participation limitations: A systematic review. *PLoS One, 11*(2). doi:10.1371/journal.pone.0147980

Walker, C. M., Sunderland, A., Sharma, J., & Walker, M. F. (2004). The impact of cognitive impairment on upper body dressing difficulties after stroke: a video analysis of patterns of recovery. *Journal of Neurology, Neurosurgery and Psychiatry, 75*(1), 43-48.

Wallace, D., Duncan, P. W., & Min Lai, S. (2002). Comparison of the responsiveness of the Barthel

Index and the Motor component of the Functional Independence Measure in stroke: The impact of using different methods for measuring responsiveness. *Journal of Clinical Epidemiology, 55*(9), 922-928. doi:10.1016/S0895-4356(02)00410-9

Wang, P., Li, H., Guo, Y., Xie, Y., Ge, R., & Qiu, Z. (2014). The feasibility and validity of the comprehensive ICF core set for stroke in Chinese clinical settings. *Clinical Rehabilitation, 28*(2), 159-171. doi:10.1177/0269215513496659

Ward, A. B., Aguilar, M., De Beyl, Z., Gedin, S., Kanovsky, P., Molteni, F., . . . Yakovleff, A. (2004). Use of botulinum toxin type A in management of adult spasticity. A European consensus statement. *Journal of Rehabilitation Medicine, 35*(2), 98-99. doi:10.1080/16501970306112

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-Item Short-Form Health Survey (SF-36) I. Conceptual framework and item selection. *Medical Care, 30*(6), 473-483.

Wellwood, I., Dennis, M. S., & Warlow, C. P. (1995). A comparison of the Barthel Index and the OPCS Disability Instrument used to measure outcome after acute stroke. *Age and Ageing, 24*(1), 54-57. doi:10.1093/ageing/24.1.54

Wheeler, S., & Acord-Vira, A. (2016). *Occupational therapy practice guidelines for adults with traumatic brain injury.* (D. Lieberman & M. Arbesman Eds.). Bethesda: AOTA Press.

White, J. H., Mackenzie, L., Magin, P., & Pollack, M. R. P. (2008). The Occupational Experience of Stroke Survivors in a Community Setting. *Occupational Therapy Journal of Research: Occupation, Participation and Health, 28*(4), 160-167. doi:10.3928/15394492-20080901-05

Wilkinson, P. R., Wolfe, C. D. A., Warburton, F. G., Rudd, A. G., Howard, R. S., Ross-Russell, R. W., & Beech, R. (1997). Longer term quality of life and outcome in stroke patients: is the Barthel index alone an adequate measure of outcome. *Quality of Life Outcomes, 6*(3), 125-130. doi:10.1136/qshc.6.3.125

Williams, A. (1990). EuroQol - a new facility for the measurement of health-related quality of life. *Health policy, 16*(3), 199-208. doi:10.1016/0168-8510(90)90421-9

Williams, G., Olver, J., Graaff, S., & Singer, B. J. (2012). The use of botulinum toxin type A in the management of adult-onset focal spasticity: A survey of Australian allied health professionals. *Australian Occupational Therapy Journal, 59*(4), 257-264. doi:10.1111/j.1440-1630.2012.01027.x

Williams, G., Singer, B. J., Ashford, S., Hoare, B., Hastings-Ison, T., Fheodoroff, K., . . . Hill, B. (2020). A synthesis and appraisal of clinical practice guidelines, consensus statements and Cochrane systematic reviews for the management of focal spasticity in adults and children. *Disability and Rehabilitation*, 1-11. doi:10.1080/09638288.2020.1769207

Williams, L. S., Weinberger, M., Harris, L. E., & Biller, J. (1999). Measuring quality of life in a way that is meaningful to stroke patients. *Neurology, 53*(8), 1839-1843. doi:10.1212/WNL.53.8.1839

Winstein, J. C., Stein, R. J., Arena, C. R., Bates, J. B., Cherney, L. L., Cramer, E. S., . . . Zorowitz, D. R. (2016). Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke, 47*(6), e98-e169. doi:10.1161/STR.0000000000000098

Wolf, S. L., Lecraw, D. E., Barton, L. A., & Jann, B. B. (1989). Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head injured patients. *Experimental Neurology, 104*, 125-132.

Wolf, T., & Koster, J. (2013). Perceived recovery as a predictor of physical activity participation after mild stroke. *Disability and Rehabilitation, 35*(14), 1143-1148. doi:10.3109/09638288.2012.720635

World Health Organisation. (2001). *International Classification of Functioning, Disability and Health*. Geneva: World Health Organisation.

World Health Organisation. (2002). *Towards a common language for functioning, disability, and health: ICF*. Geneva: World Health Organisation.

World Health Organisation. (2013). *How to use the ICF: A practical manual for using the*

International Classification of Functioning, Disability and Health (ICF). Exposure for comment. Geneva: WHO.

Xie, J., Wu, E. Q., Zheng, Z. J., Croft, J. B., Greenlund, K. J., Mensah, G. A., & Labarthe, D. R. (2006). Impact of stroke on health-related quality of life in the noninstitutionalized population in the United States. *Stroke, 37*(10), 2567-2572. doi:10.1161/01.STR.0000240506.34616.10

Yozbatıran, N., Baskurt, F., Baskurt, Z., Ozakbas, S., & Idiman, E. (2006). Motor assessment of upper extremity function and its relation with fatigue, cognitive function and quality of life in multiple sclerosis patients. *Journal of the Neurological Sciences, 246*(1-2), 117-122. doi:10.1016/j.jns.2006.02.018

Yozbatiran, N., Der-Yeghiaian, L., & Cramer, S. C. (2008). A standardized approach to performing the action research arm test. *Neurorehabilitation and Neural Repair, 22*(1), 78-90. doi:10.1177/1545968307305353

Ytterberg, C., Dyback, M., Bergstrom, A., Guidetti, S., & Eriksson, G. (2017). Perceived impact of stroke six years after onset, and changes in impact between one and six years. *Journal of Rehabilitation Medicine, 49*(8), 637-643. doi:10.2340/16501977-2258

Zaki, R., Bulgiba, A., Nordin, N., & Ismail, N. (2013). A Systematic Review of Statistical Methods Used to Test for Reliability of Medical Instruments Measuring Continuous Variables. *Iranian Journal of Basic Medical Sciences, 16*(6), 803-807.

Zaza, C., Stolee, P., & Prkachin, K. (1999). The Application of Goal Attainment Scaling in Chronic Pain Settings. *Journal of Pain and Symptom Management, 17*(1), 55-64. doi:10.1016/S0885-3924(98)00106-7

Zhang, T., Liu, L., Xie, R., Peng, Y., Wang, H., Chen, Z., . . . Li, B. (2018). Value of using the international classification of functioning, disability, and health for stroke rehabilitation assessment: A multicenter clinical study. *Medicine, 97*(42), e12802-e12802. doi:10.1097/MD.0000000000012802

Appendices

Appendix A: Publication permissions

Items that are subject to copyright with the permission obtained and the person or organisation providing that permission (with dates) and open access items with links to the licenses (with dates the materials were obtained) are presented.

Items that are subject to copyright:

- Pike, S., Lannin, N. A., Wales, K., & Cusick, A. (2018). A systematic review of the psychometric properties of the Action Research Arm Test in neurorehabilitation. *Australian Occupational Therapy Journal, 65*(5), 449-471. doi:10.1111/1440-1630.12527

  License number: 1111068-1
  License date: May 12th 2021
  Permission obtained from: John Wiley and Sons

- Pike, S., Lannin, N. A., Cameron, L., Palit, M., & Cusick, A. (2021). Chronic stroke survivors with upper limb spasticity: linking experience to the ICF. *Disability and Rehabilitation,* doi: 10.1080/09638288.2021.1894490

Open access items:

- Pike, S., Lannin, N. A., Cusick, A., Wales, K., Turner-Stokes, L., & Ashford, S. (2015). A systematic review protocol to evaluate the psychometric properties of measures of function within adult neuro-rehabilitation. *Systematic reviews*, 4(1), 86. doi:10.1186/s13643-015-0076-5

  Link to the licence: http://creativecommons.org/licenses/by/4.0
  Date materials were obtained: April 5th 2021

- Pike, S., Cusick, A., Wales, K., Cameron, L., Turner-Stokes, L., Ashford, S., & Lannin, N. A. (2021) Psychometric properties of measures of upper limb activity performance in adults with and without spasticity undergoing

neurorehabilitation– A systematic review. PLoS ONE 16(2): e0246288. doi.org/10.1371/journal.pone.0246288

Date materials were obtained: April 5th 2021

Appendix B: Ethics

The ethics approvals for Study 1, 2, 4 and 5 are presented in the table below with the corresponding study number and thesis chapter to increase ease of reader understanding. The following section provides the ethics approvals and participant information and consent forms.

Appendix Table 1. *Study ethics approvals*

Relevant study	Ethics approval details
Study 1 & 2	HREC/16/Alfred/173: The Alfred Human Research Ethics Committee
Study 3	No ethics
Study 4 & 5	HE12/077: University of Wollongong and ISLHD Health and Medical Human Research Ethics Committee
	FHEC12/152: La Trobe University Faculty Human Ethics Committee
	38/13: The Alfred Ethics Committee
	602-13: Epworth Healthcare Human Research Ethics Committee

Ethics Committee

Certificate of Authorisation of Amendments

This is to certify that amendments to

Project: **HREC/16/Alfred/173 (Local reference:442/14)**
The InTENSE trial: optimising upper limb recovery following stroke

Principal Researcher: **A/Professor Natasha Lannin**

Amendment: **Change to research personnel:**

Addition of Shannon Pike

have been authorised under the Streamlined Ethical Review Process (SERP) in accordance with your amendment application dated **11-Sep-2018** on the understanding that you observe the National Statement on Ethical Conduct in Human Research.

It is now your responsibility to ensure that all people associated with this particular research project are made aware of what has actually been approved and any caveats specified in correspondence with the Ethics Committee. Any further change to the application which is likely to have a significant impact on the ethical considerations of this project will require approval from the Ethics Committee.

[redacted]

Professor John J. McNeil
2018 Chair, Ethics Committee

Date: **08-Oct-**

All research subject to Alfred Hospital Ethics Committee review must be conducted in accordance with the National Statement on Ethical Conduct in Human Research (2007).

The Alfred Ethics Committee is a properly constituted Human Research Ethics Committee operating in accordance with the National Statement on Ethical Conduct in Human Research (2007)

APPROVAL
In reply please quote HE12/077
Further Enquiries Ph: 4221 3386
SF:MOT

11 April 2012

Mrs Shannon Pike
Wagga Wagga Ambulatory Rehabilitation Service
Wagga Wagga Base Hospital
PO Box 159
Wagga Wagga NSW 2650

Dear Mrs Pike,

Thank you for your response to the HREC letter regarding the ethics application below. I am pleased to advise that the application has been **approved.**

Ethics Number:	HE12/077
Au RED Number:	HREC/12/WGONG/22
Project Title:	Comparing two versions of the goal attainments assessment in adults with limb spasticity undergoing rehabilitation.
Researchers:	Mrs Shannon Pike, A/Professor Natasha Lannin, Professor Carolyn Unsworth, Professor Ann Cusick, Dr Louis Baggio

Sites/CIs approved:

Site	Principal Investigator for site
Wagga Wagga Base Hospital	A/Professor Natasha Lannin
Caulfield Hospital/ Alfred Health, Vic	A/Professor Natasha Lannin
Epworth Hospital, Vic	A/Professor Natasha Lannin

Documents Reviewed/Approved:

Initial Application

Participant Information Sheet v.1 10.2.2012

Patient Consent Form v.1 10.2.2012

Participant Information Sheet for Person Responsible/Carer v.1 10.2.2012

Consent Form for Person Responsible/Carer v.1 10.2.2012

Goal Attainment Scale (GAS) – Light v.1 10.2.2012

Additional Information letter dated 30 March 2012

Demographic and Clinical Information Form v.2 30.3.2012: Demographics; Clinical Information; NIH Stroke Scale; Oxfordshire Community Stroke project Classification; Mini-Mental State Exam.

Approval Date:	5 April 2012
Expiry Date:	4 April 2013

The University of Wollongong/ISLHD Health and Medical HREC is constituted and functions in accordance with the NHMRC *National Statement on Ethical Conduct in Human Research*. The HREC has reviewed the research proposal for compliance with the *National Statement* and approval of this project is conditional upon your continuing compliance with this document.

A condition of approval by the HREC is the submission of a progress report annually and a final report on completion of your project. The progress report template is available at http://www.uow.edu.au/research/rso/ethics/UOW009385.html. This report must be completed, signed by the appropriate Head of School and returned to the Research Services Office prior to the expiry date.

As evidence of continuing compliance, the Human Research Ethics Committee also requires that researchers immediately report:

- proposed changes to the protocol including changes to investigators involved
- serious or unexpected adverse effects on participants
- unforseen events that might affect continued ethical acceptability of the project.

Please note that approvals are granted for a twelve month period. Further extension will be considered on receipt of a progress report prior to expiry date.

Please note that Governance approval is required for research within NSW Ministry of Health. Before you can proceed with this research project you must complete a Site Specific Assessment (SSA) for each Local Health District included in your project. Refer to: https://ethicsform.org/Au/SignIn.aspx

For further information regarding the SSA in the ISLHD, contact:

Research Governance Officer
Illawarra Shoalhaven Local Health District
Research Directorate
Wollongong Hospital
Block C, Level 8
P: 02 4253 4876
E: Kristy.Pierce@SESIAHS.HEALTH.NSW.GOV.AU

A copy of this letter has been forwarded to the ISLHD Research Governance Officer and to the Greater Southern Area Health Service.

If you have any queries regarding the HREC review process, please contact the Ethics Unit on phone 4221 3386 or email rso-ethics@uow.edu.au.

Yours sincerely,

[redacted]

Associate Professor Sarah Ferber
Chair, UOW & ISLHD Health and Medical
Human Research Ethics Committee

cc: Governance Officer, Research Directorate, ISLHD
cc: Research Support Officer, Greater Southern Area Health Service, Research Governance Office, Bega Hospital, PO Box 173, Bega NSW 2550.

Research Services Office University of Wollongong NSW 2522 Australia
Telephone: +61 2 4221 3386 Facsimile: +61 2 4221 4338
research-services@uow.edu.au www.uow.edu.au/research

University of Wollongong

NSW GOVERNMENT Health
Murrumbidgee
Local Health District

PARTICIPATION INFORMATION SHEET FOR PATIENTS

RESEARCH TITLE: Comparing two versions of the goal attainment assessments in adults with limb spasticity undergoing rehabilitation.

PURPOSE OF THE RESEARCH

This is an invitation to participate in a study conducted by Associate Professor Natasha Lannin and Mrs Shannon Pike, an occupational therapist. The purpose of this research is to investigate the use of the Goal Attainment Scale-Light (GAS-Light) tool, to determine whether it is appropriate for use in limb spasticity management. The GAS-Light tool is used to identify goals for treatment of limb spasticity and to assess progress made towards achieving those goals. This study is important as the GAS-Light is proposed to be a fast and simple tool for clinicians and patients to use. The use of a bank of common goals will be available when completing the GAS-Light tool.

INVESTIGATORS

Associate Professor Natasha Lannin
Occupational Therapist
Alfred Clinical School
Ph: 03 94796755
N.Lannin@latrobe.edu.au

Shannon Pike
Occupational Therapist
Murrumbidgee Local Health District
Ph: 02 69386219
shannonleepike@hotmail.com

ASSOCIATE RESEARCHERS

Professor Carolyn Unsworth
Faculty of Health Sciences
La Trobe University
Ph: 03 9479 5700
c.unsworth@latrobe.edu.au

Professor Anne Cusick
School of Health Sciences
University of Wollongong
Ph: 02 42214161
acusick@uow.edu.au

Dr Louis Baggio
Rehabilitation Consultant
Murrumbidgee Local Health District
Ph: 02 69386666
Louis.baggio@gsahs.health.nsw.gov.au

METHOD AND DEMANDS ON PARTICIPANTS

If you choose to be included, as per routine spasticity management procedures you will be asked to identify goals with a clinician for the treatment of your limb spasticity during your first attendance to the spasticity management clinic. There will be a second clinician present for the goal setting and the two clinicians will fit your goals within the GAS-Light and standard GAS tools concurrently during the session. You will also be asked to complete a form to collect demographic details and clinical information about your medical condition. In addition, during your 6-8 week review your progress towards achieving your goals will be assessed by two clinicians to complete the GAS-Light and standard GAS tools. Filling in the forms will be in addition to any other tools and assessments that are usually collected as part of the routine spasticity management clinic procedures.

The duration of participation will be approximately an additional15 minutes at the initial visit, no additional time will be required at the follow-up visit as this process is part of routine procedures.

POSSIBLE RISKS, INCONVENIENCES AND DISCOMFORTS

Apart from the additional 15 minutes of your time at the initial appointment and an additional clinician being present during goal setting and assessment of goal achievement you will be participating in routine spasticity management procedures, we can foresee no risks for you. Your involvement in the study is voluntary and you may withdraw your participation from the study at any time and withdraw any data that

you have provided to that point. Refusal to participate in the study will not affect your relationship with the Wagga Wagga Base Hospital or the Murrumbidgee Local Health District. If you have any questions regarding the study, please do not hesitate to contact either investigator to discuss.

FUNDING AND BENEFITS OF THE RESEARCH

This study is unfunded and you will not be paid for taking part.

This research will provide clinicians with evidence to make an informed decision regarding the GAS-Light tool and its inclusion in limb spasticity management. Clinicians and patients in the future will benefit from a faster streamlined tool to identify goals and assess progress towards goal achievement. Findings from the study will be used for publication in journals and presentations to fellow healthcare professionals at conferences.

CONFIDENTIALITY

Your privacy and confidentiality will be protected by the research team at all times. All of your responses will be kept completely confidential and you will be identified by a participant number only, and only this number will appear with your responses. All data will be stored in locked filing cabinets, in a locked room, accessible only to the research team. After the project has been completed, Alfred Clinical School of La Trobe University will be responsible for long term storage and eventual destruction of the paper-based data.

ETHICS REVIEW AND COMPLAINTS

This study has been reviewed by the Human Research Ethics Committee (Social Science, Humanities and Behavioural Science) of the University of Wollongong. If you have any concerns or complaints regarding the way this research has been conducted, you can contact the UoW Ethics Officer on (02) 4221 4457.

Thank you for your interest in this study.

University of Wollongong

Health
Murrumbidgee
Local Health District

CONSENT FORM FOR PATIENTS

Research Title: Comparing two versions of the goal attainment assessments in adults with limb spasticity undergoing rehabilitation.

Researchers: Associate Professor Natasha Lannin and Shannon Pike

Associate Researchers: Professor Carolyn Unsworth, Professor Anne Cusick and Doctor Louis Baggio

I have been given information about "Comparing two versions of the goal attainment assessments in adults with limb spasticity undergoing rehabilitation" and discussed the research project with Shannon Pike.

I have been advised that there are no known or potential physical or psychological risks that can result from participating in this project, and that the burden which may be associated with this research is inconvenience due to an increase in time that I am required to spend at the spasticity management clinic and the presence of an additional clinician during the goal setting and goal assessment sessions. I understand that if I feel uncomfortable with a question, I may skip that question or stop responding to the question and have had an opportunity to ask Shannon Pike any questions I may have about the research and my participation.

I understand that my participation in this research is voluntary, I am free to refuse to participate and I am free to withdraw from the research at any time. My refusal to participate or withdrawal of consent will not affect my treatment in any way or my relationship with the Wagga Wagga Base Hospital or Murrumbidgee Local Health District.

If I have any enquiries about the research, I can contact Shannon Pike 02 6938 6219 or Associate Professor Natasha Lannin 03 94796755 or if I have any concerns or complaints regarding the way the research is or has been conducted, I can contact the Ethics Officer, Human Research Ethics Committee, Office of Research, University of Wollongong on 4221 4457.

By signing below I am indicating my consent to participate in the research study investigating the Goal Attainment Scale-Light (GAS-Light) tool. This tool is used to identify goals for treatment of my limb spasticity and to assess progress made towards achieving those goals. I understand that I will be asked to identify goals with two clinicians for treatment of my limb spasticity during my first attendance to the spasticity management clinic. Those goals will be fitted within the GAS-Light and standard GAS tools concurrently during the session. I understand that I will also be asked to complete a form to collect demographic details and clinical information about my medical condition. In addition, during my 6-8 week review; my progress towards achieving my goals will be assessed by two clinicians to complete both the GAS-Light and standard GAS tools. I understand that this process will require an additional 15 minutes to be spent at the initial visit to the clinic and that an additional clinician will be present. Filling in the forms will be in addition to any other tools and assessments that are part of the routine spasticity management clinic procedures.

University of Wollongong

NSW GOVERNMENT Health Murrumbidgee Local Health District

I understand that my privacy and confidentiality will be protected by the research team at all times. All of my responses will be kept completely confidential and I will be identified by a participant number only, and only this number will appear with my responses. All data will be stored in locked filing cabinets, in a locked room, accessible only to the research team. After the project has been completed, Alfred Clinical School of La Trobe University will be responsible for long term storage and eventual destruction of the paper-based data.

I understand that the data collected from my participation will be used for the purpose of a journal publication and presentations to fellow healthcare professionals, and I consent for it to be used in that manner. I also understand that no publications or presentations will use my name and that they will present only aggregated (combined) data.

Name (please print).....

Signature..... Date...../...../.....

University of Wollongong

Health
NSW GOVERNMENT | Murrumbidgee Local Health District

PARTICIPATION INFORMATION SHEET FOR PERSON RESPONSIBLE/CARER

RESEARCH TITLE: Comparing two versions of the goal attainment assessments in adults with limb spasticity undergoing rehabilitation.

PURPOSE OF THE RESEARCH

As the 'person responsible' for the patient, you are invited to consider the patient's participation in a study conducted by Associate Professor Natasha Lannin and Mrs Shannon Pike, an occupational therapist. The purpose of this research is to investigate the Goal Attainment Scale –Light (GAS-Light) tool, to determine whether it is appropriate for use in limb spasticity management. The GAS-Light tool is used to identify goals for treatment of limb spasticity and to assess progress made towards achieving those goals. This study is important as the GAS-Light is proposed to be a fast and simple tool for clinicians and patients to use. The use of a bank of common goals will be available when completing the GAS-Light tool.

INVESTIGATORS

Associate Professor Natasha Lannin
Occupational Therapist
Alfred Clinical School
Ph: 03 94796755
N.Lannin@latrobe.edu.au

Shannon Pike
Occupational Therapist
Murrumbidgee Local Health District
Ph: 02 69386219
shannonleepike@hotmail.com

ASSOCIATE RESEARCHERS

Professor Carolyn Unsworth
Faculty of Health Sciences
La Trobe University
Ph: 03 9479 5700
c.unsworth@latrobe.edu.au

Professor Anne Cusick
School of Health Sciences
University of Wollongong
Ph: 02 42214161
acusick@uow.edu.au

Dr Louis Baggio
Rehabilitation Consultant
Murrumbidgee Local Health District
Ph: 02 69386666
Louis.baggio@gsahs.health.nsw.gov.au

METHOD AND DEMANDS ON PARTICIPANTS

If you choose for the patient to be included, as per routine spasticity management procedures the patient and yourself will be asked to identify goals with a clinician for the treatment of your limb spasticity during your first attendance to the spasticity management clinic. There will be a second clinician present for the goal setting and the clinicians will fit those goals within the GAS-Light and standard GAS tools concurrently during the session. The patient and yourself will also be asked to complete a form to collect demographic details and clinical information about your medical condition. In addition, during the patients 6-8 week review the patient's progress towards goal achievement will be assessed by two clinicians to complete both the GAS-Light and standard GAS tools. Filling in the forms will be in addition to any other tools and assessments that are usually collected as part of the routine spasticity management clinic procedures.

The duration of participation will be approximately an additional15 minutes at the initial visit, no additional time will be required at the follow-up visit as this process is part of routine procedures.

POSSIBLE RISKS, INCONVENIENCES AND DISCOMFORTS

Apart from the additional 15 minutes of your time at the initial appointment and an additional clinician being present during goal setting and assessment of goal achievement you will be participating in routine spasticity management procedures, we can foresee no risks for you. The patient's involvement in the study is voluntary and you may withdraw their participation from the study at any time and withdraw any

data that they have provided to that point. Refusal to participate in the study will not affect the patients or your relationship with the Wagga Wagga Base Hospital or the Murrumbidgee Local Health District. If you have any questions regarding the study, please do not hesitate to contact either investigator to discuss.

FUNDING AND BENEFITS OF THE RESEARCH

This study is unfunded and you will not be paid for taking part.

This research will provide clinicians with evidence to make an informed decision regarding the GAS-Light tool and its inclusion in limb spasticity management. Clinicians and patients in the future will benefit from a faster streamlined tool to identify goals and assess progress towards goal achievement. Findings from the study will be used for journal publications and presented to fellow healthcare professionals at conferences.

CONFIDENTIALITY

Your privacy and confidentiality will be protected by the research team at all times. All of your responses will be kept completely confidential and you will be identified by a participant number only, and only this number will appear with your responses. All data will be stored in locked filing cabinets, in a locked room, accessible only to the research team. After the project has been completed, Alfred Clinical School of La Trobe University will be responsible for long term storage and eventual destruction of the paper-based data.

ETHICS REVIEW AND COMPLAINTS

This study has been reviewed by the Human Research Ethics Committee (Social Science, Humanities and Behavioural Science) of the University of Wollongong. If you have any concerns or complaints regarding the way this research has been conducted, you can contact the UoW Ethics Officer on (02) 4221 4457.

Thank you for your interest in this study.

University of Wollongong

NSW GOVERNMENT Health Murrumbidgee Local Health District

CONSENT FORM FOR PERSON RESPONSIBLE/CARER

Research Title: Comparing two versions of the goal attainment assessments in adults with limb spasticity undergoing rehabilitation.

Researchers: Associate Professor Natasha Lannin and Shannon Pike

Associate Researchers: Professor Carolyn Unsworth, Professor Anne Cusick and Doctor Louis Baggio.

I have been given information about "Comparing two versions of the goal attainment assessments in adults with limb spasticity undergoing rehabilitation" and discussed the research project with Shannon Pike.

I have been advised that there are no known or potential physical or psychological risks that can result from participating in this project, and that the burden which may be associated with this research is inconvenience due to an increase in time that the patient and I are required to spend at the spasticity management clinic and the presence of an additional clinician during the goal setting and goal assessment sessions. I understand that if the patient or I feel uncomfortable with a question, we may skip that question or stop responding to the question and have had an opportunity to ask Shannon Pike any questions I may have about the research and the patient's participation.

I understand that the patient's participation in this research is voluntary, I am free to refuse for the patient to participate and I am free to withdraw the patient from the research at any time. My refusal for the patient to participate or withdrawal of consent will not affect the patients or my treatment in any way or our relationship with the Wagga Wagga Base Hospital or Murrumbidgee Local Health District.

If I have any enquiries about the research, I can contact Shannon Pike 02 6938 6219 or Associate Professor Natasha Lannin 03 94796755 or if I have any concerns or complaints regarding the way the research is or has been conducted, I can contact the Ethics Officer, Human Research Ethics Committee, Office of Research, University of Wollongong on 4221 4457.

By signing below I am indicating my consent for the patient to participate in the research study investigating the Goal Attainment Scale –Light (GAS-Light) tool. This tool is used to identify goals for treatment of limb spasticity and to assess progress made towards achieving those goals. I understand that the patient and I will be asked to identify goals with two clinicians for treatment of the patient's limb spasticity during the patient's first attendance to the spasticity management clinic. Those goals will be fitted within the GAS-Light and standard GAS tools concurrently during the session. I understand that I will also be asked to complete a form to collect demographic details and clinical information about the patient's medical condition. In addition, during the patients 6-8 week review the patient's progress towards goal achievement will be assessed by two clinicians to complete both the GAS-Light and standard GAS tools. I understand that this process will require an additional 15 minutes to be spent at the initial visit to the clinic and that an additional clinician will be present. Filling in the forms will be in addition to any other tools and assessments that are part of the routine spasticity management clinic procedures.

University of Wollongong

NSW GOVERNMENT Health Murrumbidgee Local Health District

I understand that the patient's privacy and confidentiality will be protected by the research team at all times. All of the responses will be kept completely confidential and the patient will be identified by a participant number only, and only this number will appear with their responses. All data will be stored in locked filing cabinets, in a locked room, accessible only to the research team. After the project has been completed, Alfred Clinical School of La Trobe University will be responsible for long term storage and eventual destruction of the paper-based data.

I understand that the data collected from the patients participation will be used for the purpose of a journal publication and presentations to fellow healthcare professionals, and I consent for it to be used in that manner. I also understand that no publications or presentations will use my name and that they will present only aggregated (combined) data.

Participant's name (printed) .....

Name of Person Responsible (printed) .....

Relationship to participant: .....

Signature:..... Date:.....

Witness Name (printed).....

Witness Signature..... Date:.....

La Trobe University
Faculty of Health Sciences
MEMORANDUM

TO: Associate Professor Natasha Lannin School of Occupational Therapy
Professor Carolyn Unsworth
Doctor Louis Baggio

SUBJECT: *Reference:* **FHEC12/152**

Student or Other Investigator: Shannon Pike

Title: **Comparing two versions of the goal attainment assessments in adults with limb spasticity undergoing rehabilitation**

DATE: 15 August, 2012

The Faculty Human Ethics Committee's (FHEC) reviewers have considered and approved the above project that has been approved by the University of Wollongong Human Ethics Commitee.

Please note that normally La Trobe HEC's require Participant Information Statements and Consent Forms on Departmental letterhead. The Chair of FHS FHEC has waived that requirement in this case because of the prior approval by University of Wollongong Human Ethics Committee.

If you have a student/s involved in this project, a copy of this memorandum is enclosed for you to forward to the student(s) concerned.

[redacted]

Owen M Evans, PhD
Chair
Faculty Human Ethics Committee
Faculty of Health Sciences

ETHICS COMMITTEE CERTIFICATE OF APPROVAL

This is to certify that

Project No: 38/13

Project Title: Comparing two versions of goal attainment assessments in adults with limb spasticity undergoing rehabilitation.

Principal Researcher: A/Prof Natasha Lannin

Project Proposal: 38/13

Participant Information and Consent Form Version 3 **dated:** 11-Mar-2013
Participant Information and Consent Form (Person Responsible) Version 2 **dated:** 11-Mar-2013

*was considered by the Ethics Committee on **28-Feb-2013**, meets the requirements of the National Statement on Ethical Conduct in Human Research (2007) and was **APPROVED** on **14-Mar-2013***

It is the Principal Researcher's responsibility to ensure that all researchers associated with this project are aware of the conditions of approval and which documents have been approved.

The Principal Researcher is required to notify the Secretary of the Ethics Committee, via amendment or progress report, of

- Any significant change to the project and the reason for that change, including an indication of ethical implications (if any);
- Serious adverse effects on participants and the action taken to address those effects;
- Any other unforeseen events or unexpected developments that merit notification;
- The inability of the Principal Researcher to continue in that role, or any other change in research personnel involved in the project;
- Any expiry of the insurance coverage provided with respect to sponsored clinical trials and proof of re-insurance;
- A delay of more than 12 months in the commencement of the project; and,
- Termination or closure of the project.

Additionally, the Principal Researcher is required to submit

- A Progress Report on the anniversary of approval and on completion of the project *(forms to be provided)*; The Ethics Committee may conduct an audit at any time.

All research subject to the Alfred Hospital Ethics Committee review must be conducted in accordance with the National Statement on Ethical Conduct in Human Research (2007).

The Alfred Hospital Ethics Committee is a properly constituted Human Research Ethics Committee in accordance with the National Statement on Ethical Conduct in Human Research (2007).

SPECIAL CONDITIONS

None

SIGNED:

R Frew
Secretary, Ethics Committee

Please quote project number and title in all correspondence

Participant Information Statement/Consent Form

RESEARCH TITLE: COMPARING TWO VERSIONS OF GOAL ATTAINMENT ASSESSMENTS IN ADULTS WITH LIMB SPASTICITY UNDERGOING REHABILITATION.

LOCATION
Caulfield Hospital, Alfred Health Melbourne

INVESTIGATORS

Associate Professor Natasha Lannin
Occupational Therapist
Alfred Clinical School
Ph: 03 94796755

Shannon Pike
Occupational Therapist
Murrumbidgee Local Health District
Ph: 02 69386219

ASSOCIATE RESEARCHERS

Professor Carolyn Unsworth
Faculty of Health Sciences
La Trobe University
Ph: 03 9479 5700

Professor Anne Cusick
School of Health Sciences
University of Wollongong
Ph: 02 42214161

Dr Mithu Palit
Head of Neurological Rehabilitation
Caulfield Hospital
Ph: 03 9076 6416

INTRODUCTION
You are invited to take part in this research project. This is because you have spasticity in your upper limb. This project is comparing two versions of goal attainment assessments.

This Participant Information Statement/Consent Form tells you about the research project. It explains the procedures involved. Knowing what is involved will help you decide if you want to take part in the research.

Please read this information carefully. Ask questions about anything that you don't understand or want to know more about. Before deciding whether or not to take part, you may want to talk about it with a relative, friend or your local doctor.

Participation in this research is voluntary. If you don't wish to take part, you don't have to. You will receive the best possible care whether or not you take part.

If you decide to take part in the research project, you will be asked to sign the consent form. By signing it you are telling us that you:

- Understand what you have read
- Consent to take part in the research project
- Consent to participate in the procedures described
- Consent to the use of your personal and health information as described.

You will be given a copy of this Participant Information Statement and the Consent Form to keep.

PURPOSE OF THE RESEARCH
This is an invitation to participate in a study conducted by Associate Professor Natasha Lannin and Mrs Shannon Pike. Shannon Pike is conducting this research as part of a Master of Applied Science Degree supervised by Associate Professor Natasha Lannin at La Trobe University, Melbourne.

The purpose of this research is to investigate the use of the Goal Attainment Scale-Light (GAS-Light) tool, to determine whether it is appropriate for use in upper limb spasticity management. The GAS-Light

tool is used to identify goals for treatment of limb spasticity and to assess progress made towards achieving those goals. This study is important as the GAS-Light is proposed to be a fast and simple tool for clinicians and patients to use.

You have been invited to participate because you are an adult with spasticity in your upper limb which is impacting on your ability to complete everyday activities.

WHAT DOES PARTICIPATION IN THIS RESEARCH INVOLVE?

If you choose to be included, as per routine spasticity management procedures you will be asked to identify goals with a clinician for the treatment of your upper limb spasticity during your first attendance to the spasticity management clinic. There will be a second clinician present for the goal setting and the two clinicians will fit your goals within the GAS-Light and standard GAS tools concurrently during the session. You will also be asked to complete a form to collect demographic details, clinical information about your medical condition and a brief survey regarding the GAS-Light and goal setting. In addition, during your 6-8 week review your progress towards achieving your goals will be assessed by two clinicians to complete the GAS-Light and standard GAS tools. Filling in the forms will be in addition to any other tools and assessments that are usually collected as part of the routine spasticity management clinic procedures.

The duration of participation will be approximately an additional 15 minutes at the initial visit, no additional time will be required at the follow-up visit as this process is part of routine procedures.

This research project has been designed to make sure the researchers interpret the results in a fair and appropriate way and avoids the researchers or participants jumping to conclusions.

POSSIBLE RISKS, INCONVENIENCES AND DISCOMFORTS

Apart from the additional 15 minutes of your time at the initial appointment and an additional clinician being present during goal setting and assessment of goal achievement you will be participating in routine spasticity management procedures, we can foresee no risks for you.

If you are pregnant or the level of spasticity in your upper limb does not affect its use in daily activities then you are not eligible to participate.

USE OF DATA COLLECTED

Your privacy and confidentiality will be protected by the research team at all times. All of your responses will be kept completely confidential and you will be identified by a participant number only, and only this number will appear with your responses. All data will be stored in locked filing cabinets, in a locked room, accessible only to the research team. After the project has been completed, Alfred Clinical School of La Trobe University will be responsible for long term storage and eventual destruction of the paper-based data.

Findings from the study will be used for publication in paper based and online journals and presentations to fellow healthcare professionals at conferences.

FUNDING AND BENEFITS OF THE RESEARCH

This study is unfunded. A small grant has been awarded to enable participants to be offered a $20.00 voucher to assist with travel and/or parking costs.

This research will provide clinicians with evidence to make an informed decision regarding the GAS-Light tool and its inclusion in upper limb spasticity management. Clinicians and patients in the future will benefit from a faster streamlined tool to identify goals and assess progress towards goal achievement. You will not benefit from participating in this study.

Consent Form – Adult providing own consent

Title	Comparing two versions of goal attainment assessments in adults with limb spasticity undergoing rehabilitation
Principal Investigator	Associate Professor Natasha Lannin
Associate Investigator(s)	Mrs Shannon Pike, Professor Carolyn Unsworth, Professor Anne Cusick, Dr Mithu Palit
Location	Caulfield Hospital

Declaration by Participant

I have read and understood the participant information statement and consent form or someone has read it to me in a language that I understand.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I understand the purposes, procedures and risks of the research described in the project.

I freely agree to participate in the project, realising that I may withdraw at any time. I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that neither my name nor any other identifying information is used.

I understand that I will be given a signed copy of this document to keep.

Name of Participant (please print):

Signature: Date:

Name of Witness* to Participants Signature (please print):

Signature: Date:

*Witness is not to be the investigator, a member of the study team or their delegate. In the event that an interpreter is used, the interpreter may not act as a witness to the consent process. Witness must be 18 years or older.

Declaration by Senior Researcher[†]

I have given a verbal explanation of the research project, its procedures and risks and I believe that the participant has understood that explanation.

Name of Senior Researcher (please print):

Signature: Date:

[†]A senior member of the research team must provide the explanation of, and information concerning the research project.

Note: All parties signing the consent section must date their own signature.

Participant Information Statement/Consent Form – Person Responsible

RESEARCH TITLE: COMPARING TWO VERSIONS OF GOAL ATTAINMENT ASSESSMENTS IN ADULTS WITH LIMB SPASTICITY UNDERGOING REHABILITATION.

LOCATION
Caulfield Hospital, Alfred Health Melbourne

INVESTIGATORS

Associate Professor Natasha Lannin
Occupational Therapist
Alfred Clinical School
Ph: 03 94796755

Shannon Pike
Occupational Therapist
Murrumbidgee Local Health District
Ph: 02 69386219

ASSOCIATE RESEARCHERS

Professor Carolyn Unsworth
Faculty of Health Sciences
La Trobe University
Ph: 03 9479 5700

Professor Anne Cusick
School of Health Sciences
University of Wollongong
Ph: 02 42214161

Dr Mithu Palit
Head of Neurological Rehabilitation
Caulfield Hospital
Ph: 03 9076 6416

INTRODUCTION
The participant is invited to take part in this research project. This is because the participant has spasticity in their upper limb. This project is comparing two versions of goal attainment assessments.

This Participant Information Statement/Consent Form tells you about the research project. It explains the procedures involved. Knowing what is involved will help you decide if you want the participant to take part in the research.

Please read this information carefully. Ask questions about anything that you don't understand or want to know more about. Before deciding whether or not the participant can take part, you may want to talk about it with a relative, friend or the participant's local doctor.

Participation in this research is voluntary. If you don't wish the participant to take part, the participant doesn't have to. The participant will receive the best possible care whether or not they take part.

If you decide you want the participant to take part in the research project, you will be asked to sign the consent section. By signing it you are telling us that you:

- Understand what you have read
- Consent to the participant taking part in the research project
- Consent to the participant taking part in the procedures described
- Consent to the use of the participant's personal and health information as described.

You will be given a copy of this Participant Information Statement and Consent Form to keep.

PURPOSE OF THE RESEARCH
This is an invitation to participate in a study conducted by Associate Professor Natasha Lannin and Mrs Shannon Pike. Shannon Pike is conducting this research as part of a Master of Applied Science Degree supervised by Associate Professor Natasha Lannin at La Trobe University, Melbourne.

The purpose of this research is to investigate the use of the Goal Attainment Scale-Light (GAS-Light)

tool, to determine whether it is appropriate for use in upper limb spasticity management. The GAS-Light tool is used to identify goals for treatment of limb spasticity and to assess progress made towards achieving those goals. This study is important as the GAS-Light is proposed to be a fast and simple tool for clinicians and patients to use.

The participant has been invited to participate because they are an adult with spasticity in their upper limb which is impacting on their ability to complete everyday activities.

WHAT DOES PARTICIPATION IN THIS RESEARCH INVOLVE?

If you choose for the participant to be included, as per routine spasticity management procedures they will be asked to identify goals with a clinician for the treatment of their upper limb spasticity during their first attendance to the spasticity management clinic. There will be a second clinician present for the goal setting and the two clinicians will fit the participant's goals within the GAS-Light and standard GAS tools concurrently during the session. The participant will also be asked to complete a form to collect demographic details, clinical information about their medical condition and a brief survey regarding the GAS-Light and goal setting. In addition, during their 6-8 week review their progress towards achieving their goals will be assessed by two clinicians to complete the GAS-Light and standard GAS tools. Filling in the forms will be in addition to any other tools and assessments that are usually collected as part of the routine spasticity management clinic procedures.

The duration of participation will be approximately an additional 15 minutes at the initial visit, no additional time will be required at the follow-up visit as this process is part of routine procedures.

This research project has been designed to make sure the researchers interpret the results in a fair and appropriate way and avoids the researchers or participants jumping to conclusions.

POSSIBLE RISKS, INCONVENIENCES AND DISCOMFORTS

Apart from the additional 15 minutes of the participant's time at the initial appointment and an additional clinician being present during goal setting and assessment of goal achievement the participant will be participating in routine spasticity management procedures, we can foresee no risks for the participant.

If the participant is pregnant or the level of spasticity in their upper limb does not affect its use in daily activities then they are not eligible to participate.

USE OF DATA COLLECTED

The participant's privacy and confidentiality will be protected by the research team at all times. All of the participant's responses will be kept completely confidential and they will be identified by a participant number only, and only this number will appear with their responses. All data will be stored in locked filing cabinets, in a locked room, accessible only to the research team. After the project has been completed, Alfred Clinical School of La Trobe University will be responsible for long term storage and eventual destruction of the paper-based data.

Findings from the study will be used for publication in paper based and online journals and presentations to fellow healthcare professionals at conferences.

FUNDING AND BENEFITS OF THE RESEARCH

This study is unfunded. A small grant has been awarded to enable participants to be offered a $20.00 voucher to assist with travel and/or parking costs.

This research will provide clinicians with evidence to make an informed decision regarding the GAS-Light tool and its inclusion in upper limb spasticity management. Clinicians and patients in

the future will benefit from a faster streamlined tool to identify goals and assess progress towards goal achievement. You will not benefit from participating in this study.

The participant's involvement in the study is voluntary and you may withdraw the participant from the study at any time and withdraw any data that the participant has provided to that point. Refusal to participate in the study will not affect your or the participant's relationship with the Caulfield Hospital or Alfred Health.

ETHICS REVIEW AND COMPLAINTS

All research in Australia involving humans is reviewed by an independent group of people called a Human Research Ethics Committee (HREC). The ethical aspects of this research project have been approved by the HREC of Alfred Health.

This project will be carried out according to the *National Statement on Ethical Conduct in Human Research (2007).* This statement has been developed to protect the interest of people who agree to participate in human research studies.

Any questions regarding this project may be directed to the Investigators, Associate Professor Lannin, of the School of Health Sciences on 03 94796755 or Shannon Pike on 02 69386219.

If you have any complaints or queries that the investigator has not been able to answer to your satisfaction, you may contact the Research Governance Officer at Alfred Health Human Ethics on Ph: 03 9076 3619.

This study has also been reviewed by the following two Human Ethics Committees whom you may also contact if you have any concerns or complaint regarding the way this research has been conducted;

> Faculty Human Ethics Committee, Faculty of Health Sciences, La Trobe University, Victoria, 3086, please phone the secretary on (03) 9479 3583 or email health@latrobe.edu.au. (Quote application reference number FHEC12/152).
>
> Human Research Ethics Committee (Social Science, Humanities and Behavioural Science) of the University of Wollongong (UoW). Please contact the UoW Ethics Officer on (02) 4221 4457.

Thank you for your interest in this study.

Consent Form – Person Responsible

Title	Comparing two versions of goal attainment assessments in adults with limb spasticity undergoing rehabilitation
Principal Investigator	Associate Professor Natasha Lannin
Associate Investigator(s)	Mrs Shannon Pike, Professor Carolyn Unsworth, Professor Anne Cusick, Dr Mithu Palit
Location	Caulfield Hospital

Declaration by Participant

I am the Person Responsible for (the Participant).

I have read and understood the Participant Information Statement or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research described in the project.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I believe that the participation of the participant in this study is not contrary to their best interests.

I freely agree to the participant participating in this research project as described and understand that I am free to withdraw the participant at any time during the research project without affecting their future health care.

I am aware of my responsibilities as the Person Responsible for the participant and I understand that I will be assisting the participant in meeting their responsibilities whilst they are participating in this study.

I understand that I will be given a signed copy of this document to keep on behalf of the participant.

Name of Participant (please print):

Name of Person Responsible (please print):

Relationship of Peron Responsible to Participant:

Signature of Person Responsible: Date:

Name of Witness* to Person Responsible's Signature (please print):

Signature: Date:

*Witness is not to be the investigator, a member of the study team or their delegate. In the event that an interpreter is used, the interpreter may not act as a witness to the consent process. Witness must be 18 years or older.

Declaration by Senior Researcher[†]

I have given a verbal explanation of the research project, its procedures and risks and I believe that the participant has understood that explanation.

Name of Senior Researcher (please print):
Signature: Date:

[†]A senior member of the research team must provide the explanation of, and information concerning the research project.
Note: All parties signing the consent section must date their own signature.

Appendix C: Registration of Studies

- Study 1 and 2 ANZCTR registration
- Study 3 PROSPERO registration

Study 1 and 2 ANZCTR Registration

Trial registered on ANZCTR

Registration number	ACTRN12615000616572
Ethics application status	Approved
Date submitted	15/05/2015
Date registered	12/06/2015
Date last updated	9/07/2018
Type of registration	Prospectively registered

Titles & IDs

Public title	Intensive rehabilitation after botulinum toxin-A injections in stroke.
Scientific title	Effect of adding intensive upper limb rehabilitation to botulinum toxin-A on upper limb activity after stroke: the InTENSE trial.
Secondary ID [1]	Nil
Universal Trial Number (UTN)	
Trial acronym	InTENSE
Linked study record	

Health condition

Health condition(s) or problem(s) studied:

Stroke

Condition category	Condition code
Physical Medicine / Rehabilitation	Occupational therapy
Physical Medicine / Rehabilitation	Physiotherapy
Stroke	Haemorrhagic

Intervention/exposure

Study type	Interventional
Description of intervention(s) / exposure	An evidence-based therapy program is provided for 12 weeks immediately following botulinum toxin-A injection. The program commences with: (1) Up to 3 serial casts applied to place the wrist in maximum extension for 2 weeks. Cast is applied the first working day following botulinum toxin-A injection, followed by (2) 10 weeks of movement training, aimed at decreasing weakness (electrical stimulation and progressive resistance exercises) and improving movement (task-specific practice). Participants are encouraged to practice for 1 hour per day, 7 days a week during the 10 weeks (ie, approximately 70 hours in total inclusive of clinic based sessions). Participants are supported by a mix of clinic-based sessions, home visits, and phone calls. Clinic-based sessions will be conducted by a trained physical or occupational therapist (duration 1 hour, frequency decreasing (3 x per week initially, but decreasing to 1 x week by week 7)). Participants will receive a booklet outlining their movement training and a training log to record the number of minutes of practice daily.
Intervention code [1]	Rehabilitation
Comparator / control treatment	Written instructions plus follow-up telephone call post-clinic, as per usual care in Australia. Participants will receive a booklet containing 7 stretches, and 8 arm and hand exercises and a training log to record the number of minutes of practice daily.
Control group	Active

Outcomes

Primary outcome [1]	Upper limb activity will be measured using the Box and Block Test.
Timepoint [1]	Baseline (pre-randomisation), 12 weeks and 12 months
Primary outcome [2]	Individualised achievement of goals will be measured using Goal Attainment Scale
Timepoint [2]	12 weeks and 12 months
Secondary outcome [1]	Spasticity will be measured using the Tardieu scale
Timepoint [1]	Baseline (pre-randomisation) and 12 weeks
Secondary outcome [2]	Passive wrist extension will be measured using torque-controlled goniometry.
Timepoint [2]	Baseline (pre-randomisation),12 weeks and 12 months
Secondary outcome [3]	Health related quality of life will be measured using the EQ-5D
Timepoint [3]	Baseline (pre-randomisation), 12 weeks and 12 months
Secondary outcome [4]	Grip strength will be measured as a maximum voluntary contraction using dynamometry
Timepoint [4]	Baseline (pre-randomisation), 12 weeks and 12 months
Secondary outcome [5]	Pain will be measured using a 10-cm visual analogue scale
Timepoint [5]	Baseline (pre-randomisation), 12 weeks and 12 months
Secondary outcome [6]	Burden of care will be measured using the Carer Burden Scale
Timepoint [6]	Baseline (pre-randomisation), 12 weeks and 12 months

Eligibility

Key inclusion criteria	- Scheduled to receive a botulinum toxin-A injection to a muscle(s) that crosses the wrist - Agreed to receive BoNT-A injections as part of their usual care - Date of stroke three or more months prior. - Not currently receiving upper limb rehabilitation - Absence of significant cognitive impairment (as assessed by a score of less than five adjusted errors on the Short Portable Mental Status Questionnaire).
Minimum age	18 Years
Maximum age	No limit
Gender	Both males and females
Can healthy volunteers participate?	No
Key exclusion criteria	- Unable to attend clinic at least 1/wk. - Other significant upper limb impairment eg. Fracture or frozen shoulder within 6 months, severe arthritis, amputation. - Presence of any and all contraindications to botulinum toxin-A injections. - Botulinum toxin-A injections and/or serial casts in the past 6 months.

Study design

Purpose of the study	Treatment
Allocation to intervention	Randomised controlled trial
Procedure for enrolling a subject and allocating the treatment (allocation concealment procedures)	For each clinic, allocation will occur in random permuted blocks so that after every block (of 4-8 participants), the experimental and control group will contain equal numbers. Randomization will occur after injection of botulinum toxin-A. The schedule will be stored off-site and group allocation will be revealed online.
Methods used to generate the sequence in which subjects will be randomised (sequence generation)	Randomization will be computer-generated, independent and concealed. The allocation sequence has been generated and will be managed by the Griffith Randomistation Service at Griffith University.
Masking / blinding	Blinded (masking used)
Who is / are masked / blinded?	The people assessing the outcomes
Intervention assignment	Parallel
Other design features	
Phase	Not Applicable
Type of endpoint(s)	Efficacy
Statistical methods / analysis	

Recruitment

Recruitment status	Recruiting

Date of first participant enrolment			
Anticipated	3/07/2015	**Actual**	3/07/2015

Date of last participant enrolment			
Anticipated		**Actual**	

Date of last data collection			
Anticipated		**Actual**	

Sample size					
Target	136	**Accrual to date**	134	**Final**	

Recruitment in Australia	
Recruitment state(s)	NSW,SA,VIC
Recruitment hospital [1]	The Alfred - Prahran
Recruitment hospital [2]	Austin Health - Austin Hospital - Heidelberg
Recruitment hospital [3]	Royal Talbot Rehabilitation Centre - Kew
Recruitment hospital [4]	Epworth Rehabilitation Camberwell - Camberwell
Recruitment hospital [5]	Hampstead Rehabilitation Centre - Northfield
Recruitment hospital [6]	Repatriation General Hospital - Daw Park
Recruitment hospital [7]	St Joseph's Hospital - Auburn
Recruitment hospital [8]	St Vincent's Private Hospital (Darlinghurst) - Darlinghurst
Recruitment postcode(s) [1]	2010 - Darlinghurst
Recruitment postcode(s) [2]	2144 - Auburn
Recruitment postcode(s) [3]	3004 - Prahran
Recruitment postcode(s) [4]	3084 - Heidelberg
Recruitment postcode(s) [5]	3101 - Kew
Recruitment postcode(s) [6]	3124 - Camberwell
Recruitment postcode(s) [7]	5041 - Daw Park
Recruitment postcode(s) [8]	5085 - Northfield

Funding & Sponsors

Funding source category [1]	Government body
Name [1]	National Health and Medical Research Council
Address [1]	GPO Box 1421 Canberra ACT 2601
Country [1]	Australia
Primary sponsor type	University
Name	La Trobe University
Address	Bundoora, VIC, 3086
Country	Australia
Secondary sponsor category [1]	Hospital
Name [1]	Alfred Health
Address [1]	55 Commercial Road, Melbourne, Victoria 3004
Country [1]	Australia

Ethics approval

Ethics application status	Approved
Ethics committee name [1]	Alfred Hospital Human Research Ethics Committee
Ethics committee address [1]	C/- The Alfred Hospital, 55 Commercial Road, Melbourne, Victoria 3004
Ethics committee country [1]	Australia
Date submitted for ethics approval [1]	
Approval date [1]	11/12/2014
Ethics approval number [1]	442-14
Ethics committee name [2]	La Trobe University
Ethics committee address [2]	La Trobe University Bundoora, VIC, 3086
Ethics committee country [2]	Australia
Date submitted for ethics approval [2]	
Approval date [2]	05/02/2015
Ethics approval number [2]	UHEC acceptance 442/14

Summary

Brief summary	Impaired arm and hand function is a common and often devastating problem for stroke survivors. Regaining lost movement in the arm/hand is more difficult to achieve than walking, with only 5% of people with hemiplegia regaining functional use of their hand. This devastating outcome could potentially be addressed, however we do not yet know how to best increase movement in the arm and hand after stroke for patients with spasticity. There is a lack of randomized controlled trials of botulinum toxin A (BoNT-A) with a group who does not receive therapy in some dose, and so whether gains were achieved through BoNT-A or a combination of the BoNT-A and therapy cannot be determined from the studies to dateThe research project is testing whether intense therapy given after botulinum toxin injections into the arm is more helpful than just the injections alone.
Trial website	
Trial related presentations / publications	
Public notes	

Contacts

Principal investigator	
Name	A/Prof Natasha Lannin
Address	Alfred Health Occupational Therapy Department 55 Commercial Road, Prahran VIC 3181
Country	Australia
Phone	+61 417135153
Fax	
Email	n.lannin@latrobe.edu.au

Contact person for public queries	
Name	A/Prof Natasha Lannin
Address	Alfred Health Occupational Therapy Department 55 Commercial Road, Prahran VIC 3181
Country	Australia
Phone	+61 417135153
Fax	
Email	N.Lannin@latrobe.edu.au

Contact person for scientific queries	
Name	A/Prof Natasha Lannin
Address	Alfred Health Occupational Therapy Department 55 Commercial Road, Prahran VIC 3181
Country	Australia
Phone	+61 417135153
Fax	
Email	N.Lannin@latrobe.edu.au

No information has been provided regarding IPD availability

Summary results

No Results

Study 3 PROSPERO registration

NIHR | National Institute for Health Research

PROSPERO
International prospective register of systematic reviews

Citation

Shannon Pike, Natasha Lannin, Anne Cusick, Kylie Wales, Lynne Turner-Stokes, Stephen Ashford. A systematic review to evaluate the psychometric properties of measures of function within adult neuro-rehabilitation. PROSPERO 2014 CRD42014013190 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42014013190

Review question

The objectives of this systematic review will be:1. To classify the functional outcome measures reported by Ashford and Turner-Stokes 2013, according to whether activity and or participation outcomes following upper limb spasticity rehabilitation are being assessed; activity performance and participation will be defined according to the International Classification of Functioning model and
2. To locate all of the existing evidence of the properties of the outcome measures, to evaluate the strength of this evidence through the application of the Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) checklist , and come to a conclusion about the best measure available for the particular purpose of measuring activity and/or participation outcomes following upper limb spasticity rehabilitation.

Searches

A search will be conducted in MEDLINE, CINAHL and EMBASE databases. In MEDLINE, a validated search filter for finding studies on measurement properties will be used. Searches with the names of each included instruments (in the title) in combination with the terms for the study population will be conducted until each instrument has been searched.
Searches will be restricted to English language only and abstracts available.
Additional search strategy information can be found in the attached PDF document (link provided below).

Types of study to be included

All quantitative studies published in peer reviewed literature evaluating psychometric properties of the identified outcome measures will be considered. Inclusion criteria for articles will be:1. The aim of the study should be to develop or evaluate the measurement properties of a measurement instrument identified in the review published by Ashford and Turner-Stokes;2. The instrument should aim to measure activity performance or participation, as defined by the ICF.Activity performance is defined as "the execution of a task or action by an individual" Participation is defined as "involvement in a life situation".3. The instrument is evaluated in patients with spasticity (as defined by the authors of the included studies) or patients before or after botulinum toxin type - A therapy.

Condition or domain being studied

Occupational performance outcome measures in adult spasticity rehabilitation; Action Research Arm Test (ARAT), Assessment of Quality of Life, Barthel Index, Chedokee McMaster Assessment, Disability Assessment Scale, European Quality of Life Five Dimension Scale (EQ-5D), Frenchay Arm Test, Modified Frenchay Arm Test, Functional Independence Measure (FIM), Global Assessment Scale, Goal Attainment Scaling, Goal Attainment Scaling using 10-point categorical scale for daily activities, Klein-Bell Activities of Daily Living Scale, Leeds Adult Spasticity Impact Scale, Patient Disability and Carer Burden Scale, Medical Outcomes Study 36 items Short-Form Health Status Survey (SF-36), Motor Activity Log (MAL), Motor Activity Log – 28 (MAL-28), Motor Activity Log – 5 (MAL-5), Modified Motor Assessment Scale (MAS), Motricity Index, Oxford Handicap Scale, Rivermead Motor Assessment, Stroke Adapted Sickness Impact Profile (SA-SIP30), Stroke Impact Scale, nine-hole peg test, Timed 10 metre walk.

Participants/population

Inclusion criteria: Participants will be adults of 18 years of age or older presenting with upper limb spasticity. Participants will be engaging in rehabilitation programs (with or without the inclusion of Botulinum toxin type - A therapy) implemented to achieve activity and or participation goals related to their upper limb spasticity. Participants may be engaging in the rehabilitation program whilst a hospital inpatient, transitioning to home or be community dwelling.

Intervention(s), exposure(s)

This review is concerned with outcomes of upper limb spasticity rehabilitation that identify changes in the performance of an activity or participation as defined by the ICF. Studies that measure activity performance and participation will be included. Studies that measure upper limb spasticity rehabilitation outcomes through assessment of upper limb impairments only, including pain, range of movement, contracture and hypertonicity, will be excluded.

Comparator(s)/control

Not relevant.

Context

This review is concerned with outcomes of upper limb spasticity rehabilitation that identify changes in the performance of an activity or participation as defined by the ICF. Studies that measure activity performance and participation will be included. Studies that measure upper limb spasticity rehabilitation outcomes through assessment of upper limb impairments only, including pain, range of movement, contracture and hypertonicity, will be excluded.

Main outcome(s)

Comprehensive evaluation of the measurement properties of outcome measures assessing activity performance and participation goals for adults with upper limb spasticity undergoing rehabilitation.

Measures of effect

Quantitative synthesis of methodological quality of the available evidence for the psychometric properties of each measure.

Additional outcome(s)

Classification of the identified functional outcome measures according to whether activity and or participation outcomes following upper limb spasticity rehabilitation are being assessed; activity performance and participation will be defined according to the International Classification of Functioning model (World Health Organisation: International Classification of Functioning, Disability and Health. Geneva: World Health Organisation; 2001).

Measures of effect

Quantitative synthesis of methodological quality of the available evidence for the psychometric properties of each measure.

Data extraction (selection and coding)

Searches with the names of each included instruments (in the title) in combination with the terms for the study population as described above will be conducted until each instrument has been searched. Reference lists of the included articles will be reviewed manually by the first author to identify additional eligible articles. Once all searches have been exhausted, the abstracts will be downloaded into the reference management system EndNote and duplicates deleted. A study deemed a duplicate will have authors, setting and location, outcome measures implemented, date and duration of study in common.

The eligibility criteria will first be applied to the title and abstract and if deemed relevant the full manuscript will be retrieved to determine eligibility of potential studies. The initial screen and selection will be completed independently by the first author with the second author reviewing the decision. Debate on the inclusion or exclusion of studies will be resolved by an independent third reviewer and discussion between all three reviewers to reach consensus.

Details on studies that were initially selected based on title and abstract, full text articles that were retrieved and articles included in the review will be documented. Reasons for the exclusion of retrieved full text articles, particularly in the case of doubtful articles will also be recorded.

Data will be extracted from selected studies utilising a standardised data extraction form. This form will record information related to participants, study design, description of botulinum toxin therapy and rehabilitation program(s), outcome measures administered and their classification according to the ICF (activity performance and or participation focus), psychometric properties, study inclusion/exclusion criteria if available and a brief summary of the findings.

Risk of bias (quality) assessment

The quality of the individual studies will be assessed through the application of the Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) guideline for systematic reviews of patient-reported outcome measures (Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. 2018, 27(5):1147-57. doi:10.1007/s11136-018-1798-3). The determined quality of the studies will then form the basis for the planned synthesis.

Strategy for data synthesis

Once methodological quality has been determined, quantitative synthesis of the data will be completed via application of Terwee's criteria relating to good measurement properties. Terwee's criteria provides definitions relating to nine measurement properties (content validity, internal consistency, criterion validity, construct validity, reproducibility, responsiveness, floor and ceiling effects and interpretability), which are considered essential in high quality assessment tools. With the use of Terwee's quality criteria, decisions can be made regarding which functional outcome measure is the highest quality (Terwee CB, de Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC: Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007, 60:34-42).

Analysis of subgroups or subsets

Pending the results of the search strategy the population may extend to upper limb motor impairment with a sub section pertaining to upper limb spasticity created within the review.

Contact details for further information

Shannon Pike
shannon.pike@health.nsw.gov.au

Organisational affiliation of the review

La Trobe University, Bundoora Campus, Vic, Australia
http://www.latrobe.edu.au/

Review team members and their organisational affiliations

Mrs Shannon Pike. La Trobe University, Bundoora Campus Australia
Dr Natasha Lannin. La Trobe University, Alfred Health
Professor Anne Cusick. University of Wollongong
Ms Kylie Wales. University of Sydney
Dr Lynne Turner-Stokes. Kings College London
Dr Stephen Ashford. Kings College London

Anticipated or actual start date

23 June 2014

Anticipated completion date

01 March 2019

Funding sources/sponsors

None

Conflicts of interest

SP was sponsored by Ipsen Pty Ltd to present "Effective Spasticity Management for Stroke and other conditions: the Team Approach and Botulinum Toxin" to general practitioners in the Murrumbidgee Local Health District in 2010. SP has no other financial relationships with any organisations that may have an interest in the submitted work. The remaining author(s) declare that they have no competing interests.

Yes

Language

English

Country

Australia, England

Published protocol

https://doi.org/10.1186/s13643-015-0076-5

https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-015-0076-5

Stage of review
Review Completed published

Details of final report/publication(s) or preprints if available
Pike S, Cusick A, Wales K, Cameron L, Turner-Stokes L, Ashford S, Lannin, NA (2021) Psychometric properties of measures of upper limb activity performance in adults with and without spasticity undergoing neurorehabilitation– A systematic review. PLoS ONE 16(2): e0246288. https://doi.org/10.1371/journal.pone.0246288

Subject index terms status
Subject indexing assigned by CRD

Subject index terms
Adult; Muscle Spasticity; Occupational Health; Outcome Assessment (Health Care)

Date of registration in PROSPERO
26 August 2014

Date of first submission
01 November 2018

Stage of review at time of this submission

Stage	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	Yes	Yes
Data extraction	Yes	Yes
Risk of bias (quality) assessment	Yes	Yes
Data analysis	Yes	Yes

Revision note
Status of review updated

The record owner confirms that the information they have supplied for this submission is accurate and complete and they understand that deliberate provision of inaccurate information or omission of data may be construed as scientific misconduct.
The record owner confirms that they will update the status of the review when it is completed and will add publication details in due course.

Versions
26 August 2014
10 June 2015
28 November 2018

14 December 2020
13 April 2021

Appendix D: Supplementary material from Study 1

- Semi structured interview guide

Semi structured interview guide

InTENSE Clinical Trial

Optimising Upper Limb Recovery after Stroke

Qualitative Interview Guide for participant interviews

Facilitator's welcome, introduction and instructions to participants
Welcome and thank you for agreeing to take part in this interview. You have been asked to participate as your point of view is important to us. I realise you are busy and I appreciate your time.

Introduction: This interview is to help therapists and other patients better understand what it is like to receive the exercise program after Botox™ injections. We are really interested in understanding what you felt were the key features of the therapy you received in the InTENSE trial. What therapies you found beneficial, how therapy can be structured better; what outcomes you've seen; and how the physio/OT might have helped (or not).

May I record the discussion to ensure we are accurately recording what you have discussed? (if yes, switch on the recorder)

Anonymity: Despite being recorded, I would like to assure you that the discussion will be anonymous. The recording will be kept safely in a locked facility until they are transcribed word for word, then it will be destroyed. The transcribed notes will contain no information that would allow individual people to be linked to specific statements- even if you say someone's name today, we delete this in our transcriptions so that things that are discussed are not traced back to individual patients.
If there are any questions or discussions that you do not wish to answer or participate in, you do not have to do so; however please try to answer and be as involved as possible. Do you have any questions? Are you happy to continue?

Introductory question: I am just going to give you a moment to think about what it has been like for you, taking part in the InTENSE trial, but also more broadly, what is it like having spasticity. Are you happy to share this experience?

I would like to understand your views on the burden of spasticity after stroke - on you as an individual, on your family and friends, perhaps too, on the broader community and health system?

- To start, do you mind telling me a little about your stroke?
- And when did the spasticity start? How long have you had spasticity? What do you notice as the spasticity develops?
- What have you done, over the years, to help this spasticity? Has it worked?
- Have you had Botox™ injections before? [If yes: Did you have physio or OT for your arm after the injections in the past?]
- What type of support do you get from family, community health or others with looking after your health?
    - What kind of roles/responsibilities do you have in your family
    - Has having a stroke changed your role in the family? What about the spasticity- did that affect things, or were things pretty much the same?
    - Have you been able to do things since having the therapy that you had to stop doing since the stroke?
- Is there something else that you would like to say, that we have not talked about in this interview?

Appendix E: Supplementary material from Study 2

- Semi structured interview guide (see Appendix D)
- ICF linked data workbook

Body Functions (b)				Coding Examples
		b126 Temperament and personality functions		92 "My personality [didn't change]" 92 "but I'm still, I think, happy, funny" 33a "No, it was the frustration" 92 "My family and my fiancée are like, okay, can we come? Sure, but that's my decision sort of thing" 92 " I think I'm quieter because of the stroke" 33 "Admittedly, [participant name] is quite an easy person to get onwith. He's very mild since - that's the wrong thing to say, he's very mild since his stroke."
			b1265 optimism	88 " was just hoping that something would happen" 88"I thought, well something might work sooner or later"1 I was fully open to see what happened. 33 "Well, it was the final thing. I looked at it as his final chance to maybe knit a path somewhere in the brain "
			b1266 confidence	88 "Yeah, so you lost confidence.[Interviewee88]: Yeah."
			b130 energy and drive	121 "I thought I'd give it a go, purely because I wanted to see if there was a difference," 121 "I'm not sure. I was quite prepared to give it a go. Nothing really would have stopped me."

Body Functions (b)				Coding Examples
Mental Functions b1	Global Mental Functions	b130 Energy and drive functions	b1301 motivation	88 ". I don't think I realised much at all. I did what I was told" 88 "I really wanted to try to do something" 88"I now understand some of the things I was supposed to be doing, and as the comprehension got better I've been able to think about what I really should be doing" 88 "I started to realise what I was supposed to do" 88" I think I started to understand much better what I was supposed to do. I found then the exercise, well seemed to be really good, real better. But I'd say it's just maybe time, but I was doing it properly" 88 "I realised what I'm trying to do, and then things got better" 88 "I'm actually doing what I'm supposed to be" 3 "needed more new challenges" 3 "My goal" 92 "I wanted to get better really" 92 "I have a self-independenceof spirit" 92 "goals" 39 "I know myself I wouldn't do it at home" 39 "I am that person when I don't have a whipunder my hat, that means I do everything except what I supposed to do" 39 "motivation" 39 "I need somebody - maybe not only me. Maybe the people who live alone, they need somebody. Because in my situation I am just waking up and do everything when - to actually leave the home as quick as possible. Thatmeans it counts for exercises, for anything else. When I got the motivation I have to come see somebody" 1"really push myself" 1 "I try and do as much as I can myself " 1 "partly a result of me pushing myself" 1 "waspretty keen to try anything to get any sort of improvement in my movement" 1 "threw myself at it to give it the best chance possible" 1 "I can't remember what my goals were" 22 "I couldn't be bothered thinking about it, what to do and anything" 22 "hoping" 123 (goals) moving my arm and stuff like that" 33b "he just had such adetermination to make a link somewhere and to make the arm move" 33b "we were looking for the light at the end of the tunnel to hopefully achieve something." 33b "I suppose our biggest hope"- was if the fingers would just grasp something so he could use the hand even 10 per cent or 20 per cent" 33b "did give him a lot of hope" 121 "So, I've got movement up here and one day it will eventually come down to my hand. That's what I'm hoping for." 33 "That what another reason why we wanted to do the trial, to see if - well, not to prove anybody wrong, but to see if his arm would move because" 33 "Yes. That was good, actually, because she measured things and made certain [participant name] was in the right chair position and stuff like that. The table height was perfect for him to roll forward and back on with his arm when he was doing hisstretching forward exercises. No, that was very helpful for us." 121 "To get movement in my hand".
		b140 attention functions		22 "concentrate" 33 "Unfortunately, with his stroke, he gets headaches easy, so when he was concentratingfor so long and so hard, we had to cut things short a couple of times, or a few times, actually, because of theheadaches.
		b144 memory		92 "I don't remember the first couple of days, but I can remember the first week." 1 "Okay. So it was fairly early on, was it, during your [rehabilitation hospital] admission that you noticed that your arm was tighter? Interviewee[1]:I guess so - it was all a bit of a blur to me [interviewer name]." "So you can't remember much of that time? Interviewee[123]:No, definitely not. Facilitator:When did you start to become aware of the fact that something had happened? Do you remember? Interviewee[123]:Oh, probably - oh, jeez, no can't remember."

Body Functions (b)				Coding Examples
	Specific Mental Functions	b152 emotional functions		22 "at the moment I feel a little bit stuck in my four walls so to speak" 92 "I am super independent, I'm stubborn on that. But spasticity for me is a mild distraction" 92 It's like an insult. 92 "that will be a burden. " 92 "my life has definitely changed.." 1 "spasticity is a bit of a bugger some days" 1 "bloody nuisance" 1 "makes me feel better" 33 "No, it was the frustration" 121 " Yeah, that was frustrating." 33 "Mentally, he wasperfect…" 121 "Yes. So, it's been tough, it has. It's been very tough."
		b164 Higher-level cognitive functions		88 "It changed when I couldn't work it out" 88 "I just couldn't understand" 88 "that changed because I knew Icouldn't actually do em" 39 "I knew I couldn't do that what I used to do" 88 "My comprehension has moved up a huge amount" 88 "my comprehension has started to move in my mind" 88 "my mind [and my hand] didn't work properly" 1 "Generally I sort of got to a point where I accept the fact now I am the way I am and that I'm not going to get any better" 92 "they were like, well you can go to the gym. I can see their point. They've got other patients. So, I went to the gym and found, by Google, exercises." 92 "So it's pretty much -when you say by Google, you mean that you came up with the program yourself? [Interviewee92]: Yeah. Trial by error. Yep." 92 "I'm not an expert "
			b1644 Insight	121 "I think it made me more aware of what my arm's capable of doing.Facilitator: So, more understanding your limitations? Interviewee: Yes, that's right" 33 "It didn't - look, we did speak about it before we started to say that this may work or this may not work. When we had - I think halfway through it we had a little assessment and [participant name] was - oh, was it maybe at the beginning? [participant name] struggled and he couldn't do any of it. We spoke about it in depth saying – talked to each other about the fact that, we'll do this, but it maynever work, so at the end, try not to be too down on yourself. He'd done his - he'd - look he had done his fantastic best. It was just great. So we conceded defeat, but it wasn't in a bad way. It's, well, we've done thisand this is what we've achieved and maybe one day down the track a miracle might happen, and we left it atthat [laughs]."
		b167 mental function of language		88 "a year ago I was talking much, much better, but I had to think about a word because I couldn't say it. Now I generally get around 95 per cent of them, and then I've got to think about it or work it around. 33 "He finds it hard to take in information straight away. So if he sees the picture of what he's meant to do, it was somuch easier for him to achieve."
		b180 Experience of self and time functions.	b1800 experience of self	22 "Oh yeah, Trevor yes - Trevor (name for arm)" 121 "At the moment, nothing (roles/responsibilities)" 121"Do you think having a stroke's changed your role in the family? Interviewee: Definitely, yes." 121" I couldn't drive, I couldn't get around."

Body Functions (b)				Coding Examples
b2 Sensory Functions and Pain	Hearing and Vestibular	b235 Vestibular functions	b2351 balance	1 "balance"
	Additional Sensory Functions (b250- b279)	b260 proprioceptive		121 "What about knowing where your arm is without looking at it?Interviewee:No."
		b265 touch function		49 "not numb" 49 "getting more sensation going into it" 49 "I've got feeling" 121 "I can't feel my left arm, putit that way"
		b270 Sensory Functions Related to Temperature and other Stimuli	b2700 sensitivity to temperature	49 "if I ran some cold water goes on it - it feels like it's burning" 49 "if something hot falls on it, that's when itreally kills" 121 "Okay, so you've got no sensation as well? Interviewee:No, just cold." 49 "Yeah, have you always had sensation in your hand or has that been something that's… Interviewee[49]:Not much. 121 "If something cold touches it, I feel it."
			b2702 pressure	39 "You had more control over the pressure...[Interviewee39]:Yes, yes, yes. Facilitator:that you were able to apply."
	Pain	b280 Sensation of pain.		121 "I had some pain, yeah, but whenever I had pain I stopped. Facilitator:Okay, so that was something that made you stop doing the exercises? Interviewee:Yes, if something was uncomfortable, I'd stop." 121 "I think I got more sensation out of it, too. I think. Because I can feel pain, and when I felt pain, I stopped. " 39 "What's your sensation like in this hand?Do you feel it when you hook yourself? [Interviewee39]:Oh yeah, I do. Sometimes... Facilitator:It's very painful. [Interviewee39]:Yeah, because the hook has got that thing you can - when you want to take it out it won't let you..."
b4 Functions of the Cardiovascular, Haematol	Additional Functions and Sensations of the			88 "By the time you do a complete shower and get dressed again, and that sort of thing, by the time it's[laughs] - yeah, you're tired 88 "I get tired, but not exhausted" 88"its not so tiring" 121 "It was tiring somedays, yes, but that was understandable, you know?" 121 "Well, I did feel drained some days, but that's understandable."

Body Functions (b)				Coding Examples
ogical, Immunological and Respiratory Systems	Cardiovascular and Respiratory Systems	b455 exercise tolerance function	b4550 general physical endurance	1 "the best thing I can do is improve my stamina and my fitness" 1 "So you focus a lot more on your general fitness. Interviewee[1]:□do, yes."
	Functionsof the Joints and Bones	b710 Mobility of joint functions		88 "They were stretching my arm" 92 "if not it will get like a claw" 92 "range of motion" 92 "stretch" 92 "range of motion" 39 "i do all the stretches" 1 "gives my arm and fingers a good stretching" 1 "muscles a bitof a stretch" 1 "stretch it out and straighten it out" 1 "stretch my arm myself" 49 "but im improving from always stretching" 121"Just a lot of stretching" 121 "I just keep doing stretching as much as I can." (121) "The only thing is that I might do stretch my arm myself, you know like, when I'm sitting here watching TV I'll straighten it out" 1 92 "but just to combat thespasticity because if not it will get like a claw" 121 "But I still do my stretches as much as I can. "
		of joint		49 "subluxation in my shoulder"
	Muscle Functions	b730 Muscle power function		88 "I was getting a lot more movement" 88 "I've been trying to get the strength" 88 "I'm trying to get the strength" 92 "I have no strength in my arm" 92 "no strength" 92 "But in my arm, it was just so weak before and it is now a little bit of strength" 1 "My arm's still probably, pretty much hangs in its own place" 92 " it wasjust one of the dormant muscles was a little bit awake but with no strength."
			b7300 Power ofisolated muscles and muscle groups	88 "I got a little tiny movement"
			b7301 Power of muscles of onelimb	92 "My hand was paralysed"
			b7302 power of muscles of on side of the body	121 " it left me paralysed down the left side."

Body Functions (b)				Coding Examples
b7 Neuromusculoskeletal and movement related functions	Movement Functions	b735 Muscle tone functions		88 "[It got better when I was in bed it used to] relax" 88 "That would relax a bit" 92 "minor spasticity" 92 "spasticity" 92 "spasticity" 1 "spasticity" 1 "fingers became very loose" 22 "tighter" 22 "tight" 22 "tight" 49 "tightness" 123 "my arm was up here most of the time, but now its relaxed" 121 "So, it's been tight as well asweak, has it? Interviewee:□That's right, yes." 121 " I'd get, not seizures as such, but sort of like cramping and that in the leg" 49 "but I didn't notice it was cramped like that." 92 "I think it's either the same, mild, or a little bit actually decreased because mainly of my own exercise." 22 "So with effort it increases does it? Interviewee[22]:□Yeah Facilitator:□.like when you're doing things and… Interviewee[22]:□Yes."
		b750 Motor reflex functions		49 "only moves when I yawn" 49 "because I've got a reflex action"
		b760 Control of voluntary movement functions		88 "[I knew I had] no movement" 88 "like trying to get them up there and here like that" 88"I know I use all my shoulder rather than my arm" 88 "The worse thing I thought was when I had to move my arm, well I'd use my whole shoulder, going all the time" 88 [because both my mind and] my hand didn't work properly". 39 "I couldn't control my left" 39 "it starts to work. That means I could control" 1 "could maybe swing it (for walking)" 123 "move my arm" 33b "it probably moved on the table, when he was trying to move the arm towards him, maybe a centimetre or two centimetres, but we still couldn't determine whether it was the shoulder doing it or whether he was doing it." 33b "he can move his arm up a little bit, " 121 "Well, I was expecting to get movement back in my hand, my arm. But I think I've got movement in my upper arm." 121"But nothing in my hand." 121 "in the hospital I couldn't move it." 33b "we had been told a couple of years before that his arm would never move" 121 "Actually, I hopped out of bed and I fell to the ground because I couldn't use my left leg, either.it." 88 "to try and get me to just move a little bit. But not really with result. For example, only during the trial, and really just at the end, where I can be relaxed and put it in another area, rather than have to push it." 88 "my arm and my leg is working better, and the whole thing, it'smuch, much easier than it was." 33"he put his arm straight next to him, down the side of the chair, over the arm of a chair and straight down. He was able to bring it up over the arm of the chair. Now, how he did that we don't know, but that was his party trick [laughs]." 33 "He was lifting his shoulder, which is fine, but to get itover the arm of the chair, he had to bend the elbow a little and bring the arm over. He managed to do it somehow. We couldn't figure out what - whether it was him moving the body or just what it was" 33" He could move a little bit, but nothing to achieve lifting up anything or grabbing anything to help" 22 " I went to put my right hand out to get the change I couldn't, I could only put my left hand out." 22 "I still didn't think anything of it but as I was walking back I noticed my leg was starting to go "
		Involuntary movement pattern	b7651 tremor	1 "tremor in my left arm" 1 "I'll stand up and my arm starts to tremor. I have to grab it and get it under controlso it's normal." 92 "shaking"

S Body structures			Coding Examples
S4 Structures ofthe Cardiovascular, Immunological and Respiratorysystems	Structure of the cardiovascular system (s410)	Heart (s4100)	"I had a disease in my heart " 92
		Arteries (s4101)	49 " aortic dissection", 49 "repairing the heart"
S7 Structures related to movement	s720 structure of theshoulder region		92 "my shoulder"
	s730 Structure of upperextremity		92 "forearms" 92 "hands and fingers 1"arm and fingers"
	s750 structure of lowerextremity		92 "my leg, in my hamstrings" 22 "onlyfor my legs."

d Activities and Participation				Coding Examples
d1 Learning and applying knowledge	Basic Learning (d130 - d159)	d138 acquiring information		121"So, do you feel like you have learned something from being on the trial? Interviewee:haven't learned a great deal, but I hope you have."
	Applying knowledge (d160 - d179)	d163 thinking		"I knew [I had no movement]" 39 "I didn't expect much because I know that now that it is a slow process" 88"I knew I had no movement" 88 , I don't know, I haven't really thought about (changes to roles) 22 "this couldn't be stroke - because I know about strokes from - in teaching, you do…"92 "I don't know the term (spasticity) but like, okay" 39 "I'm just looking for the reasons to actually - because there's much more important things than do therapy…" 3"I know the theory with the Botox" 92 "one of the doctors said, okay, - they used the term spasticity. I don't know the term but like, okay."
		d175 problem solving		121 "So, you've got adaptive ways around things.Interviewee:That's right, yes. I've had to."
d2 General Tasks and Demands		d210 Undertaking a single task		92 ". It takes me five minutes versus a one minute normal person, but I take my time. "
		d230 Carrying out daily routine		88 "but doing anything during the day, nothing" 92 "I can do almost every normal activity on my own" 121"So, everything I've been doing, I've had to re-adjust. I do things differently these days" 33"It was a new habit that it was hard to found time for initially, but once you established it…"
		d240 Handling stress and other psychological demands		88 "I used to do it (cooking) when I was working, to stop thinking about work.
d3 Communication		d330 speaking		88 "I couldn't talk at all" 88"I was talking much, much better" 92 "I'm still suffering from expressive aphasia" 88 "talking "
		d350 conversation		92 "I talk to my parents once a week. I talk to my sisters twice a week. My fiancée every day "
	Changing and maintaining Body Position	d410 changing basic body position		1 "sitting down I'll stand up"

d Activities and Participation				Coding Examples
d4 Mobility	Carrying, Moving and Handling Objects	d430 lifting and carrying objects		88 "I help move it because I've still got a lot of strength in my left hand" (shopping) 1 "carry something in the hand" 22 "put it to my lips (cup)" 22 "put it down"
		d440 fine hand use		88 "I find it very difficult to use my right hand" 88 "I can hold onto things" 88 "I can't do a lot of the games because you need two hands, and I can't deal with that" 3 "my hand function is still not functional. I use my hand lots of things currently. Not 100 percent, far from it. It definitely helps" 3 "hand definitely improving" 39"slowly I find out it works" 39 "hooking the bait and I would never hook myself in the finger from that time." 39 "Sometimes I hook myself in the finger too..." 1 "I can grasp, like I can actually grab a hold of something"39 "I find out the way to switch it off in my - that means because that is automatically when I put something between two of them, what I was squeezing no matter what I want it or not. After simulation and all that stretching exercises it helped me"1 "trying to get my hand open and get something into my hand is difficult" 22 "hold a cup"
		d445 hand and arm use		88 " it felt like it didn't do anything anymore" 88"everything used to be terribly hard to do" 88 "I'm using bothhands a little bit" 88 "the arm moves a bit, it's just because I can move there, it's so much easier to do things" 88 "Well, I couldn't do anything, now I can do something, and that's good" 3 "improving my hand function" 3 "hand function was a little bit limited" 92 "not enough to actually do any kind of functional exercise or functional position" 39 "You need both hands to do it (fishing)" 33b "but he can't grab anything or anything like that." 33b "I suppose our biggest hope was if the fingers would just grasp something so he could use the hand even 10 per cent or 20 per cent, " 121 "So, I haven't been able to use my left arm" 121 "I haven't been able to use my left arm at all" 39 "Some improvement with my left hand." 88 "Fortunately I'mstill quite stronger than I was, and if I can do it left handed, I can do it. But anything two hands, very difficult"121 " I tried to incorporate my arm in doing things, but it couldn't. So, I couldn't grip on things." 123 "Do youthink that part of the reason all of the - how much of the reason why you need help with showering is that because of your arm?, Interviewee[123]: Yeah, I think so., Facilitator: Largely or somewhat?, Interviewee[123]: Seventy-five per cent I think, yeah. Facilitator: So it's a factor? Interviewee[123]: Yeah, definitely." 39"Yes, yes, yes. Even now is - when I stop actually that InTENSE trial still I can actually manage it. That means something stays there." 39 "when I want to open up, no please go to left one. Especially at home because that's when I feel the - because outside, outdoors, that I'm using, for safety reasons, the right one." 88 "So this hand wasn't doing very much? [Interviewee88]: No. Facilitator: This hand was doing all the work? [Interviewee88]: Yep." 33"When it camedown to the final evaluation of the program, I suppose, in a nutshell, he basically failed every step becausethe arm just didn't move and the fingers didn't move. So he was unable to lift objects up and place them over a little wall or grasp anything, which is - yeah, which was a shame, but he tried."
			d4450 pulling	39 "just using it to open up the door"
			d4451 pushing	88 "[usually my hand to] push things"

d Activities and Participation				Coding Examples
	Walking and Moving	d450 walking		92 "I can walk" 92 "learning to walk" 92 "I'm still wobbly" 92 "walking" 92" but i have to start walking" 92 "mywalking is improved" 1 "walking" 1 "I can probably walk a bit further" 1 "I did think that maybe I could walk a bit better if my arm was a bit freer" 92" my walking is improved" 121"I couldn't get around"
		d460 moving around in different locations		1 "walk up and down every aisle of the supermarket" 1 "walk a whole city block"
	Moving Around Using Transportation	d470 using transportation		he drove me and drove me back. 33"we would catch the train to [city suburb] Station and then we wouldjust catch a taxi to the hospital. It worked - it was comfortable. It was no pressure in driving up and down."
		d475 driving		92 "I can drive" 92 "I'm driving the work van" 121 "I couldn't drive, I couldn't get around." 121 "Yes, and because I don't have a car, I couldn't drive, so he had to bring me in. So, it was a big ask from him, you know?" 121 "Because if I was able to drive, I had a car, then I would have been able to come in myself. That would have made life a lot easier."
d5 Self-Care		d510 washing oneself		88 "complete shower" 88 "a shower, I just enjoy it, it's easy and not worry about" 123 "showering" 123 "showering" 121 "showering." assistance with 88 "Are you able to have a shower"
		d520 caring for body parts		22 "go to do my nails"
		d540 Dressing		88 "could pull a zip up" 88 "using clothes is much better than it was" 88 "get dressed again" 88 "Making itzip 88 "my zip, it's a big thing. Being able to just hold the jumper or a coat and pull it down, and tuck themup" 88 "I can hold those jumpers" 121"(My parents) help me to get dressed every morning" 121"Did you have any problems with getting your clothes over the cast?" 88 "get dressed yourself? [Interviewee88]: Yeah, completely."
		d550 eating		88 "every Sunday was Sunday dinner" 123 "they usually just go out with me to eat"
	Acquisition of Necessities	d610 Acquiring a place to live		22 "No, because otherwise I was trying to get my house plus his house and yeah." 88 "So we were going tostay where we're living, and then build the two units and sell them, but I got such a good price. I've alreadysold it, and we made - until when the other place is right. Facilitator: So are you going to a rental in between? [Interviewee88]: No, we're able to leave home and go straight there. Facilitator: Oh straight to the new place, and it's ready to go? [Interviewee88]: Yeah."

d Activities and Participation				Coding Examples
d6 Domestic Life		d620 Acquisition of goods and services	d6200 shopping	88 "buy some things" 88"We do that [shopping" 1 "shopping" 121 "If I want to go to the bank or shopping"1 "every Thursday I go shopping"
	Household Tasks	d630 preparing meals		88 "I used to cook then" 1 "I do all my own cooking" 121 "Meals, cooking" assistance with 88 "No, it's too difficult. I'll do it [cooking] again when I can - or if I can get to use it better".
		d640 doing housework		1 ". I do all my own washing" 1 " wash my dishes" 92 "cleaning the house, caring for the clients, cooking. I'm pretty good at that. Occasionally I'm changing the sheets, or the doona covers"
	Caring for Household Objects and Assisting Others	d650 caring for household objects		88 ". Usually [wife name] and I do what we want, we're just about to leave the house. My daughter was going to buy a new house and build it there, because it's one of those double things because she's got a big block, and I thought that – and [wife name], I thought our place is getting pretty old now, and no, it was very good when I was able to do much, but she can't do too much, and I can't. So I suggested that I buy theother half, and she was really wrapped with that"
			d6506 caring for animals	22 "He's got one dog and I've got two dogs so we put the dogs together. "
		d720 complex interpersonal interactions		121 "The thing I found frustrating was my parents.Facilitator:Okay. Interviewee:That was the biggest thing because to them - they got their minds, they got their brains, they've got all their movements to their body parts. They think - you just got to move your arm like this, that's all you've got to do. To me, it doesn't work that way. " 121 "They kept looking at me and saying, look,this is all you've got to do, just move your hand. Just tell your mind to move it. That's all you've got to do. I kept saying, it doesn't work that way. Does that make sense?" 121 "Okay, so there was a tension.... Interviewee:There was, yeah. Facilitator:..because they were helping you but also irritating you. Interviewee:That's right, yes."
	Particular Interpersonal Relationships	d760 family relationships		88 But the kids took over things that I used to do for them or for me automatically because I knew it first. It'sjust how it's worked…"
		d770 intimate relationships		121 " I separated with my wife and my three kids are still in [City name]."

d Activities and Participation				Coding Examples
d8 Major Life Areas	Education	d820 school education		123 " I was finishing high school"
	Work and Employment	d845 acquiring, keeping and terminating a job		121 "I used to work before that. I used to work as a baggage handler at the airport. Obviously when I hadthe stroke I had to stop that, I couldn't work."
		d850 remunerative employment		92 "I'm working almost full-time"
		d870 economic self-sufficiency		121 "Yep, and that - the cost of paying for parking, was ever a… Interviewee:No, I mean my dad normally paid it every day, but it was there. Facilitator:Yeah, so you would have preferred not to have it, but you tolerated it. Interviewee:Yeah." 121"That was good because every time my dad came in here we had to pay $6 for parking, too, which we had to pay. That was $6 every day, "
d9 Community, Social and Civic Life		d920 recreation andleisure		88 "draw" 88 "kept going playing games" 39 "fishing" 22 "pool" 22 "gym" 22 "we go for a coffee" 121 "Facilitator: So, it's more of a leisure… Interviewee: Yeah, pretty much, getting out and about. If I want to goto the bank or shopping." 92 "I'm not an amateur gym goer but I have a specific program with strength, power lifting, gymnastics" 123 "So it's more a leisure… Interviewee[123]:Yeah, I think so."
			d9205 socialising	88 "I can go to the shops" 121 "Yeah. To give you an example, I went to a GP couple of weeks ago and he says to me, why don't you go and do Living Longer and Living Better? I said, what's that? She says, it's a program that's run. I said, oh look, I'll give it a go. Why not? So, I went there once. With all due respect, it was all 70 and 80-year-olds. I did the whole thing, one lesson, and they said look, we come here for the social aspect of it. I said I don't want to come there for the social aspect. What am I going to do, talk with 80year-olds? It was ridiculous. Never went back."

Environmental Factors			Coding Examples
e1 Products and Technology	e110 Productsor substancesfor personal consumption	e1101 drugs	121 "Interviewee: The one thing the doctor suggested, once I when came in here, if I take antidepressants. I said, no, and they prescribed it to me. I think was a good thing because I never would have taken that. Facilitator: So… Interviewee: That was recommended to me. Facilitator: They said do you want antidepressants, you said no, they said, here's a script anyway and then you took it… Interviewee: Well, they recommended it to my GP and my GP put me onto it and that was to help my brain. Facilitator: So, that was really helpful. Interviewee: Yeah, I thought that was a step in the right direction because that was the only doctor who's told me to do that. Facilitator: Was that the doctor here when you… Interviewee: Yes. Facilitator: Ah, okay. Interviewee: Yes, Caroline. Facilitator: So, it was an unexpected… Interviewee: Yeah. I mean, I've never been a person who takes antidepressants, but the way it was explained to me was, it'll help my brain heal. Well, we'll give a go, and that's something no one else had mentioned.
	e115 Productsand technology forpersonal use in daily living	e1151 Assistive products and technology for personal use in daily living	22 "of a night I have a paint roller" 121 " have certain clamps and procedures that I do things with and I've had to adjust…" 49 " Yeah, an E-Stim - I've got one of those." 88 Oh the shower rail, I pulled at it, knocked it over - yeah, I pulled it out.88 "because I've got the AFO." 121 "So, I've been wearing an AFO" 121 "I think the electric stim that I was using felt like it was doing something because it was moving my arm." 121 "the cast wasn't for very long." 121 "Anything else that stood out for you as being particularly helpful? Interviewee: I think the major was the e-stim, yeah." 121"The e-stim, yeah, because that actually had - like, I noticed every time it hit me, my arm would move. That was a positive thing." 121 "The only thing - I mean, obviously now I haven't got an e-stim at home, so I don't do the e-stim anymore. If I had an e-stim, I probably would be still doing it. But I haven't got one, so I don't do it anymore." 121 "I don't know, is there any specialised equipment that I could have used? Any other equipment that I could have used? I don't know. Because the equipment that I used seemed quite primitive." 121 "Well, I had this wooden stand and I dragged it across a tablewith a wooden stand. To me, that seems primitive in 2018." 123 "You had had casting? Interviewee[123]: Yes." 33 "we'd done the electrode machine program where you put your hand on the machine and it moves your fingers, " 33"We had our own machine, TENS machine. It was actually identical to what [rehabilitation hospital] was, so that was no problem. Just all the other little bits of equipment were great, the little moving thing on the stick." 33"Yeah, and we had just our own drink bottles that [participant name] used to try and pick up, but yeah, no, the equipment was fine. I had problems using our own. So it was good." 121 "Yes, because Ihaven't got an electric stim at home. So, I stopped doing that, but the other exercises I keep doing, yes."

Environmental Factors			Coding Examples
	e120 Productsand technology forpersonal indoor and outdoor mobility and transportation	e1201 Assistive products and technology for personal indoor and outdoor mobility and transportation	92 "sitting in a wheelchair 88 "I've got a little cart I sometimes use….(to go to shops) 88 "I zoom off around there sometimes"
e2 Natural Environment andHuman-Made Changes to Environment	e225 climate		121 "because I have to wear long clothes here. It's been very cold and that's been one of the hardest things to adjust, when I've been living in [warmer state] for the last 20 years." 121 "The hardest thing was the weather. It was cold, and to do that, I had to take everything off. It was cold."
	e210 physical geography		1 "distance made it difficult" 33"It was just a long drive for the…" 22 "Later I found out of course they can't fly youbecause of your stroke so it took me I predict about 12 hours to get to [capital city]. " 121 "Yes, it was an hour in, an hour back."
	e310 immediate family		88 "I've got five kids" 88 "her family's big" 92 "girlfriend" 92 "sister" 92 "My family is really close-knit, happy" 92 "My girlfriend is really understanding 1 "my youngest son pops in on a regular basis" 1 "no real family for support" 1 "he does jobs for me, like cleans my garage out and he might do a bit of gardening" 22 "My daughter, when she's home she'll cook for me or sometimes she gets me out of the shower to get me onto the bed" 22 "brothers" 22 "mum" 22 "dad" 121" My parents" - help me to get dressed every morning 121"No, well I had to exercise with them because they helped me exercise. They had to put the e-stim on for me, so they sit there and watch me." 121 "I was lucky because I had my dad to bring. I was very lucky. Facilitator: So, it would have been a big issue. Interviewee: That's right, if my dad wasn't there, I wouldn't be able to get here." 123 "What sort of support do you get from family? Well, I know your mum is very involved. Interviewee[123]: Definitely." 22 "apparently my two children visited me in [capital city]. " 22"She's the daughter who's always going to another country to live…" 22 "But yeah, I'm very sorry to hear about your partner. Interviewee[22]: Oh thank you. Facilitator: How long ago was that that he passed away? Interviewee[22]: Oh about 10 weeks. Facilitator: Ooh, so very recently. Interviewee[22]: Oh yeah. It's all been yeah, a bit of a … Facilitator: A lot of change." 121 "Lucky my dad was there to help me, bring me in all those times. If it wasn't for my dad, I wouldn't have been able to get here. Facilitator: So, was that a strain for him and for you, getting - like, the travel side of things? Interviewee: It wasn't a strain, but for him to be onthe road every day or every time… Facilitator: It was commitment. Interviewee: Yeah, big commitment. Especially in [city name] with all the traffic." 121 "it was a big ask for my dad to come in here every day." 22"Yes so the other daughter, no she lives in Geelong and... and for anyone else in my family, no."
	e320 friends		39 "I ask my mate to hook it for me" 1 "Not a lot of friends. I've got a mate who drops around every now and again"22"My best friend - I was just talking about [best friend name] - she came up. Another friend and yeah. Then the next week I do remember another couple of friends visiting me but with my two children and my best friend visiting, I have no recollection of them at all."

Environmental Factors			Coding Examples
e3 Support and Relationships	e325 Acquaintances, peers, colleagues, neighbours and community members		88 "through groups"
	e340 Personalcare providersand personal assistants		88 "we've both got a help package, so it's washing and that sort of thing" 88 "laundry. Laundry [help with]" 88 "I can't put it back, so you've got to get somebody to do that" 1 "PCA come once a week and takes me shopping andpaying bills" 1 "cleans for me" 1 "She does pretty much the things I can't do like change the sheets on my bed andmakes it, vacuums for me and mop floors and clean the bathroom" 22 "home help sort of person" 123 "two carers"121 "One day a week I have a support worker that comes and picks me up and takes me wherever I want to go". 1"o that when the cleaning lady is here that she can focus on things I can't do. " 88 "So the package that comes in, is it from the council? [Interviewee88]: Yep. Facilitator: Is it just housework? [Interviewee88]: Yeah, for me."
	e355 Health professionals		88 "the physio" 39 "plus the technicality of it (health professional)" 39 "occupational therapist" 1 "physio" "39 "Muchmore important was I'm going to see somebody" 121 "Very good. It was good I came in here because they taught me how to do it properly and if I had any questions, they could explain it to me. Yeah, so it was great." 121"There were days I was at home where certain parts of the stim wasn't working properly, whether I wasn't putting it on rightor what, but when I came in here, we found the right spot again and away we went. Marked it." 121 "No, it was a guidance, you know? Facilitator:Refining the process. Interviewee:That's right, yes." 121"Well even if there were fewer, the opportunity for me to call someone and ask was there too, so that was good." 121 "Yeah, that was fine, yeah. Initially it went – the clinician came out and showed me what I could do at home and ran through the program with me. So, I was guided in that way, yes, so Iknew what to do at home." 121 "They answered any concerns or questions I had, putting it in the right spot or whyit wasn't working." 121 "Because I asked lots of doctors and neurologists, too." 33 "We had a couple of therapists down there and [participant name] got on with them both extremely well. I think it was two or three. I can't remember now. They did change a bit. But still, they were fabulous and they worked with him, oh, wonderful, wonderful." 33 "Yes. That was good, actually, because she measured things and made certain [participant name] was in the right chair position and stuff like that. The table height was perfect for him to roll forward and back on with his arm when he was doing his stretching forward exercises. No, that was very helpful for us."
	other professionals		49 "Yeah, I had acupuncture needles into my scalp - for different sections."

Environmental Factors			Coding Examples
e5 Services, systems and policies	e580 Health services, systems and policies	e5800 health services	88 "very little arm exercises" 3 "(needed) therapy" 39 "OT therapy after that, and my gym sessions" 39 "After exercises, with the girls" 1 "Then I was sent to rehab there at [rehabilitation hospital]". 49 "Because I'm still like seeing the spasticity clinic…" 88 "I tried to get physio" 1 "I still attend physio here" 3 "outpatient therapy" 39 "two groups" 49 "so my first therapy was to [ward at rehabilitation hospital]" 1 "I think I was in [acute hospital] for about three weeks, firstly in intensive care" 92" The neurologist injected 22 "Yes and then I had the first two and a half in [capital city] before I got - we go in the aeroplane [unclear] one to [home city] to then take me into the [rehabilitation hospital]." 121 "Oh, look, I think that everything helps. I think this is a step in my progress" 1 "No, mainly it's weight bearing exercises and I'm just stretching my calf muscle; normally - what we do, we put my body weight and stretching it until my heel touches the ground and then she's happy. I do a bit of walking for her " 123 "Did you have exercises or stretches… Interviewee[123]: Yes." 49"Yeah, and I've had Botox to try and - I've had Botox in my arm as well." 88 "So have you continued to have speech therapy? [Interviewee88]: No, I haven't had that." 88 " the physio I had with " 92 " after the Botox and the cast " 33 " had the Botox - he had his very first Botox injection probably " 33 "at a program place in [capital city]. We did that for quitea while" 22"Yes and apparently I got taken to [small holiday town] Hospital who then took me - I suppose I don't know if it's right or not - to [regional centre] Hospital. All I remember is the ambulance man saying we're going to take you to [capital city] Hospital, we'll be about three hours" 22 "Physio." 121"121 "Had you had Botox injectionsbefore the InTENSE trial? Interviewee: No. Facilitator: Okay, so that was your first experience of Botox. Interviewee: Yes.""

Personal Factors
92 "she has been living with me"
39 "I live alone"
39 " as a single"
123 " I was living alone"
22 "unfortunately he died so he decided in his will I could live there because it's a much bigger house than mine"
22 "My father passed away two years ago"
121 "Oh, right, so you moved down here after your stroke. Interviewee:Yes.
22 "But yeah, I'm very sorry to hear about your partner. Interviewee[22]:Oh thank you. Facilitator:How long ago was that that he passed away? Interviewee[22]:Oh about 10 weeks. Facilitator:Ooh, so very recently. Interviewee[22]:Oh yeah. It's all been yeah, a bit of a … Facilitator:A lot of change."
123 "You were living out of the house then?Interviewee[123]:Yes. Facilitator:Then you moved back? Interviewee[123]:[No…] Facilitator:Well, you're at home now aren't you? Interviewee[123]:Yeah [laughs]. Facilitator:That's all right. Interviewee[123]:Sorry, I'm just a bit confused."

Interview Guide			
Body Functions	What about the spasticity - did that affect things or were things pretty much the same?		
Activity and Participation			Have you been able to do things since having the stroke
			What kind of roles/responsibilities do you have in your family?
			Has having a stroke changed your role in the family?
Environment	e3 support and relationships	e310 immediate family	What type of support do you get from family …..with looking after your health?
		e355 Health professionals	What type of support do you get from community health …..with looking after yourhealth?
		? Others	What type of support do you get from others …..with looking after your health?
Body Structures			

Rules
Map to b735 Muscle Tone Functions when coding spasticity
Each meaningful concept can only be assigned to one code, noting that data (sentences) may hold several meaningfulconcepts that can then be assigned to the different code with the context and overall meaning dictating the final code assigned
Living situation to be documented under Personal Factors
Interval of time is not linked to ICF - as per linking rules
When the meaningful concept is referring to services or accessing services such as "physio" link to Health services code
When the meaningful concept is referring to the support provided by therapists link to code health professionals
The above two rules are based on health professionals falling under support and Health services under Servicesprovided.
Body structures: codes from this domain were applied when the participants were discussing their body part in isolation. If the body structure i.e. arm was discussed whilst referring to use or muscle tone then this was coded to Activity and Participation of Body Functions. In summary the majority of the timer this was the case so minimal coding to body structures occurred
Participants discussed their heart and aorta so this was coded under Body Structures, when participants discussed their"stroke" this term was not coded as it fits within the overarching box "the diagnosis"

Meaningful Concepts Discussed with Research Team		
Quote	**Thoughts**	**Decision**
88 "Well, I've got five kids, although, two of them in are interstate now. Usually [wife name] and I do what we want, we're just about to leave thehouse. My daughter was going to buy a new house and build it there, because it's one of those double things because she's got a big block, and I thought that – and [wife name], I thought our place is getting prettyold now, and no, it was very good when I was able to do much, but she can't do too much, and I can't. So I suggested that I buy the other half, and she was really wrapped with that"	d650 caring for household objects -no clear code for relocation that is the outcome for a limitation in previously mentioned code.	d650
88 "General ones. So the real difficult ones I can't. But the kids took overthings that I used to do for them or for me automatically because I knew it first. It's just how it's worked…" Facilitator: So it was like those roles have been transferred to the next generation? 88 "Yeah, that's right"	family relationships "kids took over""I used it d for them" activity limitation	d60 family relationships
Facilitator: So you are able to tuck your shirt into your… 88 "Yeah, not well, but yeah it's good. *Well, I couldn't do anything, now Ican do something, and that's good.* I keep trying something else. I think one of the things I find the stress, not the finger side now, I've got to walk, like trying to really do it all before time, instead of just when I do the exercises [unclear] all those - no, the each finger, you hold - bring them, but..."	hand and arm function	hand and arm function
92 "I am super independent, I'm stubborn on that. But spasticity for me is a mild distraction"	emotional functions	emotional functions
Facilitator: You don't see it as your biggest problem? 92 No, no. It's like an insult.	emotional functions	emotional functions
92 "if not it will get like a claw and that will be a burden. "	mobility of joints, emotional functions	mobility of joints, emotional functions
92 ". It takes me five minutes versus a one minute normal person, but I take my time. "	undertaking a single task	undertaking a single task
92 "my life has definitely changed.."	emotional functions	emotional functions
39 "I'm just looking for the reasons to actually - because there's muchmore important things than do therapy…"	background of accountability to commit to therapy thought functions	background of accountability tocommit to therapy thought functions

Meaningful Concepts Discussed with Research Team		
Quote	**Thoughts**	**Decision**
39 "I find out the way to switch it off in my - that means because that isautomatically when I put something between two of them, what I was squeezing no matter what I want it or not. After simulation and all that stretching exercises it helped me"	fine hand use	fine hand use
1 "spasticity is a bit of a bugger some days"	"spasticity" coded but not "bugger"b152 emotional functions	"spasticity" coded but not "bugger" b152 emotional functions
1 "bloody nuisance"	emotional functions	emotional functions
121 "Facilitator: You can't - okay. Do you feel like it's got tighter over the two years? Interviewee: No. Facilitator: It's just basically been the same… Interviewee: Yeah, it's plateaued out, yes. Facilitator: .and nothing's changed?	change in spasticity/impairment	not coded
121 "Had you had Botox injections before the InTENSE trial?Interviewee: No. Facilitator: Okay, so that was your first experience of Botox. Interviewee: Yes."	previous intervention / services	Health services
121 "Oh, right, so you moved down here after your stroke. Interviewee: Yes."So, your family - you're now living with your parents, moved in with your parents. Interviewee: Yes."	? Relocation support / services	personal factor
121 "Yes. So, it's been tough, it has. It's been very tough."		emotional functions
121 "Okay, so from objective measures, there's a change, but you haven't noticed a change… Interviewee: Not a great deal, no. Facilitator: .in what you can do or not do. Interviewee: In everyday life, no."	change in function	not coded
121 "Yes, because I haven't got an electric stim at home. So, I stopped doing that, but the other exercises I keep doing, yes."	environmental - assistive devices barrier as lack of access	e115

Meaningful Concepts Discussed with Research Team		
Quote	**Thoughts**	**Decision**
121 "It was good. It was good to - I feel like I was doing something to help. Like, you know, for future people that had a stroke, and if it doeshelp them, that's fantastic. Facilitator:So, part of your motivation was contributing - so it wasn't - this is a question later on - but the things that prompted you to participate initially was - a large part of it was to be part of finding outabout what's the most effective… Interviewee:Yes, absolutely. "	Experience of self	Experience of self
121 "To get movement in my hand.	motivation (this was a reason for participating)	motivation
121 So, you could see a benefit for yourself as well as potentiallycontributing to the body of knowledge. Interviewee:Yes.	experience of self	Experience of self
121 "That was good because every time my dad came in here we had topay $6 for parking, too, which we had to pay. That was $6 every day, which - like I said, it was a big ask for my dad to come in here every day."	Cost, last section is support from dad	economic self sufficiency, immediate family
121 "and also for education reasons. Facilitator:So, do you feel like you have learned something from being on the trial? Interviewee:I haven't learned a great deal, but I hope you have."	acquiring information	d138 acquiring info
121 "and like I said, it was explained things start from the top and they work their way down."		d163 thinking
121 Well, I mean, to see if I can get movement back after Botox. Make me able to move my arm about a lot easier. That was basically it. That was the curious bit, having Botox to my arm. Would it make it easier forme to move my arm? To answer that, I don't know if it did or not.	change in function	not coded
121 "I think it was waking up things in my arm, maybe. Facilitator:So, it felt like a change. Interviewee:Yes."	change in function	not coded
121 Yeah. But, it's knowledge, and I've been going through this for two years now. It's been interesting going through different doctors anddifferent advices.	learning	d163 thinking

Meaningful Concepts Discussed with Research Team		
Quote	**Thoughts**	**Decision**
121 "I would have pulled out, yeah. But I was quite prepared to continueto finish it, to finish the study and help in any way I could."	sense of self	
121 Yep, and that - the cost of paying for parking, was ever a… Interviewee:No, I mean my dad normally paid it every day, but it was there. Facilitator:Yeah, so you would have preferred not to have it, but you tolerated it. Interviewee:Yeah.	economic self sufficiency	economic self sufficiency
1 "The only bad thing about it was putting aside time every day to make sure you did the exercise."	time	nc
1 "So it was very time consuming? Interviewee[1]:It did; like it probably took a good hour a day to bowl it over. I was lucky because I just had Foxtel connected so I got to sit hereand watch music videos while I did them."	time	nc
123 "What about roles and responsibility within your family? Do you have any particular responsibilities or roles? Interviewee[123]:Nope."	not managing daily routine, pf	nc
123 "You were living out of the house then? Interviewee[123]:Yes. Facilitator:Then you moved back? Interviewee[123]:[No…] Facilitator:Well, you're at home now aren't you? Interviewee[123]:Yeah laughs]. Facilitator:That's all right. Interviewee[123]:Sorry, I'm just a bit confused."	relocation, personal factor	pf
123 "So do you think having your stroke has changed your roles is or doyou think it hasn't because - oh, you tell me. Do you think… Interviewee[123]:I definitely think it has."	roles	nc
3 I think steady improvement. I think - I can't tell, or maybe can tell. Very hard to tell with me, but with Botox, sort of [celebration] there improving, I reckon.	change in function	nc
39 "I was committing myself to actually do - I was waiting for that days and I was following the program, what I was told once, minimum half anhour was the first session, but completely about two hours a day."		

Meaningful Concepts Discussed with Research Team		
Quote	**Thoughts**	**Decision**
88 It's just been in the last year or so that you've noticed theimprovement? [Interviewee88]: Well, actually only about three months with the arm side,	change in function	nc
88 So we were going to stay where we're living, and then build the two units and sell them, but I got such a good price. I've already sold it, andwe made - until when the other place is right. Facilitator: So are you going to a rental in between? [Interviewee88]: No, we're able to leave home and go straight there. Facilitator: Oh straight to the new place, and it's ready to go? [Interviewee88]: Yeah. Facilitator: Oh that's fantastic.	acquiring a place to live	d610 acquiring a place to live
33 concentrating for so long and so hard, we had to cut things short acouple of times, or a few times, actually, because of the headaches.	attention	attention
33"Definitely, definitely. That was terrific. That was actually terrific because it helped [participant name] see what he was meant - what positions he was meant to be in and what - how far maybe he was tryingto push something forward or pull something back. So the diagrams were great, really good."	health professionals	health professionals
22 "But yeah, I'm very sorry to hear about your partner. Interviewee [22]: Oh thank you. Facilitator: How long ago was that that he passed away? Interviewee [22]: Oh about 10 weeks. Facilitator: oh, so very recently. Interviewee [22]: Oh yeah. It's all been yeah, a bit of a … Facilitator: A lot of change."	personal factor	PF
33 "It was just setting aside the time to practice the…"	time	nc

Appendix F: Supplementary material from Study 3

- Appendix Table 2. PRISMA-P Checklist
- Appendix Table 3. PRISMA checklist
- MEDLINE Search strategy and search terms
- Appendix Table 4. PRISMA Full text exclusion reasons
- Appendix Table 5. Methodological quality and quality criteria ratings.
- Appendix Table 6. Summary of results for included studies.
- Appendix Table 7. Terwee criteria for good measurement properties
- Appendix Table 8. Criteria to guide strength of relationship
- Appendix Table 9. ARAT COSMIN ratings
- Appendix Table 10. Quality of measurement properties (ARAT)
- COSMIN workbook

Appendix Table 2. *PRISMA-P checklist*

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item
ADMINISTRATIVE INFORMATION		
Title:		
Identification	1a	Identify the report as a protocol of a systematic review
Update	1b	If the protocol is for an update of a previous systematic review, identify as such
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number
Authors:		
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments
Support:		
Sources	5a	Indicate sources of financial or other support for the review
Sponsor	5b	Provide name for the review funder and/or sponsor
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol
INTRODUCTION		
Rationale	6	Describe the rationale for the review in the context of what is already known
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)
METHODS		
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated
Study records:		
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review

PRISMA-P Checklist

Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

PRISMA checklist

Appendix Table 3. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	3, 5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supp file 2
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5-7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7-8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7-8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7-8
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8-9, Figure 1, Supp file 2
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Supp file 4
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	33-54, Supp file 5
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Table 3
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	54-57
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	57
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	57-58
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	58

From: Moher, Liberati, Tetzlaff, Altman, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

MEDLINE Search Strategy

Filters

- Limits: published in English

("title of assessment tool") AND hasabstract AND (instrumentation[sh] OR methods[sh] OR Validation Studies OR Comparative Study[Hinderer, #19] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*(Scottish Intercollegiate Guidelines Network (SIGN)) OR clinometr*(Scottish Intercollegiate Guidelines Network (SIGN)) OR "outcome assessment (health care)"[MeSH] OR outcome assessment[tiab] OR outcome measure*(Scottish Intercollegiate Guidelines Network (SIGN)) OR "observer variation"[MeSH] OR observer variation[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR "precise values"[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR factor

analysis[tiab] OR factor analyses[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR meaningful change[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab])) NOT (("addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference, nih"[Publication Type] OR "practice guideline"[Publication Type]) OR ("animals"[MeSH Terms] NOT "humans"[MeSH Terms])

“Assessment tool” (Replace assessment tool with one below)

- “Action Research Arm Test” OR “ARAT” = Y
- “Assessment of Quality of Life” OR “AQoL” = Y
- “Barthel Index” = Y
- “Chedokee McMaster Assessment” OR “ Chedokee Assessment” OR “Chedokee McMaster Stroke Assessment” OR “CMSA” OR “CMA” = Y
- “Disability Assessment Scale” OR “DAS” = Y
- “European Quality of Life Five Dimension Scale” OR “ Euro-QoL” OR “EQ-5D” = Y
- “Frenchay Arm Test” = Y
- “Modified Frenchay Arm Test” OR “Modified Frenchay Scale” = Y

- “Functional Independence Measure” OR “FIM” = Y
- “Global Assessment Scale” OR “ GAS” = Y
- “Goal Attainment Scaling” OR “GAS” = Y
- “Goal Attainment Scaling using 10-point categorical scale for daily activities” OR “GAS using 10-point categorical scale for daily activities” = Y
- “Klein-Bell Activities of Daily Living Scale” OR “”Klein-Bell ADL Scale” = Y
- “Leeds Adult Spasticity Impact Scale” OR “LASIS” OR “Patient Disability and Carer Burden Scale” OR “8-item patient disability scale and 4-item carer burden scale” = Y
- “Medical Outcomes Study 36 items Short-Form Health Status Survey” OR “SF-36” = Y
- “Motor Activity Log” OR “Upper Extremity Motor Activity Log” OR “UE MAL” OR “MAL” = Y
- “Motor Activity Log – 28” OR “MAL-28” OR = Y
- “Motor Activity Log - 5” OR “MAL – 5” = Y
- “Modified Motor Assessment Scale” OR “MAS” = Y
- “Motricity Index” OR “MI” = Y
- “Oxford Handicap Scale” OR “OHS” = Y
- “Rivermead Motor Assessment” OR “RMA” = Y
- “Stroke Adapted Sickness Impact Profile” OR “SA-SIP30” = Y
- “Stroke Impact Scale” OR “SIS” = Y
- “9 hole peg test” OR “nine-hole peg test” OR “NHPT” = Y
- “10 metre walk” OR “10 meter walk” OR “Timed 10 metre walk” OR “Timed 10 meter walk” OR “10MWT” = Y

Appendix Table 4. *Full text exclusion reasons (PRISMA)*

Outcome Measure	Exclusion reason	n =	Outcome Measure	Exclusion reason	n =
ARAT	Psychometric properties not evaluated	24	ArmA	Psychometric properties not evaluated	2
	Conference proceedings	0		Conference proceedings	0
	Not conducted in English	15		Not conducted in English	0
	Modified or incorrect measure	3		Modified or incorrect measure	0
	Not original research	1		Not original research	0
	Protocol only	0		Protocol only	0
	Participants outside criteria	0		Participants outside criteria	0
AQoL	Psychometric properties not evaluated	2	BI	Psychometric properties not evaluated	56
	Conference proceedings	0		Conference proceedings	1
	Not conducted in English	0		Not conducted in English	23
	Modified or incorrect measure	0		Modified or incorrect measure	15
	Not original research	0		Not original research	5
	Protocol only	0		Protocol only	0
	Participants outside criteria	0		Participants outside criteria	17
CMSA	Psychometric properties not evaluated	14	DAS	Psychometric properties not evaluated	3
	Conference proceedings	0		Conference proceedings	1
	Not conducted in English	0		Not conducted in English	1
	Modified or incorrect measure	6		Modified or incorrect measure	5
	Not original research	0		Not original research	0
	Protocol only	0		Protocol only	0
	Participants outside criteria	1		Participants outside criteria	0
EQ-5D	Psychometric properties not evaluated	3	FAT	Psychometric properties not evaluated	1
	Conference proceedings	0		Conference proceedings	0
	Not conducted in English	8		Not conducted in English	0
	Modified or incorrect measure	3		Modified or incorrect measure	1
	Not original research	2		Not original research	0
	Protocol only	0		Protocol only	0
	Participants outside criteria	4		Participants outside criteria	0
mFAT	Psychometric properties not evaluated	6	FIM	Psychometric properties not evaluated	44
	Conference proceedings	4		Conference proceedings	1

Outcome Measure	Exclusion reason	n =	Outcome Measure	Exclusion reason	n =
GAS	Not conducted in English	3	GAS-10 point	Not conducted in English	25
	Modified or incorrect measure	4		Modified or incorrect measure	19
	Not original research	1		Not original research	0
	Protocol only	0		Protocol only	0
	Participants outside criteria	0		Participants outside criteria	36
	Psychometric properties not evaluated	20		Psychometric properties not evaluated	0
	Conference proceedings	2		Conference proceedings	0
	Not conducted in English	2		Not conducted in English	0
	Modified or incorrect measure	2		Modified or incorrect measure	0
	Not original research	3		Not original research	0
	Protocol only	2		Protocol only	0
	Participants outside criteria	9		Participants outside criteria	0
Global Assessment Scale	Psychometric properties not evaluated	1	Klein Bell ADL Scale	Psychometric properties not evaluated	3
	Conference proceedings	0		Conference proceedings	0
	Not conducted in English	0		Not conducted in English	3
	Modified or incorrect measure	5		Modified or incorrect measure	2
	Not original research	0		Not original research	1
	Protocol only	0		Protocol only	0
	Participants outside criteria	0		Participants outside criteria	6
LASIS	Psychometric properties not evaluated	2	SF-36	Psychometric properties not evaluated	21
	Conference proceedings	0		Conference proceedings	0
	Not conducted in English	0		Not conducted in English	2
	Modified or incorrect measure	0		Modified or incorrect measure	5
	Not original research	0		Not original research	0
	Protocol only	0		Protocol only	0
	Participants outside criteria	0		Participants outside criteria	1
MAL	Psychometric properties not evaluated	10	MAL-5	Psychometric properties not evaluated	0
	Conference proceedings	2		Conference proceedings	0
	Not conducted in English	6		Not conducted in English	0
	Modified or incorrect measure	13		Modified or incorrect measure	0
	Not original research	1		Not original research	0
	Protocol only	0		Protocol only	0

Outcome Measure	Exclusion reason	n =	Outcome Measure	Exclusion reason	n =
	Participants outside criteria	0		Participants outside criteria	0
	Psychometric properties not evaluated	0		Psychometric properties not evaluated	19
	Conference proceedings	0		Conference proceedings	2
	Not conducted in English	1		Not conducted in English	3
MAL-28	Modified or incorrect measure	1	MI	Modified or incorrect measure	5
	Not original research	0		Not original research	0
	Protocol only	0		Protocol only	0
	Participants outside criteria	0		Participants outside criteria	1
	Psychometric properties not evaluated	51		Psychometric properties not evaluated	22
	Conference proceedings	13		Conference proceedings	1
	Not conducted in English	9		Not conducted in English	0
NHPT	Modified or incorrect measure	0	OHS	Modified or incorrect measure	1
	Not original research	2		Not original research	0
	Protocol only	1		Protocol only	0
	Participants outside criteria	4		Participants outside criteria	0
	Psychometric properties not evaluated	0		Psychometric properties not evaluated	9
	Conference proceedings	0		Conference proceedings	4
	Not conducted in English	0		Not conducted in English	4
PDC/CBS	Modified or incorrect measure	0	RMA	Modified or incorrect measure	9
	Not original research	0		Not original research	0
	Protocol only	0		Protocol only	0
	Participants outside criteria	0		Participants outside criteria	0
	Psychometric properties not evaluated	9		Psychometric properties not evaluated	0
	Conference proceedings	4		Conference proceedings	0
	Not conducted in English	4		Not conducted in English	2
RMA-UL	Modified or incorrect measure	8	SA-SIP	Modified or incorrect measure	0
	Not original research	0		Not original research	1
	Protocol only	0		Protocol only	0
	Participants outside criteria	0		Participants outside criteria	0
SIS	Psychometric properties not evaluated	17	10MWT	Psychometric properties not evaluated	13
	Conference proceedings	9		Conference proceedings	1

Outcome Measure	Exclusion reason	n =	Outcome Measure	Exclusion reason	n =
	Not conducted in English	4		Not conducted in English	3
	Modified or incorrect measure	25		Modified or incorrect measure	0
	Not original research	1		Not original research	0
	Protocol only	0		Protocol only	0
	Participants outside criteria	0		Participants outside criteria	2
	Psychometric properties not evaluated	7			
	Conference proceedings	0			
	Not conducted in English	4			
UL-MAS	Modified or incorrect measure	1			
	Not original research	0			
	Protocol only	0			
	Participants outside criteria	0			

ARAT = Action Research Arm Test, ArmA = Arm Activity Measure, AQoL = Assessment of Quality of Life, BI = Barthel Index, BI (C&W) = Barthel Index - Collin & Wade version, CMSA = Chedoke-McMaster Stroke Assessment, DAS = Disability Assessment Scale, EQ-5D = EuroQol – 5 dimension, FAT = Frenchay Arm Test, mFAT = modified Frenchay Arm Test, FIM = Functional Independence Measure, GAS = Goal Attainment Scale, GAS – 10pt = Goal Attainment Scale – 10 point, Global Ax = Global Assessment Scale, KleinBell ADL = Klein-Bell Activities of Daily Living scale, LASIS = Leeds Adult Spasticity Impact Scale, SF-36 = Medical Outcome Study 36-Item Short-Form Health Survey, MAL = Motor Activity Log, MAL-5 = Motor Activity Log - 5, MAL-28 = Motor Activity Log - 28, MI = Motricity Index, NHPT = Nine Hole Peg Test, OHS = Oxford Handicap Scale, PDS/CBS = Patient Disability Scale / Carer Burden Scale, RMA = Rivermead Motor Assessment, RMA-UL = Rivermead Motor Assessment - Upper Limb, SA-SIP = Stroke-Adapted Version of the Sickness Impact Profile, SIS = Stroke Impact Scale, 10MWT = Ten Metre Walk Test, UL MAS = Upper Limb Motor Assessment Scale.

Appendix Table 5. *Methodological quality and quality criteria ratings*

Action Research Arm Test	Content validity	Structural validity	Internal consistency	Reliability		Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Beebe & Lang (2009)							*Very good - (3/5)*	*Doubtful + (2/2)*
Burridge et al., (2009)							*Adequate - (4/8)*	
De Weerdt & Harrison (1985)							Very good + (1/1)	Inadequate + (1/1)
Dromerick et al., (2006)							Very good - (2/3)	
Fleming et al,. (2014)							*Adequate + (1/1)*	
Lang et al., (2008)*								
Lang et al., (2006)							*Very good - (5/7)*	*Doubtful + (2/2)*
Lyle (1981)	Inadequate	Inadequate +						
Morris et al., (2013)							Doubtful - (1/2)	
Rabadi & Rabadi (2006)							Very good – (1/2)	Adequate + (1/1)
Rand & Eng (2015)							Very good + (1/1)	
Yozbatiran et al., (2008)				Adequate +	Adequate +			

Arm Activity Measure	Content validity	Structural validity	Internal consistency	Reliability		Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Ashford et al., (2015)	*Very good +/-*							
Ashford et al., (2016)		*Adequate +*						
Ashford et al., (2014)								*Inadequate + (2/2)*
Ashford, Slade et al., (2013)	*Adequate +*							
Ashford, Turner-Stokes, et al., (2013)		*Adequate +*	*Very good +*			*Adequate +*	*Inadequate + (2/2)*	*Adequate + (2/2)*

Assessment of Quality of Life	Content validity	Structural validity	Internal consistency	Reliability		Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest		

Hawthorne et al., (2009)								Very good + (1/1)	
Hawthorne et al., (1999)	Inadequate								
Sturm et al., (2002)**								Adequate + (1/1)	
Barthel Index	**Content validity**	**Structural validity**	**Internal consistency**	**Reliability**			**Measurement error**	**Construct validity**	**Responsiveness**
				Inter	**Intra**	**Retest**			
Ali et al., (2013)								Adequate – (0/1)	
Filiatrault et al., (1991)								Adequate + (2/2)	Inadequate ?
Kwon et al., (2004)								Very good + (2/2)	
Mahoney & Barthel (1965)	Inadequate								
Wallace et al., (2002)									Doubtful - (0/1)
Wellwood et al., (1995)								Doubtful + (1/1)	
Barthel Index (Collin & Wade)	**Content validity**	**Structural validity**	**Internal consistency**	**Reliability**			**Measurement error**	**Construct validity**	**Responsiveness**
				Inter	**Intra**	**Retest**			
Barer & Murphy (1993)		Inadequate ?						Inadequate + (2/2)	
Collin et al., (1988)	Inadequate			Inadequate ?					
Dennis et al., (2000)								Inadequate + (4/4)	
Green et al., (2001)						Inadequate ?	Adequate +		
Houlden et al., (2006)									Doubtful + (1/1)
Sarker et al., (2012)								Doubtful + (2/2)	
van der Putten et al., (1999)									Doubtful – (1/2)
Wade & Hewer (1987)		Adequate +						Doubtful + (1/1)	
Wilkinson et al., (1997)								Doubtful + (8/8)	
Chedoke-McMaster Stroke Assessment	**Content validity**	**Structural validity**	**Internal consistency**		**Reliability**		**Measurement error**	**Construct validity**	**Responsiveness**

				Inter	Intra	Retest			
Dang et al., (2011)								Doubtful - (1/2)	
Gowland (1990)	Inadequate								
Gowland et al., (1993)				II very good + AI adequate +	Very good +	Adequate -		Adequate + (4/4)	Adequate + (1/1)
Moreland et al., (1993)	Inadequate								

Disability Assessment Scale	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Brashear et al., (2002)	*Inadequate ?*			*Adequate -*	*Adequate ?*				
Doan et al., (2012)								*Adequate + (2/2)*	

EuroQol – 5 Dimension	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Alderman et al., (2001)								Adequate - (0/4)	
Barton, Sach, Doherty, et al., (2008)								Very good + (1/1)	
Barton, Sach, Avery, et al., (2008)								Adequate + (1/1)	
Doan et al., (2012)								*Adequate + (2/2)*	
Dorman et al., (1999)								Adequate + (1/1)	
Dorman et al., (1998)						Doubtful + patient - proxy			
Dorman et al., (1997)								Doubtful + (6/7)	
Fisk et al., (2005)						Adequate +		Adequate + (6/6)	
Gillard et al., (2015)								*Adequate + (1/1)*	
Kohn et al., (2014)								Adequate - (1/2)	
Kuspinar & Mayo (2013)	Doubtful +/-							Adequate + (2/2)	
Kuspinar et al., (2014)								Very good	

Moore et al., (2004)								+ (1/1) Inadequate - (1/3)	
Nicholl et al., (2001)								Doubtful - (0/2)	
Peters et al., (2014)									Inadequate – (0/2)
Pickard et al., (2005)									Adequate + (11/13)
Salter et al., (2008)	Doubtful ?								
Williams (1990)	Inadequate -								
Xie et al., (2006)								Very good + (1/1)	

modified Frenchay Arm Test	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Heller et al., (1987)				Doubtful ?		Doubtful ?			
Heller et al., (1987)								Doubtful - (0/1)	

Functional Independence Measure	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Brown et al., (2015)								Very good + (3/3)	
Corrigan et al., (1997)								Doubtful + (9/9)	
Cullen et al., (2014)								Very good + (2/2)	
Cuthbert et al., (2015)								Adequate – (0/1)	
Egan et al., (2014)								Doubtful + (1/1)	
Grant et al., (2014)								Adequate – (0/1)	
Hall et al., (1993)								Inadequate + (6/6)	
Hamilton & Granger (1994)				Doubtful +					
Heinemann et al., (1993)		Adequate +							
Heinemann et al., (1994)		Adequate +							

Heinemann et al., (1997)								Inadequate + (1/1)	
Houlden et al., (2006)									Doubtful - (2/3)
Keith et al., (1987)	Inadequate								
Kuys et al., (2009)								Adequate - (0/1)	
Oczkowski & Barreca (1993)								Adequate + (1/1)	
Ouellette et al., (2015)								Doubtful – (0/1)	
Rabadi & Vincent (2013)								*Adequate + (1/1)*	*Doubtful + (1/1)*
Sharrack et al., (1999)		Inadequate +	Adequate +	Adequate +	Adequate +			Adequate - (0/1)	Doubtful + (1/1)
Stineman et al., (1996)		Adequate +	Adequate +						
van der Putten et al., (1999)									Doubtful - (1/2)

Goal Attainment Scale	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Bovend'Eerdt et al., (2011)				Adequate -			Adequate ?		
Brock et al., (2009)								Inadequate + (4/5)	
Doig et al., (2010)								Adequate + (1/1)	Inadequate + (1/1)
Joyce et al., (1994)				Doubtful ?				Doubtful - (4/7)	
Khan et al., (2008)								Very good - (1/3)	Doubtful + (2/2)
Lannin (2003)									Doubtful + (1/1)
Malec (1999)								Very good + (1/1)	
Malec et al., (1991)								Very good + (4/4)	
Turner-Stokes et al., (2010)								*Doubtful - (3/7)*	

Motor Activity Log	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Chen et al., (2012)							Adequate ?		

Dromerick et al., (2006)								Very good - (1/2)	
Harris & Eng (2007)								*Very good - (3/7)*	
Uswatte et al., (2006)		Inadequate ?							
Uswatte & Taub (2005)	Inadequate								

Motor Activity Log-28	**Content validity**	**Structural validity**	**Internal consistency**	**Reliability**			**Measurement error**	**Construct validity**	**Responsiveness**
				Inter	**Intra**	**Retest**			
Uswatte et al., (2006)	Inadequate ?	Inadequate ?	Very good +			Adequate Patient + Proxy -		Inadequate Patient + (3/4) Proxy – (2/4)	

Motricity Index	**Content validity**	**Structural validity**	**Internal consistency**	**Reliability**			**Measurement error**	**Construct validity**	**Responsiveness**
				Inter	**Intra**	**Retest**			
Bohannon (1999)			Doubtful +					Inadequate - (0/1)	
Collin & Wade (1990)				Doubtful ?				Adequate + (1/1)	
Demeurisse et al., (1980)^	Inadequate								
Jacob-Lloyd et al., (2005)								Very good – (1/2)	Doubtful - (0/1)
Stone et al., (1993)								Doubtful + (1/1)	
Wade & Hewer (1987)								Doubtful + (1/1)	

Nine Hole Peg Test	**Content validity**	**Structural validity**	**Internal consistency**	**Reliability**			**Measurement error**	**Construct validity**	**Responsiveness**
				Inter	**Intra**	**Retest**			
Beebe & Lang (2009)								*Very good – (3/5)*	*Doubtful + (2/2)*
Benedict et al., (2011)								Adequate + (6/8)	
Costelloe et al., (2008)								Adequate - (2/3)	
Goodkin et al., (1988)								Adequate + (2/2)	
Heller et al., (1987)				Doubtful ?		Doubtful ?			
Heller et al., (1987)								Doubtful + (1/1)	

Jacob-Lloyd et al., (2005)								Very good - (1/2)	Doubtful + (1/1)
Morris et al., (2013)								Doubtful - (0/1)	
Poole et al., (2010)								Adequate - (6/10)	
Schwid et al., (2002)							Adequate +		
Oxford Handicap Scale	**Content validity**	**Structural validity**	**Internal consistency**	**Reliability**			**Measurement error**	**Construct validity**	**Responsiveness**
				Inter	**Intra**	**Retest**			
Rigby et al., (2009)								Doubtful - (0/1)	
Simon et al., (2008)								Adequate + (2/2)	
Rivermead Motor Assessment	**Content validity**	**Structural validity**	**Internal consistency**	**Reliability**			**Measurement error**	**Construct validity**	**Responsiveness**
				Inter	**Intra**	**Retest**			
Adams, Pickering, & Taylor (1997)		Inadequate -							
Adams, Pickering, Ashburn, et al., (1997)		Inadequate -							
Jones (1998)								Very good +(3/3)	
Lincoln & Leadbitter (1979)	Inadequate								
Sackley (1990)								Adequate + (2/2)	
Rivermead Motor Assessment – Upper Limb	**Content validity**	**Structural validity**	**Internal consistency**	**Reliability**			**Measurement error**	**Construct validity**	**Responsiveness**
				Inter	**Intra**	**Retest**			
Adams, Pickering, & Taylor (1997)		Inadequate +							
Adams, Pickering, Ashburn, et al., (1997)		Inadequate -							
Collin & Wade (1990)								Adequate + (1/1)	
Lincoln & Leadbitter (1979)	Inadequate								
Morris et al., (2013b)								Doubtful - (0/1)	
Sackley (1990)								Adequate + (2/2)	

SA-SIP30	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Doan et al., (2012)								*Adequate + (1/1)*	
Edwards et al., (2006)								Adequate + (2/2)	
Salter et al., (2008)	Doubtful ?								
van Straten et al., (1997)^	Inadequate								

SF-36	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Anderson et al., (1996)			Doubtful +					Adequate - (1/2)	
Dorman et al., (1999)								Adequate + (1/1)	
Dorman et al., (1998)			Very good +			Patient adequate + Proxy Adequate -			
Duncan et al., (1997)								Doubtful + (1/1)	
Findler et al., (2001)								Adequate + (2/2)	
Freeman et al., (2000)			Very good +					Adequate + (7/7)	Doubtful - (0/1)
Freeman et al., (1996)								Very good - (2/3)	
Guilfoyle et al., (2010)		Very good -	Very good +					Doubtful + (1/1)	
Hagen et al., (2003)			Very good +					Inadequate - (1/3)	Doubtful - (0/2)
Herrmann et al., (1996)								Very good - (1/2)	
Hobart et al., (2002)		Adequate ?	Very good +						
Mackenzie et al., (2002)		Very good +						Adequate - (0/1)	
Madden et al., (2006)								Very good - (0/1)	Inadequate - (0/1)
Moore et al., (2004)								Inadequate	

Murrell et al., (1999)						Doubtful ?		- (1/6)	
O'Mahony et al., (1998)**									
Pittock et al., (2004)								Very good - (1/2)	
Riazi et al., (2003)								Very good + (1/1)	
Robinson Jr et al., (2009)								Adequate - (0/1)	
Salter et al., (2008)	Doubtful ?								
Vickrey et al., (1995)								Adequate + (1/1)	
Vickrey et al., (1997)			Very good +			Adequate +		Very good - (4/8)	
Ware & Sherbourne (1992)	Inadequate								
Williams et al., (1999)								Inadequate - (0/1)	

Stroke Impact Scale	**Content validity**	**Structural validity**	**Internal consistency**	**Reliability**			**Measurement error**	**Construct validity**	**Responsiveness**
				Inter	**Intra**	**Retest**			
Duncan et al., (2003)	Inadequate	Very good +							
Duncan et al., (2002)				Adequate ?				Adequate + (8/8)	
Duncan et al., (2005)			Very good +			Doubtful +			
Duncan et al., (1999)^^	Inadequate								
Eriksson et al., (2013)								Very good + (4/5)	
Jenkinson et al., (2013)		Very good +	Very good +						
Kwon et al., (2006)								Adequate + (4/4)	
Lai et al., (2002)								Doubtful + (1/1)	
Salter et al., (2008)	Doubtful ?								
Wolf & Koster (2013)								Doubtful + (1/1)	

Ten Metre Walk Test				**Reliability**					**Responsiveness**

	Content validity	Structural validity	Internal consistency	Inter	Intra	Retest	Measurement error	Construct validity	
Donovan et al., (2008)								Adequate - (0/2)	
Hirsch et al., (2014)						Adequate +		Very good + (1/1)	
Kuys et al., (2009)								Adequate + (1/1)	
Miller et al., (2013)								Adequate + (2/2)	
Mudge & Stott (2009)								Doubtful - (0/1)	
Salbach et al., (2001)									Doubtful – (2/5)
Schmid et al., (2012)								Adequate - (0/1)	
Scrivener et al., (2014)									Doubtful + (1/1)

Upper Limb – Motor Assessment Scale	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness
				Inter	Intra	Retest			
Carr et al., (1985)	Inadequate			Inadequate ?					
Johnson & Selfe (2004)			Very good +						
Khan et al., (2013)		Adequate +						Very good + (1/1)	
Kuys et al., (2009)								Adequate - (0/1)	
Lannin (2004)		Very good ?	Very good +						
Loewen & Anderson (1988)				Doubtful ?	Doubtful ?				
Loewen & Anderson (1990)								Adequate - (2/6)	
Miller et al., (2010)		Very good +	Very good +						
Pickering et al., (2010)		Inadequate -							
Sabari et al., (2005)		Doubtful -							

Terwee quality criteria: '+' = sufficient, '?' = indeterminant, '-' = insufficient. 'Very good', 'adequate', 'doubtful', 'inadequate' as per (Prinsen et al., 2018; C.B. Terwee et al., 2007)

Italicised data indicates study sample included participants with identified upper limb spasticity.

*Internal consistency evidence strength cannot exceed structural validity as per COSMIN guidelines and has been reduced accordingly.

**Interpretability data available not included in table.

^Included for content validity (measure development) only, excluded for other psychometric property analysis due to failure to meet eligibility criteria regard language study conducted in.

^^Included for content validity (measure development) only excluded for other psychometric property analysis due to failure to meet modified or incorrect measure (incorrect version).

AI = Chedoke-McMaster Stroke Assessment Activity Inventory, II = Chedoke-McMaster Stroke Assessment Impairment Inventory, SA-SIP30 = Stroke-Adapted Version of the Sickness Impact Profile, SF-36 = Medical Outcome Study 36-Item Short-Form Health Survey.

Appendix Table 6. *Summary of results for included studies*

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
ARAT (Obs perf measure of upper limb activity)	Modification of the UEFT	Guttman Scaling criteria met	-	ICC 0.99	ICC 0.99	-	-	Measures of; • Activity and UL dexterity r = -0.25-0.95 • Global measures of function r = 0.2 – 0.6 • Impairment r = 0.03–0.86 Predicting UL use R^2 = 0.6 (P<0.001)–0.776, (P<0.01) Not predictive HRQoL	ES = 0.52–1.018 RR= 5.20–7.067 SRM = 0.68 R^2 = .56, P<.001 T=1, n=28, z=4.60	MCID 12, 17 Floor effect 0% Ceiling effect 0-41%	(Beebe & Lang, 2009; Burridge et al., 2009; De Weerdt & Harrison, 1985; Dromerick et al., 2006; Fleming et al., 2014; Lang et al., 2006; Lyle, 1981; Morris et al., 2013; Rabadi & Rabadi, 2006; Rand & Eng, 2015; Yozbatiran et al., 2008)
ArmA (Self-report measure of upper limb activity –	Tool development methodology supports content validity.	Unidimensional passive and active scales.	Passive subscale α = 0.85, active subscale α = 0.96			Weighted kappa passive subscale 0.90 (CI 0.68 – 1.12), active		Passive subscale with; • passive items on LASIS Rho 0.50; p = 0.01,	Significant difference between responder and non-responder groups for passive	Ceiling effect (37%) active function subscale.	(Ashford et al., 2015; Ashford et al., 2016; Ashford et al., 2014; Ashford, Slade, et

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
active and passive)	Items derived from patient goals (categorised into passive function, active function, symptoms, cosmesis, impairment) informed standardised measure development .	Passive subscale conforms to Rasch model. Able to differentiate at least two groups of patients.				subscale 0.93 (CI 0.71 – 1.15)		• active items LASIS Rho 0.02; p = 0.9 Active function subscale with; • active LASIS Rho 0.48; p = 0.01 • DASH active items Rho 0.63 p = 0.01 • passive LASIS items Rho 0.23 p = 0.078	function subscale (8 wks) U = 98.5; p = 0.01), not shown for the active function subscale (U=163.4; p=0.35) as expected. More responsive than LASIS, DASH active items, Barthel Index . Detected change post focal spasticity intervention	No floor effect on either subscale.	al., 2013; Ashford & Turner-Stokes, 2013)
AQoL (Self-report measure	-	-	-	-	-	-	-	Measures of; • Handicap LHS r = 0.83 • Disability BI r = 0.77 • Impairment r = -0.69	-	No floor or ceiling effect (1-2%)	(Hawthorne et al., 2009; Hawthorne et al., 1999;

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
of global HRQoL)								• Mood r = -0.63- -0.60 • Global measure of function SF-36 r = 0.15–0.83 Discriminates between stroke types more efficiently than SF-36 Sensitive in TBI ES -0.80 Predictor of death and/or institutionalization			Sturm et al., 2002)
BI (Self-report or obs perf measure of global ADL)	-	-	-	-	-	-	-	Greater associated with proxy ratings of QoL than patient (r = 0.68) • m-FIM r_s=0.9479 p<0.0001 • MRS r_s=-0.8856 p<0.0001. • Fugl-Meyer r rho = 0.60 • Functional test for the hemiparetic UL Rho = .61	Kendall's statistic W = 0.39 t =0.63	Ceiling effect	(Ali et al., 2013; Filiatrault et al., 1991; Kwon et al., 2004; Wallace et al., 2002; Wellwood et al., 1995)

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
								• disability r=-0.73 p<0.001.			
BI (C&W) (Self-report or obs perf measure of global ADL)	-	One major factor. Guttman scaling criteria not met.	-	Disagreement of 1 - 9 points between self-report, nurse, 2 skilled observers. Agreement lower for transfers, feeding, dressing, grooming, toileting.	-	>75% agreement, kappa - 0.99 to 0.81.	0.4 (95%CI 0.01-0.90) Difference of 4/20 points likely to represent genuine change	Predicts DC destination and LOS ≥ 0.70, Stronger relationships with measures of disability and physical abilities Measures of; • Activity = 0.73-0.83 • Disability r_s = 0.73-0.8 all P<0.01. • Global measures of function r_s = 0.33 -0.81, • Mental health r_s = -0.2 - -0.4	TBI FIM change scores r^2 = 0.733 MS ES = 0.37 Stroke ES = 0.52-0.95	Floor and ceiling effects	(Barer & Murphy, 1993; Collin et al., 1988; Dennis et al., 2000; Green et al., 2001; Houlden et al., 2006; Sarker et al., 2012; van der Putten et al., 1999; Wade & Hewer, 1987; Wilkinson et al., 1997)
CMSA (Obs perf measure of impairment and mobility)	-	-	-	ICC = 0.88 (95% CI 0.76–0.94) to 0.99 (95% CI 0.98–1.00)	ICC 0.93 (95% CI 0.85–0.96) to 0.98 (95% CI 0.95–0.99)	Disability Total 0.98 (95% CI 0.95–0.99)	-	• Fugl-Meyer Test r = 0.95 p<0.001 • FIM r = 0.79, p<0.05 (disability inventory) • Gowlands predictive equations; not supported	CMSA disability inventory (0.53 F 37.25 p<0.001) more responsive than FIM (0.39 F	-	(Dang et al., 2011; Gowland et al., 1993)

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
									19.40 p<0.001).		
DAS (Self-report or obs perf measure of upper limb activity – active and passive)	-	-	-	Kendall W Hygiene 0.626 (95% CI 0.297–1.00 p<0.001), dressing 0.494 (95% CI 0.234–1.00 p<0.001) limb position 0.557 (95% CI 0.264–1.00 p<0.001) pain 0.772 (95% CI 0.366 – 1.00 p<0.001)	Weighted Kappa hygiene 0.520 (95% CI 0.239 – 0.802), P .994, Dressing 0.530 (95% CI 0.278 – 0.782) P 0.478, Limb position 0.775 (95% CI 0.560-0.991) P 0.998, Pain 0.776 (95% CI 0.533–1.0) P 0.992.	-	-	Increasing disability in DAS was associated with diminishing EQ-5D index scores (P < .002) Increasing disability is associated with reduction in HRQoL and caregiver burden (P<.05)	-	-	(Brashear, Zafonte, et al., 2002; Doan et al., 2012)
EQ-5D (Self-report measure	Includes • 4/10 domains most important to people	-	-	-	-	ICC = 0.63–0.80 (individual domains)	-	No significant differences between staff and participant ratings Measures of	EQ-5D mean change -0.01 (95%CI -0.03 to 0.01),	Index floor 0%, ceiling 13.3%	(Alderman et al., 2001; Barton, Sach, Avery, et

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
of global HRQoL)	with MS and 2/7 domains not prioritized by people with MS • 6/9 items related to domains considered important for people post-stroke.					ICC = 0.81-0.86 (overall) Self-report more reliable than proxy		• global function r = 0.33 - 0.77 • UL activity/dexterity r = 0.56 • disability (P<.002) • self-care r = 0.64 • activities r = 0.60 • carer burden (P<.05) • cognition/psychological r = 0-0.56 • Mobility r = -0.69-61 • pain r = 0.71 • general health status r = 0.80 • Less sensitive than FAMS and SF-54 EQ5D index stronger association with measures of disability and ADL VAS stronger correlations with changes in mental functioning Discriminant validity; • baseline stroke severity and type	VAS mean change -1.88 (95%CI -5.12 to 1.37) No significant change over 12 mo. EQ5D greatest change scores amongst generic measures.	VAS floor 0% % ceiling 3%. Individual subscales: floor effect 2-34%, ceiling effect 7-68%	al., 2008; Barton, Sach, Doherty, et al., 2008; Doan et al., 2012; Dorman et al., 1999; Dorman et al., 1998; Dorman et al., 1997; Fisk et al., 2005; Gillard et al., 2015; Kohn et al., 2014; Kuspinar et al., 2014; Kuspinar & Mayo, 2013; Moore et al., 2004; Nicholl et al., 2001; Peters et al., 2014; Pickard et al., 2005; Salter et al., 2008; Xie et al., 2006)

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
								• spasticity. Mean difference 0.07 (CI -0.12 to -0.03) • stroke (average 6.9% lower index, 7.2% lower VAS) • functional walking capacity and general health perception			
FAT (Obs perf measure of upper limb activity)	-	-	-	-	-	-	-	-	-	-	-
mFAT (Obs perf measure of upper limb activity)	-		-	Rho = 0.83–0.99	Rho = 0.68–0.90	Rho = 0.75–0.99	-	Less sensitive than NHPT 52% within normal NHPT scores.	-	Floor effect 30% Ceiling effect 34%	(Heller et al., 1987)
FIM (Obs perf measure of global ADL)	-	Two factor (motor and cognitive) structure supported – (89.4% of total variance	FIM total α = 0.94 – 0.98, FIM motor α = 0.93 – 0.97, FIM cognitive α = 0.93 – 0.94	FIM total ICC = 0.99, FIM motor ICC = 0.96 FIM cognitive ICC = 0.91	FIM total ICC = 0.94	-	-	Predictive of; • LOS • Discharge destination (X_2 = 69.4, P<0.001, AUC = 0.76, sensitivity = 0.76, specificity = 0.64).,	FIM total ES= 0.46–0.72 FIM Cognitive ES = 0.35-0.43	Floor and ceiling effect	(Brown et al., 2015; Corrigan et al., 1997; Cullen et al., 2014; Cuthbert et al., 2015; Egan et al., 2014; Grant et

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
		explained). 3 or 4 factor structure proposed (.>50% of variance explained)						• Minutes and type of assistance and supervision(p<0.0001), (p<0.0032), (p<0.0063) • Predicting RTD sensitivity 72%, specificity 73%, ROC 0.71 • More superior prediction than SF-36 Measures of • Activity r = 0.21 P<.01 to 0.54 P<.001 • Disability r = 0.64-0.96 • Work r=-0.59, p<0.001–0.64 p=0.001 • Independence r=-0.44 p=0.001 • Global measures of function r = 0.63–0.92 • Robotic measures = -0.21- -0.79 • QoL r = 0.41-0.86 • Depression r = 0.27- -.43 (p = .001)	SRM 0.53 More responsive than BI (ES = 0.4 SRM 1.0, t 1.1z 1.1; variance ratio 0.39 F 19.40 p<0.001). Less responsive than • GAS • CMSA disability inventory		al., 2014; Hall et al., 1993; Hamilton & Granger, 1994; Heineman n et al., 1997; Heineman n et al., 1993, 1994; Houlden et al., 2006; Kuys et al., 2009; Oczkowski & Barreca, 1993; Ouellette et al., 2015; Rabadi & Vincent, 2013; Sharrack et al., 1999; van der Putten et al., 1999)

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
								Low correlations with; • DC walking speed r = 0.25 (<0.001). • PTA length, cognition, consciousness • employment at 2 years • emotional well being • FIMm with nursing contact time			
Global Ax Scale (Self-report measure of response to treatment)	-	-	-	-	-	-	-	-	-	-	-
GAS (Self-report or obs perf measure of individual goal	Majority of goal areas matched necessary domains.	-	-	ICCA,k 0.478 r = 0.92 - 0.94	-	-	LOA: -1.52 +/- 24.54	2 mo GAS predicted final GAS 0.66 70% agreement between participant	ES 0.90 SRM: 2.4 t value 10.0 z value 1.4	-	(Bovend'Eerdt et al., 2011; Brock et al., 2009; Doig et al., 2010; Joyce et al., 1994;

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
attainment)								and significant other ratings. Greater goal achievement was associated with; • higher mobility (FIM motor r = 0.55) • less depression (r = 0.46) • better self-efficacy r = 0.46) 6 months post. Measure of; • Activity (observed) r = -0.0039 to 0.77 • perceived activity and participation r = 0.45 – -0.51 p<0.005 • global clinical impression (clinical judgement of efficacy) r = 0.81, • vocational independence and outcome r = -0.34 - -0.69) • MAS 0.35, GAS change score and	Pre T-score 36.9 (6.3) post Ix 52.8 (6.2) t = -9.65 p<0.01 More responsive than FIM, BI		Khan, Pallant, et al., 2008; Lannin, 2003; Malec, 1999; Malec et al., 1991; Turner-Stokes et al., 2010)

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
								reduction in spasticity rho 0.28 p 0.04. No relationship with; • LOS -0.13			
Klein-Bell ADL Scale (Self-report measure of global ADL)	-	-	-	-	-	-	-	-	-	-	-
LASIS (Self-report measure of upper limb activity – active and passive)	-	-	-	-	-	-	-	-	-	-	-

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
MAL (Self-report measure of upper limb activity)	-	Items removed due to missing data; write on paper (48%), put on makeup / shaving cream (20%). 92% item-total correlations >0.5.	-	-	-	-	MDC AOU 16.8%, QOM 15.3%	Measures of; • Activity r = 0.61-0.82 (p<0.01) • Participation r = 0.23 (p<0.05) • Impairment r = -0.06 - 0.84 (p<0.01)	-	Nil floor/ceiling effect	(Chen et al., 2012; Dromerick, et al., 2006; Harris & Eng, 2007; Uswatte et al., 2006)
MAL-5 (Self-report measure of upper limb activity)	-	-	-	-	-	-	-	-	-	-	-

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
MAL-28 (Self-report measure of upper limb activity)	-	-	α = 0.94, caregiver 0.95	-	-	Patient reported QOM/AOU ICC 0.82/0.79, caregiver reported QOM/AOU ICC 0.72/0.66.	-	Patient and participant reported QOM/AOU with; • accelerometry r = 0.52/0.47, p<0.01, r = 0.61/0.57, p<0.01, • SIS hand function r = 0.72/0.68, p<0.01, 0.40/0.35, <0.01	-	-	(Uswatte et al., 2006)
MI (Obs perf measure of impairment and mobility)	-	-	α = 0.968	MI arm spearman rho = 0.88	-	-	-	Measures of; • Activity r = 0.73–0.76, • Dexterity r = 0.36–0.53 • Global function r = 0.61–0.77 • Impairment r = 0.74–0.94, Predictor of independence	ES 0.49	Ceiling effect 18% No floor effect	(Bohannon, 1999; Collin & Wade, 1990; Jacob-Lloyd et al., 2005; Stone et al., 1993; Wade & Hewer, 1987)

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
NHPT (Obs perf measure of upper limb activity)	-		-	Rho = .83–0.99	Rho = 0.68–0.90	Rho = 0.75–0.99	20% score change indicates true change	Measures of; • Activity r = 0.36–0.76 • Grip and dexterity r_s =0.61–0.95 • Self-reported hand use r_s = 0.53–0.66 • Global measures of function r = -0.19–0.61 • Disability r = 0.63 • Cognition r = -0.20 - -0.65 p<0.01 • HRQoL r = 0.08 • Age and impairments r = -0.05-0.48 9HPT more sensitive than Frenchay Arm Test, Did not predict overall HRQoL r = -0.08	ES = 0.52–0.66	Floor effect = 75% (stroke) Nil floor or ceiling (MS)	(Beebe & Lang, 2009; Benedict et al., 2011; Goodkin et al., 1988; Heller et al., 1987; Jacob-Lloyd et al., 2005; Morris et al., 2013; Poole et al., 2010; Schwid et al., 2002)

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
OHS (Self-report measure of global ADL)	-	-	-	-	-	-	-	Not predictive of caregiver burden (Relatives Stress Scale P<0.0001; Bakas Caregiver Outcomes Scale P 0.059). Predicted number of services required (13% of variance F=9.53 d.f.1=1, d.f.2=62 P=0.001, 14%) and amount of time provided (26.5% variance F=26.39, d.f.1=1, d.f.2=64 P<0.001 and 28.2%).	-	-	(Rigby et al., 2009; Simon et al., 2008)
PDS/CBS (Self-report measure of upper limb activity – active and passive)	-	-	-	-	-	-	-	-	-	-	-

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
SIS (Self-report measure of global HRQoL)	-	V3 developed following deletion of 5 items secondary to misfit to the construct. Unidimensional domains, ranging in item difficulty and able to discriminate. 8 domains can form single index (α = 0.93, accounting for 68.76% of variance)	α = 0.86 – 0.89 excluding emotion domain, 0.76 – 0.83	Self and proxy agreement ICC = 0.50-0.83.	-	ICC = 0.62-0.94 (Mail) ICC = 0.91-0.98 (telephone)	-	Measures of; • Activity (patient reported) r = 0.42-0.77, (proxy reported) r = 0.37–0.78 • Global measures of function r = 0.40-0.98 Perceived recovery predicted perceived participation. Discriminant validity; • Stroke (mean 9 points lower) • Disability (Kruskal-Wallis test 0.0002 – 0.0694) Telephone administration; • FIM and SF-36 r = 0.362-0.858	-	No evidence of floor or ceiling effect (complete SIS) Ceiling effect 32.2% (participation domain)	(Duncan et al., 2003; Duncan et al., 2002; Duncan et al., 2005; Eriksson et al., 2013; Jenkinson et al., 2013; Kwon et al., 2006; Lai et al., 2002; Wolf & Koster, 2013)

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
SA-SIP30 (Self-report measure of global HRQoL)	-	-	-	-	-	-	-	Greater disability (DAS) associated with higher SA-SIP30 (P<.05) 6 mo SA-SIP30 was not predicted by admission NIHSS r = 0.11 Accounted for 53% of variance in predicting Reintegration to Normal Living R2 =0.63, P<0.0001).	-	-	(Doan et al., 2012; Edwards et al., 2006)

SF-36 (Self-report measure of global HRQoL)	-	Inconsistent support for two factor structure of scale in TBI Scaling assumptions were not fully satisfied for scale or summary scores in stroke.	α >0.7 to 0.96 for domains excluding general health and vitality	-	-	Self-reported ICC = 0.30 to 0.96, Proxy ICC = 0.24 to 0.76, Self and proxy ICC = 0.28 to 0.80	-	Physical and mental construct validity supported – limitations in social functioning Measures of; • Global measures of function r = 0.16–0.69 • Disability (F = 19.7 – 48.8, df = 5, all p<10) • positive correlation with work and study • Depression r = >0.50 • Cognition <0.50 • no correlation with age • fatigue r = -0.31- -0.72 • summary scores had weak to no association with patients rating of severity of symptoms and quality of life (r = 0.06 - -0.69) More sensitive than BI Discriminates between TBI, MS and stroke severity Inconsistent discrimination in TBI severity using summary scores	Dimensions ES = 0.01–0.30 SRM = 0.39 – 0.02 (1-3mo) SRM = -0.15 to 0.88 (3 - 6mo) Less responsive than FIM	Stroke (< 6 mo) floor effect = 23 – 85%, ceiling effect = 16 – 54% Stroke (> 6mo) Floor effect = 17 – 61%), Ceiling effect = 16 - 52%). MS Floor effect = 21 – 85% Ceiling effect = 18 – 74% TBI Floor effect = 44 -	(Anderson et al., 1996; Dorman et al., 1999; Dorman et al., 1998; Duncan et al., 1997; Findler et al., 2001; Freeman et al., 2000; Freeman et al., 1996; Guilfoyle et al., 2010; Hagen et al., 2003; Herrmann et al., 1996; Hobart et al., 2002; Mackenzie et al., 2002; Madden et al., 2006; Moore et al., 2004; Murrell et al., 1999; O'Mahony et al., 1998; Pittock et al., 2004; Riazi et al., 2003; Robinson

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
										57% Ceiling effect = 17 – 38% MID: Physical functioning 4-9, role physical 6-8, social functioning 6-7, PCS 6 points.	Jr et al., 2009; Salter et al., 2008; Vickrey et al., 1997; Vickrey et al., 1995; Williams et al., 1999)

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
10MWT (Obs perf measure of impairment and mobility)	-	-	-	-	-	ICC 0.946–0.979	-	Measures of; • Activity r = -0.31 P<0.05 • Participation r = -0.21 • Mobility activity r = 0.41–0.71 • Impairment r = 0.044-0.175 • Predicts DC walking speed (10MWT and MAS Item 2) R^2 = 0.36. LOA 10MWT/6 min walk test -13.7 to 13.4 m/min) No relationship between different gait velocities	ES = 0.55–1.44 mES 0.45 SRM 0.93	Floor effect 2.8 - 66.8% Nil ceiling effect.	(Donovan et al., 2008; Hirsch et al., 2014; Kuys et al., 2009; Miller et al., 2013; Mudge & Stott, 2009; Salbach et al., 2001; Schmid et al., 2012; Scrivener et al., 2014)

Measurement tool	Content validity	Structural validity	Internal consistency	Reliability			Measurement error	Construct validity	Responsiveness	Interpretability	References
				Inter	Intra	Retest					
UL-MAS (Obs perf measure of upper limb activity)	-	Unidimensional scale Inconsistencies within hierarchical scoring of items Wrist radial deviation in > 65 years not indicative of motor function	Summed α = 0.83 Item 6 0.893, item 7 0.889, item 8 0.854	% agreement mean 88 – 95 (40 - 100%) Kappa 0.93 – 1.0, mean % agreement 88-95	Kendall's (Tau) 0.74 -0 1.00	-	-	predictive and discriminant validity within UL function Did not predict walking speed r = 0.06 – 0.09 Correlation with global measures of function <0.70 Rasch based more precise than summative scoring, admission (15% RP, 1.15; 95% CI: 1.01, 1.40) and discharge (11%, RP, 1.11; 95% CI: 1.02, 1.23)	-	Floor effect 0 – 38%. Ceiling effect 0 – 67%.	(Carr et al., 1985; Johnson & Selfe, 2004; Khan et al., 2013; Kuys et al., 2009; Lannin, 2004; Loewen & Anderson, 1988; Loewen & Anderson, 1990; Miller et al., 2010; Pickering et al., 2010; Sabari et al., 2005)

ARAT = Action Research Arm Test, ArmA = Arm Activity Measure, AQoL = Assessment of Quality of Life, BI = Barthel Index, BI (C&W) = Barthel Index - Collin & Wade version, CMSA = Chedoke-McMaster Stroke Assessment, DAS = Disability Assessment Scale, EQ-5D = EuroQol – 5 dimension, FAT = Frenchay Arm Test, mFAT = modified Frenchay Arm Test, FIM = Functional Independence Measure, GAS = Goal Attainment Scale, GAS – 10pt = Goal Attainment Scale – 10 point, Global Ax = Global Assessment Scale, KleinBell ADL = Klein-Bell Activities of Daily Living scale, LASIS = Leeds Adult Spasticity Impact Scale, SF-36 = Medical Outcome Study 36-Item Short-Form Health Survey, MAL = Motor Activity Log, MAL-5 = Motor Activity Log - 5, MAL-28 = Motor Activity Log - 28, MI = Motricity Index, NHPT = Nine Hole Peg Test, OHS = Oxford Handicap Scale, PDS/CBS = Patient Disability Scale / Carer Burden Scale, RMA = Rivermead Motor Assessment, RMA-UL = Rivermead Motor Assessment - Upper Limb, SA-SIP = Stroke-Adapted Version of the Sickness Impact Profile, SIS = Stroke Impact Scale, 10MWT = Ten Metre Walk Test, UL MAS = Upper Limb Motor Assessment Scale.

Appendix Table 7. *Terwee criteria for good measurement properties*

Measurement property	Rating	Criteria
Structural validity	+	**CTT** CFA: CFI or TLI or comparable measure > 0.95 OR RMSEA < 0.06 OR SRMR < 0.08[a] **IRT/Rasch** No violation of unidimensionality[b]: CFI or TLI or comparable measure > 0.95 OR RMSEA < 0.06 OR SRMR < 0.08 *AND* no violation of local independence: residual correlations among the items after controlling for the dominant factor < 0.20 OR Q3's < 0.37 *AND* no violation of monotonicity: adequate looking graphs OR item scalability > 0.30 *AND* Adequate model fit IRT: χ^2 > 0.001 Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized values > −2 and < 2
	?	CTT: not all information for '+' reported IRT/Rasch: model fit not reported
	-	Criteria for '+' not met
Internal consistency	+	At least low evidence[c] for sufficient structural validity[d] AND Cronbach's alpha(s) ≥0.70 for each unidimensional scale or subscale[e]
	?	Criteria for "At least low evidence[c] for sufficient structural validity[d]" not met
	-	At least low evidence[c] for sufficient structural validity[d] AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale[e]
Reliability	+	ICC or weighted Kappa ≥0.70
	?	ICC or weighted Kappa not reported
	-	ICC or weighted Kappa <0.70
Measurement error	+	SDC or LoA < MIC[d]
	?	MIC not defined
	-	SDC or LoA > MIC[d]
Hypotheses testing for construct validity	+	The result is in accordance with the hypothesis[f]
	?	No hypothesis defined (by the review team)
	-	The result is not in accordance with the hypothesis[f]
Cross-cultural validity\measurement invariance	+	No important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's R^2 <

Measurement property	Rating	Criteria
		0.02)
	?	No multiple group factor analysis OR DIF analysis performed
	-	Important differences between group factors OR DIF was found
Criterion validity	+	Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70
	?	Not all information for '+' reported
	-	Correlation with gold standard < 0.70 OR AUC < 0.70
Responsiveness	+	The result is in accordance with the hypothesis[f] OR AUC ≥ 0.70
	?	No hypothesis defined (by the review team)
	-	The result is not in accordance with the hypothesis[f] OR AUC < 0.70

The criteria are based on, e.g., Terwee et al., (2007) and Prinsen et al., (2018)

AUC area under the curve, CFA confirmatory factor analysis, CFI comparative fit index, CTT classical test theory, DIF differential item functioning, ICC intraclass correlation coefficient, IRT item response theory, LoA limits of agreement, MIC minimal important change, RMSEA root mean square error of approximation, SEM standard error of measurement, SDC smallest detectable change, SRMR standardized root mean residuals, TLI Tucker–Lewis index

"+" = sufficient, "−" = insufficient, "?" = indeterminate

[a]To rate the quality of the summary score, the factor structures should be equal across studies

[b]Unidimensionality refers to a factor analysis per subscale, while structural validity refers to a factor analysis of a (multidimensional) patient- reported outcome measure

[c]As defined by grading the evidence according to the GRADE approach

[d]This evidence may come from different studies

[e]The criteria 'Cronbach alpha < 0.95' was deleted, as this is relevant in the development phase of a PROM and not when evaluating an existing PROM

[f]The results of all studies should be taken together and it should then be decided if 75% of the results are in accordance with the hypotheses

Appendix Table 8. *Criteria to quantify strength of relationship*

Effect size	Cohen's d criteria: <0.5 for poor, 0.5-0.75 for moderate, 0.75-0.9 for good, and >0.90 for excellent.
Correlations	Cohen's r criteria: 0.10 - <0.30 small, 0.30 - <0.50 medium, ≥0.50 large

Appendix Table 9. *COSMIN ratings (ARAT)*

Study author, year	Reliability					Validity				Responsiveness
	Internal Consistency	Inter rater	Intra rater	Test retest	Measurement Error	Content Validity	Structural Validity	Hypothesis Testing	Cross cultural Validity	
Barden et al., (2014)								Fair		Fair
Barden et al., (2012)								Fair		
Barreca et al., (2005)								Good		Good
Barreca et al., 2006 (a) (2006)								Fair		Fair
Barreca et al., (2006) (b)								Poor		
Beebe & Lang (2009)								Fair		Fair
Blennerhassett et al., (2010)								Fair		Fair
Burridge et al., (2009)								Fair		
Celik et al., (2010)								Poor		
de Weerdt & Harrison (1985)								Fair		Good
Dromerick et al., (2006)								Good		
Edwards et al., (2012)								Good		Fair
Fleming et al., (2014)								Fair		
Lang et al., (2006)								Fair		Fair

Study author, year	Internal Consistency	Reliability			Measurement Error	Validity				Responsiveness
		Inter rater	Intra rater	Test retest		Content Validity	Structural Validity	Hypothesis Testing	Cross cultural Validity	
Lyle (1981)							Excellent			
McDonnell et al., (2006)				Fair				Fair		
Morris et al., (2013)								Fair		
Notley et al., (2007)								Fair		
O'Dell et al., (2013)								Fair		Poor
Page et al., (2015)		Good						Good		
Page et al., (2012)		Fair			Fair			Fair		
Rabadi & Rabadi (2006)								Fair		Fair
Rand & Eng (2015)								Fair		
Stinear et al., (2012)								Poor		
Urbin et al., (2015)								Fair		
Yozbatiran et al., (2008)		Good	Good					Fair		

Criterion validity not included as no agreed gold standard in upper limb measurement exits,
Lang, Edwards, Birkenmeier & Dromerick 2008 not included in table as Interpretability not given COSMIN rating

Appendix Table 10. *Quality of measurement properties (ARAT)*

Study author, year	Content Validity	Structural validity (CTT)	IRT/Rasch Analysis	Internal consistency	Construct validity	Measurement Error	Reliability (intra)	Reliability (Inter)	Reliability (retest)	Responsiveness	Interpretability	Floor or ceiling effect
Barden et al., (2014)	0	0	0	0	?	0	0	0	0	?	0	-
Barden et al., (2012)	0	0	0	0	?	0	0	0	0	0	0	-
Barreca et al., (2006) (a)	0	0	0	0	?	0	0	0	0	+	0	0
Barreca et al., (2006) (b)	0	0	0	0	?	0	0	0	0	0	0	0
Barreca et al., (2005)	0	0	0	0	+	0	0	0	0	+	0	0
Beebe & Lang (2009)	0	0	0	0	?	0	0	0	0	?	0	0
Blennerhassett et al., (2010)	0	0	0	0	?	0	0	0	0	?	0	0
Burridge et al., (2009)	0	0	0	0	?	0	0	0	0	0	0	0
Celik et al., (2010)	0	0	0	0	?	0	0	0	0	0	0	-
de Weerdt & Harrison (1985)	0	0	0	0	+	0	0	0	0	?	0	0
Dromerick et al., (2006)	0	0	0	0	?	0	0	0	0	0	0	-
Edwards et al., (2012)	0	0	0	0	?	0	0	0	0	?	0	+/-
Fleming et al., (2014)	0	0	0	0	?	0	0	0	0	0	0	+
Lang et al., (2008)	0	0	0	0	0	0	0	0	0	0	+	0

Study author, year	Content Validity	Structural validity (CTT)	IRT/Rasch Analysis	Internal consistency	Construct validity	Measurement Error	Reliability (intra)	Reliability (Inter)	Reliability (retest)	Responsiveness	Interpretability	Floor or ceiling effect
Lang et al., (2006)	0	0	0	0	?	0	0	0	0	+	0	0
Lyle (1981)	0	0	+	0	0	0	0	0	0	0	0	0
McDonnell et al., (2006)	0	0	0	0	?	0	+	0	0	0	0	0
Morris et al., (2013)	0	0	0	0	-	0	0	0	0	0	0	0
Notley et al., (2007)	0	0	0	0	?	0	0	0	0	0	0	0
O'Dell et al., (2013)	0	0	0	0	+	0	0	0	0	?	0	0
Page et al., (2015)	0	0	0	0	+	0	+	0	0	0	0	0
Page et al., (2012)	0	0	0	0	?	?	+	0	0	0	0	0
Rabadi & Rabadi (2006)	0	0	0	0	?	0	0	0	0	?	0	0
Rand & Eng (2015)	0	0	0	0	?	0	0	0	0	0	0	0
Stinear et al., (2012)	0	0	0	0	?	0	0	0	0	0	0	0
Urbin et al., (2015)	0	0	0	0	+	0	0	0	0	0	0	0
Yozbatiran et al., (2008)	0	0	0	0	?	0	+	+	0	0	0	0

+ = positive rating, ? = indeterminant rating, - = negative rating, 0 = no information available.

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
9HPT	Beebe & Lang (2009)	33	stroke 1 - 6 months post	inpatient rehab and community	Construct validity - hypothesis testing, responsiveness	CTT	Yes completed MAS as additional testing to represent sample,Mean (SD): 0.58 (0.67) Range 0-3
9HPT	Benedict, Holtzer, Motl, Foley, Kaur, Hojnacki & Weinstock-Guttman (2011)	211	Multiple Sclerosis	outpatients	Construct validity - hypothesis testing	CTT	not stated
9HPT	Costelloe, O'Rourke, McGuigan, Walsh, Tubridy & Hutchinson (2008)	n = 200 at baseline and n=150 at follow-	Multiple Sclerosis	outpatients	Construct validity - Hypothesis testing	CTT	nil stated
9HPT	Goodkin, Hertsgaard & Seminary (1988)	21 patients in control (63 successive administrations) and prospective group 68 patients (320	Multiple Sclerosis	outpatient clinic	predictive validity – hypothesis testing, interpretability	CTT	nil stated
9HPT	Heller, Wade, Wood, Sunderland, Hewer & Ward (1987) (a)	reliability n=10 inter rater n=20 obs, test retest	stroke - acute andchronic	hospital, outpatient rehab	reliability - test retest, inter rater	CTT	
9HPT	Heller, Wade, Wood, Sunderland, Hewer & Ward (1987) (b)	sensitivity n= 56	stroke - acute andchronic	hospital, outpatient rehab	hypothesis testing - sensitivity	CTT	
9HPT	Jacob-Lloyd, Dunn, Brain & Lamb (2005)	58	acute stroke	inpatient rehabilitation	floor & ceiling, responsiveness, (convergent validity) hypothesis testing	CTT	

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
9HPT	Morris, van Wijck Joice & Donaghy (2013)	85	stroke 6 months post	cohort of community dwelling clients	hypothesis testing	CTT	spasticity not stated in inclusion/exclusion criteria
9HPT	Poole, Nakamoto, McNulty, Montoya, Weill,Dieruf & Skipper (2010)	56	Multiple Sclerosis	outpatient clinics	hypothesis testing	CTT	
9HPT	Schwid, Goodman, McDermott, Bever & Cook (2002)	27	Multiple Sclerosis	outpatient but details not given	measurement error	CTT	not specifically stated
AQoL	Hawthorne, Gruen & Kaye (2009)	56	TBI > 6mo	community	hypotheses testing	CTT	not reported
AQoL	Sturm, Osborne, Dewey,Donnan, Macdonnell & Thrift (2002)	93	stroke <6mo		construct validity, interpretability	CTT	not reported
ARAT	Beebe & Lang (2009)	33	stroke 1 - 6 months post	inpatient rehab and community	hypothesis testing, responsiveness	CTT	completed MAS as additional testing to represent sample, MEAN (SD): 0.58 (0.67) Range 0-3
ARAT	Burridge, Turk, Notley, Pickering & Simpson (2009)	17	chronic stroke	outpatient rehabilitation	hypothesis testing	CTT	spasticity not stated in inclusion/exclusion criteria
ARAT	de Weerdt & Harrison (1985)	53	acute stroke	general teaching hospital	responsiveness, hypothesistesting	CTT	spasticity not stated in inclusion/exclusion criteria
ARAT	Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, Edwards (2006)	39	stroke	part of VECTORS trial - inpatient rehab	hypothesis testing	CTT	not stated

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
ARAT	Fleming, Newham, Roberts-Lewis, Sorinola (2014)	33	chronic stroke > 3months	outpatient rehabilitation	hypothesis testing - predictive	CTT	excluded if MAS > 4otherwise included
ARAT	Lang, Edwards, Birkenmeier, Dromerick (2008)	52 - but 12 only included in MCID analysis	acute stroke	inpatient rehabilitation	Minimal clinically importantdifference	CTT	spasticity not stated in inclusion/exclusion criteria
ARAT	Lang, Wagner, Dromerick & Edwards (2006)	50	acute stroke	Barnes-Jewish Hospital St Louis	responsiveness, Construct validity – hypothesis testing	not IRT	yes one measure - elbow joint spasticity
ARAT	Lyle (1981)	20	mixed neuro - TBI,stroke, aneurysm ranging from 42 years to one monthpost mean 46 months	outpatient	structural validity	CTT	not stated
ARAT	Morris, van Wijck Joice & Donaghy (2013)	85	stroke 6 months post	cohort of community dwelling clients	hypothesis testing	CTT	spasticity not stated in inclusion / exclusion criteria
ARAT	Rabadi & Rabadi (2006)	104	acute stroke	inpatient rehab	hypothesis testing, responsiveness	CTT	spasticity not stated in inclusion / exclusion criteria
ARAT	Rand & Eng (2015)	32	acute and 12 mo post	community	hypothesis testing - predictive validity	CTT	nil stated
ARAT	Yozbatiran, Der-Yeghiain & Cramer (2008)	12	stroke >6mo		interrater reliability, intra raterreliability, hypothesis testing	CTT	spasticity not stated in inclusion / exclusion criteria

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
ArmA	Ashford, Slade & Turner-Stokes 2013	experienced prof n=10, survey with clinicians n=36, patient and carers n=13 pairs			Content validity - tool development		**Y**es
ArmA	Ashford, Jackson & Turn**er-Stokes**	16	mixed neuro (stroke TBI)		Content validity		**Y**es
ArmA	Ashford, Siegert & Alexandrescu 2016	92	mixed neuro (stroke, TBI, MS,other)	rehabilitation	Structural validity	IRT	**Y**es
ArmA	Ashford, Slade, Nair & Turner-Stokes 2014	58	MS, stroke, ABI	inpatient and community	responsiveness	CTT	**Y**es
ArmA	Ashford, Slade, Turner-Stokes 2013	46 clinicians, 26 patients / carers	stroke		content validity		**Y**es
ArmA	Ashford, Turner-Stokes, Siegert & Slade 2013	92	mixed neuro		structural validity (unidimensionality), constructvalidity (convergent and divergent), reliability (test-retest), internal consistency, responsiveness, interpretability		
BI	Ali, Fulton, Quinn, Bradyon behalf of VISTA (2013)	867	stroke 3 months	assumed hospital	concurrent validity – hypothesis testing	CTT	not reported
BI	Filiatrault et al (1991)	18	stroke	hospital	hypothesis testing responsiveness	CTT	not reported
BI	Kwon et al (2004)	1680	stroke	hospital	hypothesis testing	CTT	not reported
BI	Mahoney & Barthel (1965)	not given	not given		content validity,		not reported

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
BI	Wellwood et al (1995)	152	stroke	hospital	hypothesis testing	CT	not reported
BI (C&W)	Barer & Murphy 1993	730	stroke	hospital	hypothesis testing, predictive validity, structural validity, responsiveness	CTT	not reported
BI (C&W)	Collin et al 1988	25	mixed (stroke n=13, TBI n=11, other n=1)	hospital	reliability, content validity	CTT	not reported
BI (C&W)	Dennis et al 2000	417	stroke	mixed	hypothesis testing	CTT	not reported
BI (C&W)	Dennis, Wellwood & Warlow 1996	152	chronic stroke > 1year post		convergent construct validity	CTT	not reported
BI (C&W)	Green et al 2001	22	stroke	community	test retest measurement error	CTT	not reported
BI (C&W)	Houlden, Edwards, McNeil, Greenwood 2006				Responsiveness	CTT	not reported
BI (C&W)	Sarker, Rudd, Douiri, Wolfe 2012	238	stroke	community	hypothesis testing	CTT	not reported
BI (C&W)	Wade & Hewer 1987	976	acute stroke	hospital andcommunity	validity – predictive	CTT	not stated
BI (C&W)	Wallace et al 2002	372	stroke <6mo	hospital	Responsiveness	CTT	not reported
BI (C&W)	Wilkinson et al 1997	106	stroke	community	hypothesis testing	CTT	not reported
DAS	Brashear, Zafonte, Corcoran, Galvez-Jimenez, et al 2002	9 patients, 10 raters completed 2 assessments of each pt	> or equal to 6 months post stroke	university medical centre - outpatients	reliability - inter intra rater	CTT	yes all pts
DAS	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	279	chronic mean > 5years	community	hypothesis testing	CTT	yes

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
EQ-5D	Alderman et al 2001	11	mixed	rehabilitation	hypothesis testing	CTT	not reported
EQ-5D	Barton 2008b: comparison of EQ-5D and EQ6D >/=45	62	stroke	community	hypothesis testing	CTT	not reported
EQ-5D	Barton, Sach, Doherty, Avery, Jenkinson, Muir et al 2008a	57	stroke	community	hypothesis testing	CTT	not reported
EQ-5D	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	279	chronic mean > 5 years	community	hypothesis testing	CTT	yes
EQ-5D	Dorman et al 1999	≥531 (p2150)	stroke	hospital	hypotheses testing	CTT	not reported
EQ-5D	Fisk, Brown, Sketris, Metz & Stadnyk 2005	187	Multiple Sclerosis	outpatient clinics	Construct validity - hypothesis testing test retest reliability	CTT	not specifically stated
EQ-5D	Gillard et al 2015	274 without 54 with spasticity	stroke	community	hypothesis testing	CTT	yes
EQ-5D	Kohn, Sidovar, Kaur, Zhu & Coleman 2014	3044	MS	community	hypothesis testing, interpretability	CTT	not reported
EQ-5D	Kuspinar & Mayo 2013	185	MS	community	hypothesis testing, **content validity**	CTT	not reported
EQ-5D	Kuspinar, Finch, Pickard & Mayo 2014	189	MS	community rehabilitation	hypothesis testing – discriminant validity	CTT	not reported
EQ-5D	Moore, Wolfson, Alexandrov & Lapierre 2004	114	MS	community	hypothesis testing	CTT	not reported
EQ-5D	Nicholl et al 2001	88	MS		hypothesis testing floor and ceiling	CTT	not reported

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
EQ-5D	Peters 'Change in healthstatus' (2014) Peters, Crocker, Jenkinson, Doll& Fitzpatrick 2014	93 index, 82 VAS	stroke		responsiveness	CTT	not reported
EQ-5D	Xie et al 2006	Sample of 38640 (n=1040 with stroke)	stroke	community	hypothesis testing	CTT	not reported
FIM	Brown, Therneau, Schultz, Niewczyk & Granger 2015	148 367	stroke	hospital	hypothesis testing – predictive ability	CTT	not reported
FIM	Corrigan, Smith-Knapp & Granger 1997	95	TBI	community	hypotheses testing	CTT	not reported
FIM	Cullen, Krakowski & Taggart 2014	59	TBI	rehabilitation	hypothesis testing	CTT	not reported
FIM	Cuthbert, Harrison-Felix,Corrigan, Bell, Haarbauer-Krupa & Miller 2015	64081	TBI	community	hypothesis testing	CTT	not reported
FIM	Egan, Davis, Dubouloz, Kessler & Kubina 2014	55	stroke greater than6	community	hypothesis testing	CTT	not reported
FIM	Grant, Goldsmith & Anton 2014	11, 983	stroke < 6 mo	rehabilitationsubacute	hypothesis testing – predictive validity	CTT	not reported
FIM	Hall, Hamilton, Gordon & Zasler 1993	332	TBI	hospital	hypothesis testing interpretability	CTT	not reported
FIM	Hamilton & Granger 1994	1018	stroke <less than 6	<6mo	inter rater reliability	CTT	not reported
FIM	Hawthorne et al 2009				content validity		
FIM	Heinemann et al 1994	9961	mixed (stroke 36%)	<6 mo	structural validity	IRT	not reported
FIM	Heinemann et al 1997	129	TBI	<6mo	hypothesis testing	CTT	not reported

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
FIM	Heinemann, Linacre, Wright, Hamilton & Granger 1993	10092	stroke	rehab inpatients	structural validity	IRT	not reported
FIM	Houlden, Edwards, McNeil, Greenwood 2006	411	TBI, stroke	<6mo inpatients	responsiveness, interpretability	CTT	not reported
FIM	Kuys, Bew, Lynch, Morrison & Brauer 2009	120	acute stroke	hospital/rehabilitation	Predictive validity - hypothesis testing	CTT	nil stated
FIM	Oczkowski & Barreca 1993	113	acute stroke	hospital patients	Predictive validity - hypothesis testing	CTT	not reported
FIM	Ouellette, Timple, Kaplan, Rosenberg & Rosario 2015	407	stroke	hospital	Predictive validity - hypothesis testing	CTT	not reported
FIM	Rabadi & Vincent 2013	76	MS	community	hypothesis testing, responsiveness	CTT	yes
FIM	Sharrack, Hughes, Soudain & Dunn 1999	inter rater reliability 64, intra-rater 35, internal consistency 64, responsiveness = 25 structural validity = 25 hypothesis testing = 50	MS	community	inter, intra rater reliability,hypotheses testing, responsiveness, internal consistency	CTT	not reported
FIM	Stineman et al 1996		stroke, TBI	< 6mo	internal consistency, structural validity	CTT	not reported
GAS	Bovend'Eerdt, Dawes, Izadi, Wade 2011	29 patients 112 goals	neurological disorders - stroke n= 27, TBI n=1 MS n=1	hospital patients	reliability, measurement error	CTT	not reported

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
GAS	Brock, Black, Cotton, Kennedy, Wilson & Sutton 2009	45 stroke survivors and 23primary carers	stroke - one month - 6months	inpatient with community follow up	Concurrent validity - hypothesis testing	CTT	not reported
GAS	Doig, Fleming, Kuipers & Cornwell 2010	14	TBI	community	responsiveness, confirmatory validity - hypothesis testing	CTT	not reported
GAS	Joyce, Rockwood & Mate-Kole 1994	16	TBI <5 months post, 1 SAH 1 stroke 1 hypoxia	inpatient rehabilitation	content and construct validity, inter rater reliability, feasibility/clinical utility	CTT	not reported
GAS	Lannin 2003	n=12 goals n= 36	>6 months assumed nil **data given**	community	responsiveness	CTT	spasticity not stated in inclusion / exclusion criteria
GAS	Malec 1999	88	75% TBI, 17% stroke, 8% anoxiaplus other	community	Concurrent validity - hypothesis testing	CTT	not reported
GAS	Malec, Smigielski & DePompolo 1991	14	TBI n=12, stroken=3 anoxia n=1	outpatient	Predictive and construct validity – hypothesis testing	CTT	not reported
GAS	Khan, Pallant & Turner-Stokes 2008	n= 24 (203 goals)	MS - 58.3% secondary progressive, 21% relapsing remitting21% primary progressive	inpatient rehabilitation	Concurrent validity - hypothesis testing, responsiveness	CTT	not reported
GAS	Turner-Stokes, Baguley, De Graaff, Katrak, Davies, McCrory & Hughes 2010	90	stroke - at least 6months ago	outpatient	Concurrent validity - hypothesis testing, responsiveness	CTT	yes

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
MAL	Chen, Wolf, Zhang, Thompson & Winstein 2012	116	stroke 3-9 mo post	community	measurement error hypothesis testing	CTT	not specifically stated
MAL	Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, Edwards 2006	39	**S**troke	part of VECTORS trial - inpatient rehab	hypothesis testing	CTT	not stated
MAL	Harris & Eng 2007	93	chronic stroke > 1year	community dwelling	validity - hypothesis testing	CTT	yes
MAL	Uswatte, Taub, Morris, Light & Tompson 2006	n=222	subacute stroke 3- 12 months post	check EXCITE trialdata	structural validity	CTT	nil specifically stated
MAL-28	Uswatte, Taub, Morris, Light & Tompson 2006	patients n = 222 caregiver n = 185	subacute stroke 3- 12 months post		Internal consistency, validity - convergent & divergent, test -retest reliability	CTT	nil specifically stated
modified Frenchay Arm Test	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987	varied for each psychometric property - reliability 10; sensitivity 56	stroke - acute andchronic	hospital, outpatient rehab	reliability - test retest, inter rater, validity and sensitivity (appraised as predictive validity)	CTT	
Motricity Index	Bohannon 1999	10	acute stroke	hospital	construct validity – hypothesis testing internal consistency	CTT	not stated
Motricity Index	Collin & Wade 1990	20	acute stroke	inpatient rehabilitation	reliability, concurrent validity – hypothesis testing	CTT	Nil stated specifically
Motricity Index	Jacob-Lloyd, Dunn, Brain & Lamb 2005	58	acute stroke	inpatient rehabilitation	floor & ceiling, responsiveness, convergent validity – hypothesis testing	CTT	
Motricity Index	Stone, Patel & Greenwood 1993	84	acute - chronic stroke with **visual neglect**	hospital	Predictive validity - hypothesis testing	CTT	not stated

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
Motricity Index	Wade & Hewer 1987	976	acute stroke	hospital andcommunity	validity – predictive	CTT	not stated
OHS	Rigby, Gubitz, Eskes, Reidy, Christian, Grover & Phillips 2009	155	**S**troke	hospital and community follow up	predictive validity – hypothesis testing	CTT	nil stated
OHS	Simon, Kumar & Kendrick 2008	baseline n = 105 1st follow up n= 74 last follow up n = 53 carers	informal carers of first ever stroke survivors	acute through to community	Predictive validity - hypothesis testing	CTT	nil stated
RMA - 3 sections separated	Adams, Pickering & Taylor 1997 (1)	51	acute stroke	hospital patients	structural validity	CTT	not reported
RMA - 3 sections separated	Adams, Pickering, Ashburn & Lincoln	ranged 206(6 mo post) to 83 (12 mo post)	stroke - 6 mo post	community	structural validity	CTT	not reported
RMA - 3 sections	Collin & Wade 1990	reliability n=20 hypothesis testing (concurrent and predictive validity)	acute stroke	inpatient rehabilitation	concurrent validity	CTT	Nil stated specifically
RMA - 3 sections separated	Jones 1998	29	stroke 6 weeks post	inpatient rehabilitation	predictive validity – hypothesis testing	CTT	not reported
RMA- UL	Morris, van Wijck Joice & Donaghy 2013	85	stroke 6 months post	cohort of community dwelling clients	hypothesis testing	CTT	spasticity not stated in inclusion / exclusion criteria
RMA - 3 sections separated	Sackley 1990	90	stroke - varied 6-9 wks post to 26 wks post	hospital patients	concurrent validity - hypothesis testing	CTT	not reported
SA-SIP	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	279	chronic mean > 5years	community	hypothesis testing	CTT	**Y**es

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
SA-SIP	Edwards, Hahn, Baum &Dromerick 2006	219	stroke >6mo	community	construct validity	CTT	not reported
SF-36	Anderson 1996	90	stroke **>** 6 mo	community	hypothesis testing	CTT	not reported
SF-36	Dorman et al 1999	≥531 (p2150)	**S**troke	hospital	hypothesis testing	CTT	not reported
SF-36	Dorman, Slattery, Farrell, Dennis & Sandercock 1998	209	stroke > 6mo	community	reliability, internal consistency	CTT	not reported
SF-36	Duncan 1997		> 6 mo post stroke	community	hypothesis testing, floor andceiling	CTT	not reported
SF-36	Findler 2001		TBI		hypothesis testing	CTT	not reported
SF-36	Freeman, Hobart, Langdon & Thompson 2000	149	MS	rehab - O/P, I/P	internal consistency, hypothesis testing, responsiveness	CTT	not reported
SF-36	Freeman, Langdon, Hobart & Thompson 1996	50	MS >6	community	hypothesis testing interpretability	CTT	not reported
SF-36	Guilfoyle, et al 2010	453	TBI	outpatient clinic	structural validity, internal consistency, hypothesis testing,interpretability	CTT	not reported
SF-36	Hagen, Bugge & Alexander 2003	>136	stroke <6	hospital/rehabilitation	internal consistency, hypothesis testing, responsiveness	CTT	not reported
SF-36	Hermann, Vickrey, Hays,Cramer, Devinsky, Meador, Perrine, Myers & Ellison 1996	85	MS	community	hypothesis testing	CTT	not reported
SF-36	Hobart, Williams, Moran & Thompson 2002	177	**S**troke	inpatients (3hospitals)	structural validity, internal consistency	CTT	not reported

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
SF-36	MacKenzie 2002 'Using the SF-36'	1197	TBI 1 year post	community	structural validity, hypothesis testing	CTT	not reported
SF-36	Madden, Hopman, Bagg, Verner, O'Callaghan 2006	116	**S**troke	rehabilitation	hypothesis testing, responsiveness	CTT	not reported
SF-36	Moore, Wolfson, Alexandrov & Lapierre 2004	114	MS	community	hypothesis testing	CTT	not reported
SF-36	Murell, Kenealy, Beaumont & Lintern 1999	22 pts, tested 5 times each	MS	hospital	retest reliability floor and ceiling	CTT	not reported
SF-36	O'Mahoney 1998	104	Stroke	not reported	floor and ceiling	CTT	not reported
SF-36	Pittock 2004	185	MS	>6mo	hypothesis testing	CTT	not reported
SF-36	Riazi et al 2003	638	MS	mixed, in, outpatients and community	Discriminant validity hypothesis testing -	CTT	not reported
SF-36	Robinson, Zhao, Kim & Revicki 2009	249	MS	community	hypothesis testing, interpretability (MID)	CTT	not reported
SF-36	Salter et al 2008	not reported	not reported		content validity		not reported
SF-36	Vickery, Hays, Genovese, Myers & Ellison 1997	171, 84	MS	community >6mo (10 yrs)	internal consistency, test retest reliability, hypothesis testing	CTT	not reported
SF-36	Vickrey 1995	179	MS	community >6mo (10 **y**s)	hypothesis testing	CTT	not reported
SF-36	Ware & Sherbourne 1992	not reported	not reported	not reported	content validity		
SF-36	Williams 1999	71	< 6 mo stroke	hospital	hypothesis testing	CTT	not reported
SIS	Duncan et al 1999	91	stroke <6mo		content validity		
SIS	Duncan, Bode, Lai, Perera GAIN 2003	696 pts 1264 returned questionnaires	acute stroke 1-3 mo	community	structural validity	IRT	nil stated

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
SIS	Duncan, Lai, Tyler, Perera, Reker & Studenski 2002	287 pts andtheir proxies	stroke <6mo	community	reliability, concurrent validity - hypothesis testing	CTT	nil reported
SIS	Duncan, Reker, Kwon, Lai, Studenski, Perera, Alfrey & Marquez 2005	28 mail 30 telephone	stroke >3mo	community	internal consistency, retest reliability	CTT	not reported
SIS	Eriksson, Baum, Wolf & Connor 2013	116	Stroke	community	hypothesis testing	CTT	not reported
SIS	Jenkinson, Fitzpatrick, Crocker & Peters 2013	73	chronic stroke	community	internal consistency, structural validity, convergent validity - hypothesis testing	CTT	nil reported
SIS	Kwon, Duncan, Studenski, Perera, Lai & Reker 2006	95	stroke <6mo	community	validity - construct, convergent and discriminant of telephone mode	CTT	not reported
SIS	Lai, Studenski, Duncan & Perera 2002	81	stroke >3mo	community	hypothesis testing (sensitivity,discriminant validity), interpretability floor and ceilingeffect	CTT	not reported
SIS	Salter et al 2008	not given	not given		content validity		not reported
SIS	Wolf & Koster 2013	96	chronic stroke > 6mo	community	predictive validity – hypothesis testing	CTT	nil reported
10MWT	Donovan, Lord, McNaughton & Weatherall 2008	30	chronic stroke > 6months	community dwelling	Concurrent validity - hypothesis testing	CTT	
10MWT	Hirsch, Williams, Norton & Hammond 2014	23	TBI - acute	inpatient rehabilitation - hospital	reliability - test-retest, concurrent validity	CTT	nil stated
10MWT	Kuys, Bew, Lynch, Morrison & Brauer 2009	120	acute stroke	hospital/rehabilitation	Predictive validity - hypothesis testing	CTT	nil stated

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
10MWT	Miller, Combs, Van Puymbroeck, Altenburger, Kean, Dierks & Schmid 2013	77	chronic stroke	community dwelling	validity - hypothesis testing	CTT	not stated
10MWT	Mudge & Stott 2009	49	6 months post stroke	community dwelling	Predictive validity - hypothesis testing	CTT	no nil stated
10MWT	Salbach, Mayo, Higgins, Ahmed, Finch & Richards 2001	50	acute stroke	hospital andcommunity	Responsiveness	CTT	nil stated
10MWT	Schmid, Van Puymbroeck, Altenburger, Dierks, Miller, Damush & Williams 2012	77	chronic stroke > 6months post	community	Concurrent validity - hypothesis testing	CTT	nil stated
10MWT	Scrivner, Schurr & Sherrington 2014	190	acute stroke	hospital	responsiveness, floor and ceiling effect	CTT	nil stated
UL MAS	Carr, Shephard, Nordholm & Lynne 1985	n=5 participants, n=40 observations	stroke 14 weekspost	community	inter rater reliability	CTT	not reported
UL MAS	Khan, Chien & Brauer 2013	497	acute stroke timeframe not given	inpatient rehabilitation	structural validity, hypothesis testing	IRT	not reported
UL MAS	Kuys, Bew, Lynch, Morrison & Brauer 2009	105	acute stroke	hospital/rehabilitation	Predictive validity - hypothesis testing	CTT	nil stated
UL MAS	Lannin 2004	27	acute stroke	rehabilitation unit	internal consistency, Structural validity	CTT	nil stated
UL MAS	Loewen & Anderson 1988 (full MAS with clonus section only omitted).	7 patients acute stroke 14 therapists with 637 observations	acute stroke	hospital	inter and intra rater reliability	CTT	nil stated

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
UL MAS	Loewen & Anderson 1990 (full MAS minus clonus section)	50 initially 30 at discharge	acute stroke	hospital	predictive validity	CTT	nil stated
UL MAS	Miller, Slade, Pallant, Galea 2010	80 participants 140 observations	acute/subacute stroke	inpatient rehabilitation	structural validity, discriminant validity - hypothesis testing internal consistency	IRT	not reported
UL MAS	Pickering, Hubbard, Baker & Parsons 2010	25 participants 40 ax results	acute stroke (within first month)	inpatient rehabilitation	structural validity	IRT	not reported
UL MAS	Sabari, Lim, Velozo, Lehman, Kieran & Lai 2005	100	acute/subacute stroke	rehabilitation	structural validity	IRT	not reported
UL MAS	Johnson & Selfe 2004	26	acute stroke - within first 3 months	hospital patients	internal consistency	CTT	not reported
Excluded at COSMIN update							
9HPT	Alusi, Worthington, Glickman, Findley & Bain (2000)	30	Multiple Sclerosis	outpatient clinics	construct validity – hypothesis testing	CTT	
9HPT	Fisk, Brown, Sketris, Metz & Stadnyk (2005)	187	Multiple Sclerosis	outpatient clinics	Construct validity - hypothesis testing	CTT	not specifically stated
9HPT	Marrie & Goldman (2011)	44	Multiple Sclerosis	not stated	hypothesis testing	CTT	
9HPT	Rossier & Wade (2002)	43	MS	community dwelling	Concurrent validity – hypothesis testing	CTT	not specifically stated

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
9HPT	Sunderland, Tinson, Bradley & Hewer (1989)	38	acute stroke	hospital setting	hypothesis **testing,** responsiveness, floor and ceiling effect	CTT	yes, initially 12 participants (31%) and over the 6 months 7 (22%) not rated with MAS but rating given for abnormal resistanceto movement - mild or severe at each joint.
AQoL	Turner-Stokes, Baguley, De Graaff, Katrak, Davies, McCrory & Hughes (2010)	90	stroke - at least 6months ago	outpatient	Concurrent validity - hypothesis testing, responsiveness	CTT	yes
ARAT	Barden, Baguley, Nott, Chapparo (2014)	28 stroke = 22 TBI =6	ABI -	outpatient	Concurrent validity - hypothesis testing responsiveness	CTT	yes present
ARAT	Barden, Nott, Heard, Chapparo & Baguley (2012)	38	UL spasticity postABI	outpatient spasticity clinics - community based pts	hypothesis testing, responsiveness	CTT	yes
ARAT	Barreca, Stratford, Lambert, Masters & Streiner (2005)	39	group 1: acute/mild mod n=24 (8 weeks or less poststroke CMSA arm and hand between7 and 11) group 2 chronic/severe n=15 (3months+ post stroke CMSA arm and hand stage 5 or less)	inpatient/outpatient rehabilitation facilities	Construct validity - hypothesis testing, responsiveness	CTT	spasticity not stated in inclusion/exclusion criteria
ARAT	Barreca, Stratford, Masters, Lambert, Griffiths (2006)	105	stroke	4 inpatient and outpatient rehabilitation facilities	responsiveness, hypothesis testing	CTT	spasticity not stated in inclusion / **exclusion criteria**

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
ARAT	Barreca, Stratford, Masters, Lambert, Griffiths, McBay (2006)	39	acute (< 8 weeks post) and chronic (3 months > 1 year) stroke	not stated	hypothesis testing	CTT	spasticity not stated in inclusion / **exclusion criteria**
ARAT	Blennerhassett, Avery & Carey (2010)	22	acute stroke	inpatient rehabilitation	hypothesis testing, responsiveness	CTT	spasticity not stated in inclusion / **exclusion criteria**
ARAT	Celik, O'Malley, Boake, Levin, Yozbatiran & Reistetter (2010)	9	chronic stroke	assumed community rehabilitation	hypothesis testing,	CTT	participants mildly impaired no mention of spasticity in criteria
ARAT	Edwards, Lang, Wagner,Birkenmeier & Dromerick (2012)	51 at day 0 and day 14, 40 at day 90	acute stroke	Inpatient	Concurrent validity - hypothesis testing responsiveness	CTT	spasticity not stated in **in**clusion / exclusion criteria
ARAT	McDonnell, Hillier, Ridding & Miles (2006)	17	subacute stroke	not stated	hypothesis testing, reliability	CTT	spasticity not stated in inclusion / **exclusion criteria**
ARAT	Notley, Turk, Pickering, Simpson & Burridge (2007)	10	acute/chronic stroke	outpatient rehabilitation	hypothesis testing	CTT	spasticity not stated in inclusion / **exclusion criteria**
ARAT	O'Dell, Kim, Rivera, Fieo, Christos, Polistena, Fitzgerald, Gorga (2013)	32	chronic stroke (>6months)	community dwelling	hypothesis testing, responsiveness, ceiling effect	CTT (IRT process for AMAT-9 only but ARAT not included in that analysis)	included if > 3 months post botulinum toxin injections as considered part of rehabilitation but specific focus with spasticity not present

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
ARAT	Page, Hade & Persch (2015)	32	chronic stroke	community	reliability, concurrent validity	CTT	excessive spasticity excluded MAS greater than or equal2 NO evidence of included spasticity
ARAT	Page, Levine, Hade (2012)	29	chronic stroke >1year post	outpatient rehabilitation clinics	concurrent validity - hypothesistesting & intrarater reliability, measurement error	CTT	excluded if excessivespasticity MAS greater than or equalto 3 but no evidence of people with spasticity included.
ARAT	Stinear, Barber, Petoe, Anwar & Byblow (2012)	40	acute stroke	hospital	Predictive validity - hypothesis testing	CTT	
ARAT	Urbin, Waddell & Lang (2015)	35	n=8 <30 days n=27 ? 6 mo post stroke	inpatient and community	hypothesis testing	CTT	
BI	Ashford, Turner-Stokes, Siegert & Slade 2013	92	stroke, TBI, MS	outpatient rehabilitation units	Responsiveness	CTT	**Y**es
BI	Chen, Chen, Hreha, Goedert & Barrett 2015	108	stroke	hospital - rehabilitation unit	structural validity, hypothesis testing	IRT / CTT	not reported
BI	Duncan, Lai, Tyler, Perera, Reker & Studenski 2002	287 pts andtheir proxies	acute stroke	community	concurrent validity – hypothesis testing	CTT	nil reported
BI	Khan, Pallant & Turner-Stokes 2008	n= 24 (203 goals)	MS - 58.3% secondary progressive, 21% relapsing remitting 21% primary progressive	inpatient rehabilitation	Concurrent validity - hypothesis testing responsiveness	CTT	not reported

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
BI	Lai, Studenski, Duncan & Perera 2002	81	3 months post stroke	community	Sensitivity, discriminant - hypothesis testing, interpretability floor and ceiling effect	CTT	not reported
BI	Maujean, Davis, Kendall,Casey & Loxton 2014	80	stroke	community	hypothesis testing	CTT	not reported
BI	Sprigg, Selby, Fox, Berge, Whynes, Philip & Bath 2013	2238	stroke	community	hypothesis testing	CTT	not reported
BI	Wolf & Koster 2013	96	chronic stroke > 6mo	community	predictive validity – hypothesis testing	CTT	nil reported
BI (C&W)	Rossier & Wade 2002	43	MS		responsiveness, floor and ceiling	CTT	not reported
CMSA	Dang, Ramsaran, Street, Syed, Barclay-Goddard, Stratford & Miller 2011	74	acute stroke	hospital patients	Predictive validity - hypothesis testing	CTT	not reported
CMSA	Gowland 1990	not reported	stroke		content validity		not reported
CMSA	Gowland, Stratford, Ward, Moreland, Torresin, Van Hullenar, Sanford, Barreca, Vanspall & Plews 1993	n= 28 responsiveness n=32 for validity and reliability	stroke - subacute 236 weeks post	rehabilitation - inpatient and dayhospital	reliability (inter, intra, test- retest), validity (construct and concurrent), responsiveness	CTT	yes 13 spastic hemiplegia
CMSA	Moreland et al 1993	not reported	stroke		content validity		not reported
CMSA	Coderre, Zeid, Dukelow, Demmer, Moore, Demers, Bretzke, Herter, Glasgow, Norman, Bagg & Scott 2010	52	acute 2-3 weeks post	**H**ospital	hypothesis testing	CTT	nil stated

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
CMSA	Manns, Tomczak, Jelani, Cress & Haennel 2009	10	stroke > 6 months	community	Concurrent validity - hypothesis testing	CTT	not reported
CMSA	Oczkowski & Barreca 1993	113	acute stroke	hospital patients	Predictive validity - hypothesis testing	CTT	not reported
CMSA	Semrau, Herter, Scott, Dukelow 2015	76	acute stroke < 6 mo	hospital andcommunity	hypothesis testing	CTT	not stated
DAS	Brashear, Gordon, Elovic, Kassicieh, et al 2002	122	chronic stroke 6 months post	not stated assumed community	validity - hypothesis testing	CTT	yes all pts
EQ-5D	Ali, Fulton, Quinn, Bradyon behalf of VISTA 2013	867	stroke 3 months	assumed hospital	concurrent validity – hypothesis testing	CTT	not reported
EQ-5D	Jenkinson, Fitzpatrick, Crocker & Peters 2013	73	chronic stroke	community	Convergent validity hypothesis testing	CTT	not stated
EQ-5D	Sprigg, Selby, Fox, Berge, Whynes, Philip & Bath 2013	2238	stroke	community	hypothesis testing	CTT	not reported
FIM	Cusick, Lannin, Hanssen & Allaous 2014	33	TBI	**H**ospital	hypothesis testing	CTT	not reported
FIM	Herrman, Black, Lawrence, Szekely & Szalai 1998	3mo n = 150 1 year n = 136	acute stroke	**H**ospital	Concurrent validity - hypothesis testing	CTT	nil stated
FIM	Khan, Pallant & Turner-Stokes 2008	n= 24 (203 goals)	MS - 58.3% secondary progressive, 21% relapsing remitting21% primary progressive	inpatient rehabilitation	Concurrent validity - hypothesis testing, responsiveness	CTT	not reported
FIM	Kwon, Duncan, Studenski, Perera, Lai & Reker 2006	95	stroke	**I**npatient	hypothesis testing	CTT	not reported

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
FIM	McNett, Amato, Gianakis, Grimm, Philippbar, Belle & Moran 2014	33	TBI	**H**ospital	predictive and concurrentvalidity	CTT	not reported
FIM	Perrin, Niemeier, Mougeot et al 2015	100	TBI	**H**ospital	hypothesis testing	CTT	not reported
FIM	Rabadi & Rabadi 2006	104	stroke			CTT	not reported
FIM	Semrau, Herter, Scott, Dukelow 2015	76	acute stroke < 6 mo	hospital andcommunity	hypothesis testing	CTT	not stated
FIM	Tyryshkin, Coderre, Glasgow, Herter, Bagg, Dukelow & Scott 2014	154	stroke	**H**ospital	hypothesis testing	CTT	measured MAS on demographic table .05 - 0.45 yes
FIM	Wu, Burgard & Radel 2014	15	stroke	inpatient rehabilitation	hypothesis testing	CTT	not reported
GAS	Barden, Baguley, Nott, Chapparo 2014	28 stroke=22 TBI=6	ABI -	outpatient	Concurrent validity - hypotheses testing, responsiveness	CTT	yes present
GAS	Barden, Baguley, Nott, Chapparo 2014 (b)	28	mixed - 22 stroke 6TBI	community rehabilitation	hypothesis testing, responsiveness	CTT	**Y**es
GAS	Barden, Nott, Heard, Chapparo & Baguley 2012	38	UL spasticity postABI	outpatient spasticity clinics - community based pts	hypothesis testing, responsiveness	CTT	**Y**es
GAS	Malec 2001	96	ABI - stroke, TBI,anoxia and other <10%		Concurrent validity - hypothesis testing	CTT	not reported
Global Ax	Brashear, Gordon, Elovic, Kassicieh, Marciniak, Do, Lee, Jenkins, Turkel, 2002	122	chronic stroke 6 months post	not stated assumed community	validity - hypothesis testing	CTT	yes all pts
LASIS	Ashford, Slade, Nair & Turner-Stokes 2014	48	UL spasticity mixed - TBI, stroke, MS 2% MND	outpatient spasticity clinic	Responsiveness	CTT	**Y**es

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
LASIS	Ashford, Turner-Stokes, Siegert & Slade 2013	92	stroke, TBI, MS	outpatient rehabilitation units	hypothesis testing, responsiveness	CTT	**Y**es
LASIS	Barden, Baguley, Nott & Chapparo 2014	28					
LASIS	Turner-Stokes, Baguley, De Graaff, Katrak, Davies, McCrory & Hughes 2010	90	stroke - at least 6months ago	outpatient	Concurrent validity - hypothesis testing, responsiveness	CTT	**Y**es
MAL	Atler, Malcolm & Griefe 2015	12	chronic stroke > 6months	community	hypothesis testing	CTT	nil reported
MAL	Borstad & Nichols-Larsen 2016	12	stroke > 3months	community	Construct validity - hypothesis testing	CTT	nil reported
MAL	Celik, O'Malley, Boake, Levin, Yozbatiran & Reistetter 2010	9	chronic stroke	assumed community rehabilitation	hypothesis testing,	CTT	participants mildly impaired no mention of spasticity in criteria
MAL	Harris & Eng 2006	93	chronic stroke > 1yr	community	hypothesis testing	CTT	yes measured
MAL	Mark, Woods, Mennemeier, Abbas & Taub 2006	15	chronic stroke > 6months	rehabilitation clinic	hypothesis testing	CTT	nil specifically stated
modified Frenchay Arm Test	Sunderland, Tinson, Bradley & Hewer 1989	38	acute stroke	hospital setting	hypothesis testing, responsiveness, floor and ceiling effects	CTT	yes, initially 12 participants (31%) and over the 6 months 7 (22%) not rated with MAS but rating given for abnormal resistanceto movement - mild or severe at each joint.

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
Motricity Index	Kopp, Kunkel, Flor, Platz, Rose, Mauritz, Gresser, McCulloch & Taub 1997	33	sub-acute stroke	inpatient neurological rehabilitation centre	Validity - concurrent, internal consistency	CTT	nil stated specifically
Motricity Index	Smith-Arena, Edelstein, Rabadi, 2006	39	acute stroke	acute stroke rehabunit	validity - hypothesis testing	CTT	nil specified
Motricity Index	Sunderland, Tinson, Bradley & Hewer 1989	38	acute stroke	hospital setting	hypothesis testing, validity,responsiveness, floor and ceiling effects	CTT	yes, initially 12 participants (31%) and over the 6 months 7 (22%) not rated with MAS but rating given for abnormal resistanceto movement - mild or severe at each joint.
OHS	Dennis, Wellwood & Warlow 1996	152	chronic stroke > 1year post		convergent construct validity	CTT	not reported
OHS	Gubitz, Reidy, Christian & Phillips 2012	450	acute stroke	hospital	predictive validity – hypothesis testing	CTT	nil stated
OHS	Herrman, Black, Lawrence, Szekely & Szalai 1998	3mo n = 150 1 year n = 136	acute stroke	hospital	hypothesis testing	CTT	nil stated
OHS	Pittock, Meldrum, Dhuill, Hardiman & Moroney 2003	105 at 2 weeks and 94 at 6 mo	stroke acute - subacute	community	hypothesis testing	CTT	not reported
Patient disability scale	Barden, Baguley, Nott, Chapparo 2014	28 stroke = 22 TBI =6	**M**ixed	outpatient	Concurrent validity - hypotheses testing, responsiveness	CTT	yes present
Patient disability scale	Barden, Nott, Heard, Chapparo & Baguley 2012	38	ABI	outpatient spasticity clinics - community based pts	hypothesis testing, responsiveness	CTT	**Y**es

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
RMA - UL	Meldrum, Pittock, Hardiman, Dhuill, O'Regan & Moroney 2004	114	stroke >6mo	community	predictive ability – hypothesis testing	CTT	not reported
RMA - all 3	Pittock, Meldrum, Dhuill, Hardiman & Moroney 2003	105 at 2 weeks and 94 at 6 mo	stroke acute - subacute	community	hypothesis testing	CTT	not reported
RMA 3 sections separated	Taylor, Ashburn & Ward 1994	38	stroke - acute	hospital patients	hypothesis testing	CTT	not reported
SF-36	Corrigan & Bogner 2004				hypothesis testing	CTT	not reported
SF-36	Rudick, Miller, Hass et al2007	2113	MS	community	hypothesis testing	CTT	not reported
SIS	Ali, Fulton, Quinn, Bradyon behalf of VISTA 2013	ranged 867 least	stroke >3mo	assumed hospital	concurrent validity – hypothesis testing	CTT	nil reported
SIS	Boger, Hankins & Latter 2015	74	stroke >3mo	community	hypothesis testing	CTT	not reported
SIS	Ellis, Sukal, DeMott & Dewald 2008	11	chronic stroke >2yrs	community	Concurrent validity - hypothesis testing	CTT	yes measured with MAS
SIS	O'Dell, Kim, Rivera, Fieo, Christos, Polistena, Fitzgerald, Gorga 2013	32	chronic stroke >6mo	community dwelling	hypothesis testing	CTT (IRT process for AMAT-9 only but ARAT not included in that analysis)	included if > 3 months post botulinum toxin injections as considered par of rehabilitation but specific focus with spasticity not present
10MWT	Baer & Smith 2001	164	acute stroke	hospital	predictive validity – hypothesis testing	CTT	nil stated

Characteristics of Included Studies							
Measure	Author	Sample Size	Diagnosis	Setting	Measurement Properties	Notes	Spasticity
10MWT	Bower, McGinley, Miller, Clark 2014	30	stroke > 3 mo	community	hypothesis testing, test retest reliability, measurement error	CTT	not reported
10MWT	Combs, Dugan, Passmore, Reisner, Whipker, Yingling & Curtis 2010	16	stroke - greater than 6 mo	community	hypothesis testing, measurement error	CTT - study found in SIS searchnot 10MWT	not reported
10MWT	Smith & Baer 1999	229	acute stroke	hospital	predictive validity – hypothesis testing	CTT	nil stated
10MWT	Vernon, Paterson, Bower, McGinley, Miller, Pua & Clark 2015	30	stroke > 3months	community	hypothesis testing, retest reliability and measurement error not captured as duplicate to Bower et al 2014)	CTT	not reported
10MWT	Wolf, Catlin, Gage, Gurucharri, Robertson & Stephen 1999	28 with stroke 28 without impairment	chronic stroke	community	interrater reliability, concurrent validity – hypothesis testing	CTT	nil stated
UL MAS	Horgan, Cunningham, Coakley, Walsh, O'Regan & Finn 2006	41	>11 months poststroke	rehabilitation and community	concurrent validity, clinical utility	CTT	nil stated

Reliability					
Measure	**Author**	**COSMIN Score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee Rating**
10MWT	Hirsch, Williams, Norton & Hammond 2014	fair - excellent	test re test reliability 10MWT self-selected pace ICC 0.946 (excellent) and 0.976 using Pearsonproduct moment value, 10MWT fastest pace ICC 0.961 (excellent) and PPM 0.979. SSP trial1 was faster than following 5 and FP not significantly different between trials.	1. EXCELLENT 2 N/A 3 EXCELLENT -power analysis for sample size calculated needed 10-15 as per calculation and recruited 23. 4 EXCELLENT 5. FAIR insufficient information provided to assess 6 EXCELLENT 7 GOOD as would have been stable but nil evidence provided 8. 1 min rest between Too short FAIR 9 EXCELLENT 10 EXCELLENT 11. EXCELLENT 12 n/a 13 n/a 14 n/a COSMIN UPDATE 1. Adequate 2. very Goode) 3. very good 4 very good 5 na 6 na 7 na 8. very good	plus
9HPT	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987	fair - excellent (sample not included)	test retest spearman rho correlations for all 4 tests (9HP, Frenchay Arm Test, grip strength and finger tapping) Interrater 0.83 – 0.99, testretest 0.68 – 0.90, test retest 0.75 – 0.99 (Spearman Rho). good inter-observer and test retest reliability for each test.	1. Not stated but can be deduced GOOD 2. GOOD 3. Poor n=10 4. EXCELLENT 5 Excellent 6 Excellent 7 GOOD assumable pts stable 8 EXCELLENT 9 GOOD assumable testconditions similar 10 many individual studies reported so detaillacking to really appraise age, generalisability etc FAIR 11. FAIR no ICC calculated	? No ICC or weighted kappa reported spearman rho only
ARAT	Yozbatiran, Der-Yeghiain & Cramer 2008	poor - excellent sample excluded good	interrater reliability total scores (individual Ax components scores listed in article) 2 examiners; spearman's r assessor 1 scores = to assessor 2: .96, spearman's p assessor 1 scores = assessor 2: <0.0001, ICC assessor 1 = assessor 2: 0.9986. Intrarater reliability assessor 1 scores equal to assessor 2, spearman's r: 0.99, spearman's p: <0.0001, ICC:0.99. ARAT is capable of detecting change in range of clinically significant values. for ARATthis is -2.4 to 2.8 a range less than minimum clinically important difference defined by van derLee as 5.7 points increasing confidence that clinically significant changes are not a result of measurement error.	1. excellent 2. good 3. poor <30, 4. excellent 5. excellent 6. excellent 7. good 8. excellent 9. good 10. excellent 11. EXCELLENT 12. excellent weighted Kappa calculated 13 N/A 14 N/A COSMIN update 1. adequate 2 very good 3. adequate 4. adequate 5. na 6 na 7 na 8 very good ADEQUATE and +	plus

Reliability					
Measure	**Author**	**COSMIN Score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee Rating**
ArmA	Ashford, Turner-Stokes, Siegert &	adequate	retest reliability - weighted kappa coefficients forsubscales were passive: 0.90; confidence interval 0.68 - 1.12, active 0.93 (CI 0.71 - 1.15)	1 adequate assumable patients stable as repeated one day apart 2very good 3 adequate assumable test conditions the same 4 n/a 5 very good weighted Kappa 6 very good 7 very good 'quadratic' 8 very good	plus
BI	Collin, Wade, Daves Horne1988	inadequate	Range of disagreement between self-report, nurse, 2 skilled observers was from 1 to 9 points. 72% of patients had agreement betweennurse and two skilled observers of 2 points or less, 92% of patients 4 points or less. agreement was lower for transfers, feeding, dressing, grooming and toileting. found asking a nurse or relative was as reliable as testing but quicker. Difference of 4/20 points likely to represent genuine change	COSMIN UPDATE 1 doubtful unclear if pts stable as ratings took place over three days in acute admission 2 very good 3 adequate assumable test conditions similar - as asking pt etc then assume same methods 4. n/a 5 ordinal inadequate no kappa calculated only percentage agreement. Poor; Patients may have changed from baseline o 72 hrs post discharge for OT assessment; additionally kendalls coefficient not as favorable. stat. Methodological flaw - authors state instructions slightly modified from study to publication therefore we don't know how this might have further affected results.	?
BI(C&W)	Green et al 2001	inadequate	No kappa for Bowels and toilet due to high prevalence of the item. Bladder had low agreement -0.09 and walking was 0. Appearedthat the BI had some good test-retest reliabilityoverall but this was only for % agreement BI agreement >75%	COSMIN UPDATE 1. adequate 2 very good 3 adequate 4. n/a 5 only % agreement for total score, weighted kappa for individual questions inadequate 6 inadequate as per above 7 good	inadequate ?
CMA	Gowland, Stratford, Ward, Moreland, Torresin, Van Hullenar, Sanford, Barreca, Vanspall & Plews 1993	fair - excellent impairment - intra and inter rater very good, disability inter and test-retest adequate	impairment inventory shoulder pain intra ICC0.96 95% CI 0.92-0.98 inter ICC 0.95 95%CI 0.91-.098 postural control Intra ICC 0.96 95% CI 0.93-0.98 inter ICC 0.92 95%CI 0.84-0.96, Arm intra ICC 0.95 95%CI 0.89-0.97 inter ICC 0.88 95% CI 0.76 - 0.94 Hand intra ICC 0.93 95% CI 0.85-0.96 Inter ICC 0.93 95%CI 0.84-0.96 leg intra ICC 0.98 95%CI 0.96-0.99 inter ICC 0.8595%CI 0.73-0.93 Foot intra ICC 0.98 95%CI 0.95-0.99 inter ICC 0.96 95%CI 0.94-0.98 total score intra ICC 0.98 95% CI 0.95-0.99 Inter ICC 95% CI 0.94-0.98 DISABILITY Inventory; gross	1. yes 3 did not have dc data EXCELLENT 2. GOOD 3. FAIR sample size 4. EXCELLENT 5. GOOD 6. EXCELLENT 7 GOOD 8 EXCELLENT 9 GOOD assumable that conditions were similar particularly for impairment as was videotaped but activity as repeated and nil specific information to support similarities given 10. EXCELLENT 11 excellent COSMIN UPDATE INTRA and INTERRATER IMPAIRMENT INVENTORY 1. very good 2 very good 3 very good 4 very good 8 very good INTER AND TEST RETSET OF DISABILITY Adequate as assumable test conditions similar and patients stable all else very good	plus

Reliability					
Measure	**Author**	**COSMIN Score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee Rating**
			motor function inter ICC 0.98 95% CI 0.97-0.99 test retest ICC 0.96 95%CI 0.93-0.98, walking inter ICC 0.98 95%CI 0.95-0.99 test- retest ICC 0.98 95%CI 0.96-0.99 total scores inter ICC 0.99 95%CI 0.98-1.00 test-retest ICC 95%CI 0.95-0.99. summary - the total scores ranged from 0.97 to 0.99 (minimum set apriori at 0.90) and for the dimensions and indexes from 0.85 to 0.98 (minimum set a priori at 0.80), the lower 95%CI of the ICCs for the total scores ranged from 0.94 to 0.98 (lowest acceptable value set a priori at 0.80). all valueshigher than apriori		
DAS interrater	Brashear, Zafonte, Corcoran, Galvez-Jimenez, Gracies, Gordon, Mcafee,	adequate	interrater reliability of DAS for all parameters (hygiene, dressing, limb position and pain) therewas good agreement among raters at evaluation1 and 2. Mean of evaluations 1 and 2 Kendall W and (95% CI, P) hygiene 0.626 (0.297-1.000 P<0.001), dressing 0.494 (0.234-1.00 P<0.001) Limb position 0.557 (0.264-1.00 P<0.001) pain 0.772 (0.366-1.00 P<0.001).	1. EXCELLENT 2. EXCELLENT 3 N = 10 raters (but 40 ratings) 4. EXCELLENT 5. EXCELLENT 6. EXCELLENT 7. GOOD 8 EXCELLENT 9. GOOD 10. EXCELLENT 11. N/a 12 EXCELLENT 13 EXCELLENT 14 GOOD COSMIN UPDATE 1adequate 2. Very good 3. adequate 4. N/A 5. very good 6 very good, same as Kendall W 7 adequate	minus (50 per cent above and 50 per cent below 0.70 weighted kappa)
DAS intrarater	Brashear, Zafonte, Corcoran, Galvez-Jimenez, Gracies, Gordon,	good - excellent	Intrarater reliability 78% (31/40) of the evaluations indicated either good (weighted k ≥0.4) or excellent (weighted k ≥ 0.75) whencomparing trained health professionals	1. EXCELLENT 2. EXCELLENT 3 N = 10 raters (but 40 ratings so what does n apply to???) 4. EXCELLENT 5. EXCELLENT 6. EXCELLENT 7. GOOD 8 EXCELLENT 9. GOOD 10. EXCELLENT 11. N/a 12 EXCELLENT 13 EXCELLENT 14 GOOD COSMIN UPDATE 1. adequate 2 very good 3 adequate 4 na 5 very good 6 very good 7 adequate	minus (50 per cent above and 50 per cent below 0.70 weighted kappa)
EQ-5D	Dorman, Slattery, Farrell, Dennis & Sandercoc	doubtful	pt all K greater than 0.70 except usual activities 0.66 (.54 to .78), and greatest .85 (.72 to .94) (mobility) in comparison to proxy .31 (.00 to .66) to .63 (.50 to .77), proxy nil greater than 0.70	COSMIN UPDATE 1adequate 2 very good 3 very good 4 n/a 5very good 6 doubtful 7n/a 8 very good	proxy - patient +

Reliability					
Measure	**Author**	**COSMIN Score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee Rating**
EQ-5D	Fisk, Brown,	adequate	ICC 0.83	COSMIN UPDATE 1. adequate 2 very good 3 adequate 4 adequate 5 n/a 6 n/a 7 n/a 8 very good	plus
FIM	Hamilton & Granger	fair (doubtful)	inter rater FIM ICC = 0.96 motor and 0.91 cognitive	doubtful	plus
FIM	Sharrack, Hughes,		Inter-rater FIM total ICC=0.99	1. n/a 2 very good 3. very good 4 very good 5 n/a 6 n/a 7 n/a 8 very good	plus
FIM	Stineman et al 1996	good	18-item 0.94 – 0.97, FIMm 0.93-0.97m FIMc 0.93 – 0.94	COSMIN UPDATE 1. Adequate	plus
Frenchay Arm Test	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987	fair - excellent	for all 4 tests (9HP, Frenchay Arm Test, grip strength and finger tapping) tested on 3 occasions by 2 different raters blinded to previous results; 0.83-0.99 between observer 1/test 1 and observer 2/ test 2; 0.68-0.90 between observer 1/test 1 and observer 1/test3; 0.75-0.99 between observer 2/test 2 and observer 1 / test 3. all statistically significant, strong associations (p> 0.025 - p < 0.001) - indicating good inter-observer and test retest reliability for each test.	1. Not stated but can be deduced GOOD 2. GOOD 3. Poor n=10 4. EXCELLENT 5 Excellent 6 Excellent 7 GOOD assumable pts stable 8 EXCELLENT 9 GOOD assumable test conditions similar 10 many individual studies reported so detail lacking to really appraise age, generalisability etc FAIR 11 test re test spearman rho calculated no ICC nor kappa 12 n/a FAIR 13 n/a 14 n/a COSMIN UPDATE 1. adequate 2 verygood 3 adequate 4 doubtful 5 na 6 na 7 na 8.doubtful	? No ICC or weighted kappa reported spearman rho only
GAS	Bovend'Eerdt, Dawes, Izadi, Wade 2011	good (excluding sample)	mixed model ICC 0.478 (fair to good reliability)between therapist and the masked assessor scoring procedures.	1. EXCELLENT 2. GOOD, 3 POOR sample size n=29, 4 EXCELLENT 5 EXCELLENT 6 EXCELLENT 7 FAIR 8 EXCELLENT 9 GOOD assumed but nil data re same 10 EXCELLENT 11 EXCELLENT 12 N/A 13 N/A 14 N/A COSMIN UPDATE 1. doubtful 2 very good 3 adequate 4 very good 5 na 6 na 7 na 8 very good	minus
GAS	Joyce, Rockwood & Mate-Kole 1994	fair	inter rater reliability of GA scaling was 0.92 onadmission and 0.94 at discharge.	1. GOOD 2. n/a 3. poor sample size 4 excellent 5. good 6. fair time interval not stated 7 good 8 fair 9 good 10 fair- minimal methodological data to assess what specifically happened in regards to reliability 11 n/a 12 ? 13 COSMIN UPDATE 1. adequate 2 doubtful (ax on admission and dc study reports LOS so assume that was GAS program length 3 adequate 4 5 6 7 as assume ICC but not stated 8 doubtful	plus As appears to beassociation not ICC and no description given to clarify

Reliability					
Measure	**Author**	**COSMIN Score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee Rating**
MAL-28	Uswatte, Taub, Morris, Light & Thompson	fair - excellent	pt MAL-28 QOM and AOU from delayed-treatment participants were reliable QOM: ICC 0.82 AOU ICC 0.79. Caregiver MAL QOM ICC 0.72, AOU ICC 0.66 - summary pt MAL morereliable than caregiver MAL	1. GOOD 2. GOOD 3 EXCELLENT sample size 4 EXCELLENT 5 GOOD 6 EXCELLENT 7 GOOD 8 EXCELLENT 9 GOOD 10 FAIR minimal details provided 11 type 3,1 EXCELLENT COSMIN UPDATE 1. adequate 2. very good 3. adequate 4. very good 5 na 6 na 7 na	plus - pt reported + caregiver -
MI	Collin & Wade 1990	poor - excellent	interrater reliability MI arm Spearman rho 0.88(p< 0.001) MI leg 0.87 (p< 0.001) MI side 0.88(p<0.001)	1. n = 20 original n for multi property study 36 EXCELLENT 2. GOOD 3 POOR 4 EXCELLENT 5 EXCELLENT 6 EXCELLENT 7 good 8 EXCELLENT 9 GOOD 10 EXCELLENT 11 FAIR - computes as continuous scores 12 n/a 13 n/a 14 GOOD not described COSMIN UPDATE 1. adequate 2. very good 3. adequate 4doubtful 5 na 6 na 7na 8very good	COSMIN UPDATE
SF-36	Dorman, Slattery, Farrell, Dennis & Sandercock 1998	adequate	Test retest reliability patient reported ICC = 0.30to 0.81, proxy 0.24 to 0.76, both 0.28 to 0.80.	COSMIN UPDATE 1. adequate 2 very good 3 very good 4 adequate 5 n/a 6 n/a 7 n/a 8 very good	proxy - 5/8 only met criteria , + except mental health
SF-36	Murell, Kenealy, Beaumont & Lintern 1999	doubtful	N=22 for test/re-test, so poor. Five/eight SF-36 scales reached the lower reliability criteriaset (0.6-0.69), although only physical functioning, vitality, and general health domains produced consistency to the traditionally required standard of 0.70	COSMIN UPDATE 1. doubtful 2 very good 3 adequate 4 doubtful Pearson product moment tests 5 n/a 6 n/a 7 n/a 8 very good	?
SF-36	Vickery, Hays, Genovese,	adequate	ICC scores ranged from 0.64 (social function) 0.66 (Role limitations - physical) to 0.96(physical function). 2/8 less than 0.7	COSMIN UPDATE 1. very good 2 very good 3 very good 4 adequate 5 n/a 6 n/a 7 n/a 8 very good	plus (excl 2 subscales)

Reliability					
Measure	**Author**	**COSMIN Score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee Rating**
SIS	Duncan, Lai, Tyler, Perera, Reker & Studenski 2002	good - excellent	proxies rated patients as more impaired in 5 domains, biases between patient and proxies low (effect sizes -0.1 to 0.4). Strength of agreement ranged from ICC 0.50 to 0.83 - moderate to excellent. Best agreements were for observable physical behaviours and worst agreements on more subjective domains e.g. memory and thinking, communication, emotion and strength. Pts with more sever Rankin scale (4-5) the difference in patient-proxy reporting forADL/IADL and SIS-16 were significantly larger than in an individual with less severe stroke. stroke severity did not affect other domains. regression models did not systematically identifyany pt or proxy factors that affected the observed differences.	1. excellent 2 n/a 3 excellent 4 excellent 5 excellent 6 excellent 7 good 8 excellent 9 good 10 excellent 11 excellent COSMIN UPDATE: 1. Adequate 2 very good 3 adequate 4 very good 5 n/a 6 n/a 7 n/a 8very good	plus COSMIN UPDATE (?) ICC 0.5 to 0.83 ? 5/8 less than 0.70
SIS	Duncan, Reker, Kwon, Lai, Studenski, Perera, Alfrey & Marquez 2005	poor - excellent	ICC mail (all responders) ranged 0.62 - 0.94 when exclude outlier lowest is 0.75. pt who reported their health did not change at all withinthe week ICC ranged 0.41 (when outlier removed 0.62) to 0.97. Telephone all 0.91 to 0.98 pts who reported their health did not change at all 0.77 to 0.98. summary test retestin mail group (0.77-0.99) good to excellent and for telephone group (0.90 - 0.99) excellent	1. EXCELLENT 2 FAIR not clear how handled 3 POOR sample size 4. EXCELLENT 5. GOOD assumable independent 6. EXCELLENT one week interval 7. GOOD - they did screen with a global question have they changed but not using a validated outcome measure 8. Excellent time interval appropriate 9. EXCELLENT 10. FAIR primarily male sample 11. GOOD model not described. 12 n/a 13 n/a 14 n/aCOSMIN UPDATE : 1 adequate 2 very good 3 very good 4 adequate 5 n/a 6 n/a 7 n/a 8 doubtful due to sample	plus
UL MAS	Loewen & Anderson 1988 (b) intra rater	poor - excellent	1. not stated but assumed 0 GOOD 2. n/a 3. EXCELLENT 4. EXCELLENT 5 GOOD 6 EXCELLENT 7 n/a - video reviewed again 8 EXCELLENT 9 excellent 10 FAIR 11 n/a 12 Kendall Tau?? Correlation coefficient 13 neitherso assign poor Kappa not calculated nor percentage agreement FAIR 14 GOOD COSMIN UPDATED 1 very good 2 very good 3very good 4 n/a 5	Intrarater: Kendall's rank order correlation (Tau) Item 6: 0.74 - 1.00, Item 7: 1.00, Item 8 1.0(Kendall's Tau must be greater than or equal to 0.69 to be significant at the 0.05 level. MORE detail re full MAS:MAS interrater: 80% of the 637 comparisons were in the excellent range, kappa 0.75 - 1.00 with mean K ranging from 0.73 - 0.96. Spearman rank-order correlation coefficients for total mMAS was 0.83 to 1.00 with a median of 0.97. individual items of each scale for mMAS ranged from 72% (side lying) to 100% (hand movements, advanced	Kendalls rank order coefficient was used

Reliability					
Measure	**Author**	**COSMIN Score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee Rating**
			very good 6 chase Kendall Tau - doubtful	handactivities) with mean kappa values ranging from 0.56 to 1.00.Intrarater: total score per therapist mMAS 85% of 98 kappa values in excellent agreement range, spearman rank-order correlations coefficients was 0.81 to 1.00 with a median of Intrarater for individual item of each scale, Kendalls Tauonly 5 of the 114 (4%) were non-significant (less than or equalto 0.69). Both mMAS (full MAS minus clonus section) and BI were found to have inter and intra rater reliability and are a measure of motor recovery and functional independence. Both performed similarly.	
UL MAS	Loewen & Anderson 1988(a) inter rater	fair - excellent	1. not stated but assumed 0 GOOD 2. n/a 3. EXCELLENT 4. EXCELLENT 5 GOOD 6 EXCELLENT 7 n/a - video reviewed again 8 EXCELLENT 9 excellent 10 FAIR 11 n/a 12 EXCELLENT 13 FAIR unweighted kappa 14 GOOD COSMIN UPDATE 1 very good 2 very good 3 very good 4 n/a 5 very good 6 doubtful 7 adequate 8 doubtful	concise: Interrater reliability: Percentage agreement between therapists Item 6: 96.2, Kappa 0.93, Item 7: 100, Kappa 1.0, Item 8: 100, Kappa 1.0.m	plus as reference given for useof unweighted kappa
UL-MAS	Carr, Shephard, Nordholm & Lynne 1985 (a) inter rater	poor - excellent (not including sample)	2. 20 Raters assessed 5 patients and the following is the percentage of agreement of theraters with the initial rater (criterion rating) Item6: mean 88 (70-100), item 7: mean 95 (85-100)item 8 mean 88 (40-100). Recommend clinicians should become familiar with the itemson at least 6 patients before formally using in clinical practice.	1. EXCELLENT as per detailed in table nil missing 2. n/a 3. FAIR n=40 observations 4. EXCELLENT 5. EXCELLENT 6. FAIR time interval NOT stated 7. GOOD assumable pts were stable 8. EXCELLENT 9. GOOD assumable test conditions 10. EXCELENT 11. n/a 12. Poor % agreement only 13. poor % agreement only 14 good	?
UL-MAS	Carr, Shephard, Nordholm & Lynne 1985 test retest	fair - excellent	whole MAS no UL-MAS 0.87 - 1.00 with an average of 0.98 Inter rater agreement mean %agreement item 6 88, item 7 95, item 8 88	1. EXCELLENT 2. GOOD not described but clear excluded 3.Poor sample size n=14 4. EXCELLENT 5. GOOD assumableindependent 6. EXCELLENT 7. GOOD 8. EXCELLENT - timeinterval 4 weeks but given chronic stage this is considered appropriate 9 GOOD 10 EXCELLENT 11 FAIR Pearson product moment correlation (not ideal) COSMIN UPDATE 1. adequate 2 very good 3 adequate 4, 5, 6 inadequate percentage agreement calculated	inadequate

Reliability					
Measure	**Author**	**COSMIN Score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee Rating**
Excluded at COSMIN update					
10MWT	Bower, McGinley, Miller, Clark 2014	good - excellent (sample Fair	Test retest reliability ICC (85% CI) 0.97 (0.93, 0.98)	1. EXCELLENT 2 GOOD 3. sample size (30) but power calculation given to support sample EXCELLENT, 4. EXCELLENT 5. EXCELLENT 6. EXCELLENT 7 EXCELLENT -self reported global rating of change was completed with nil participant recording any change 8 EXCELLENT 9 EXCELLENT - study protocol implemented re ordering of tests 10 EXCELENT 11EXCELLENT	plus
10MWT	Wolf, Catlin, Gage, Gurucharri , Robertson	fair - excellent	1. GOOD 2. N/A 3. GOOD 4. EXCELLENT 5 FAIR 6 n/a 7. GOOD 8 n/a 9. GOOD 10 FAIR - Trained raters, convenience high level of function participants 11 EXCELLENT 12 n/a 13 n/a 14 n/a	ICC without impairment 0.980 with stroke 0.998. 2 of 4 investigators randomly assigned to participants and both scored each measure on separate scoring sheets.	plus howeveras each subgroup less than n=50 is a ?
ARAT	McDonnell ,Hillier, Ridding & Miles 2006	poor - excellent sample excluded - fair	all measures were reliable between testing sessions for the unaffected hand (mean ICC(3,1) = 0.86+/- 0.007 range 0.78-0.99) and for theaffected hand (excluding time shift) (mean ICC(3,1) = 0.93 +/- 0.05, range 0.83-0.99).	1. not stated but can be deduced GOOD 2 not stated but canbe deduced GOOD 3 POOR n=17 4 EXCELLENT 5 FAIR noinfo provided 6 EXCELLENT 7 FAIR 8 EXCELLENT 9 FAIR 10 FAIR due to lack of detail of methods stated in paper 11 GOOD 12 n/a 13 n/a 14 n/a	plus
ARAT	Page, Hade & Persch 2015	fair - excellent (good sample excl)	intrarater reliability ICC 0.99 (95%CI 0.98, 0.99)compared to w/h UE FM 0.95 (95%CI 0.92, 0.99)	1. GOOD nil evidence 2 N/A 3. FAIR sample n=32 4. EXCELLENT 5. GOOD assumable independent 6. EXCELLENT 7. GOOD assumable stable given chronicity and statement verified by medical records and clinicians 8. EXCELLENT 9. GOOD 10. EXCELLENT 11. EXCELLENT	plus
ARAT	Page, Levine, Hade 2012	poor-excellent - sample excluded fair	Intrarater reliability using ICC ARAT pre test 1(mean +-SD) 22.4 +/- 15.1, pre test 2 21.9 +/- 16.1 ICC: 0.71 (0.53-0.89 95%CI) minimal detectable change 22.54. Finding that UE FMand w/h UE FM had higher intrarater reliabilitythan ARAT.	1. GOOD % missing items not described but can be deduced from paper 2. GOOD 3. POOR 4. EXCELLENT 5. GOD assumable independent administrations. 6 EXCELLENT 7 EXCELLENT 8 EXCELLENT 9 FAIR 10 EXCELLENT (appears to be typo as cut off for MMSE greater than or equal to 70). 11. GOOD 12 N/A 13 N/A 14 N/A	plus

Reliability					
Measure	**Author**	**COSMIN Score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee Rating**
BI	Loewen & Anderson 1988	fair - excellent	1. not stated but assumed 0 GOOD 2. n/a 3.EXCELLENT 4. EXCELLENT 5 GOOD 6 EXCELLENT 7 n/a video reviewed again 8 EXCELLENT 9 excellent 10 FAIR 11 n/a 12 EXCELLENT 13 FAIR unweighted kappa 14GOOD	concise: Interrater reliability: Agreement between 5 therapistsusing BI total score to rate 7 patients (10 comparisons per patient totaling 70 comparisons, 51% of Kappa values in excellent range K 0.75 - 1.00) 49% in good-fair (K 0.41-0.74) 0% in poor. Spearman rank-order correlations for BI was 0.91to 1.00 with a median of 0.96. Interrater reliable between 5 therapists and individual BI items to rate 7 pts. percentages ofagreement ranged from 69% (w/c transfer to/from bed) to 100% (bowel control, bladder control, feeding, and ascendingand descending stairs), mean Kappa ranged from 0.47 - 1.00.INTRARATER: BI total score per therapist (n=5) 83% of the 35 Kappa values were in excellent agreement range, spearman rank-order correlation 0.95 to 1.00. INTRARATER per individual item: Kendalls rank order (Tau) ranged from 0.42 (item 3) to 1.00, majority 1.0 according to Tau scale only2 (4%) of the 50 values were non-significant. (Kendall's Tau must be greater than or equal to 0.69 to be significant at the 0.05 level. Both full MAS minus clonus section) and BI were found to have inter and intra rater reliability and are a measureof motor recovery and functional independence. Both performed similarly.	plus but is unweighted

Internal Consistency					
Outcome measure	**Author**	**Findings**	**COSMIN**	**Study results**	**Terwee Rating**
ArmA	Ashford, Turner-Stokes, Siegert & Slade 2013	1 yes reflective 1 very good 2 very good	very good	Cronbach's alpha for passive subscale 0.85, for active subscale 0.96	plus
FIM	Stineman et al 1996	structural validity good, internal consistency good; good for structural validity as technically PCA analysis is viewed separately to CFA or EFA, good rating for internal consistency as it links to factor analysis	adequate	18-item 0.94 – 0.97, FIMm 0.93-0.97m FIMc 0.93 – 0.94	plus
MAL-28	Uswatte, Taub, Morris, Light & Tompson 2006	1 yes reflective model 2. not stated GOOD 3. N/A 4. EXCELLENT 5 EXCELLENT 6.excellent 7 no POOR 8 FAIR very challenging to determine which statistics related to which version of MAL 9 EXCELLENT COSMIN UPDATE 1. very good 2. very good 3, na 4 na	poor – excellent	MAL-28 had high internal consistency, Cronbach's α 0.94 for both QOM and AOU, for caregiver high internal consistency also Cronbach's α 0.95 for both AOU and QOM scales	plus
Motricity Index	Bohannon 1999	1 yes 2 Good 3 n/a 4 poor sample size n=10 5. poor 6 poor 7 assumed due to wording excellent 8. fair - very minimal information regarding methodology to identify potential flaws 9 excellent COSMIN UPDATE 1. doubtful 2. very good 3 na 4 na - no evidence that scale is unidimensional in this study	poor - excellent (sample excluded)	Cronbach's alpha of the MI scores (UE) were 0.968	COSMIN UPDATE, no evidence on structural validity within this study but ? Elsewhere so +
SF-36	Anderson 1996	COSMIN UPDATE unidimensionality not checked doubtful 2) very good 3	doubtful	For all eight scales except vitality, internal consistency by Cronbach's alpha satisfied Nunally's criterion of 0.7 (results)	plus
SF-36	Dorman, Slattery, Farrell, Dennis & Sandercock 1998	COSMIN UPDATE 1 very good (unidimensional screen from other papers) 2. very good 3 n/a 4 n/a	very good	Cronbach's alpha for each domain >or equal to 0.80	plus

Internal Consistency					
Outcome measure	**Author**	**Findings**	**COSMIN**	**Study results**	**Terwee Rating**
SF-36	Freeman, Hobart, Langdon & Thompson 2000	1. very good (through intercorrelations within dimensions) 2. very good	very good	Internal consistence for each of 8 dimensions high with alpha coefficients ranging between 0.77 to 0.94	plus
SF-36	Hagen, Bugge & Alexander 2003	COSMIN UPDATE 1.very good (item total correlations occurred accepting 0.4 (checking unidimensionality) 2. very good 3 n/a 4 n/a	very good	alpha 0.7 excluding general health at 3 mo 0.665	plus
SF-36	Vickery, Hays, Genovese, Myers & Ellison 1997	COSMIN UPDATE 1.very good (item total correlations occurred accepting 0.4 2. very good 3 n/a 4 n/a	COSMIN UPDATE 1. very good	Cronbach's alpha ranged from 0.79 (social function) to 0.96 (physical function)	plus
SF-36	Guilfoyle 2010	1. very good 2 very good 3 very good unidimensionality was checked		Domain scores showed excellent internal consistency with alpha values all exceeding 0.8.	plus
SF-36	Hobart et al 2002	1 very good 2 very good	very good	Alpha coefficients ranged from 0.68 (general health) to 0.90 (physical functioning) indicating that most scales generated reliable scores (Table 4). One scale (GH) failed to satisfy the criterion of 0.7 and 2 scales (GH and SF) just failed to satisfy the more stringent criterion of 0.8.	plus
SIS	Duncan, Reker, Kwon, Lai, Studenski, Perera, Alfrey & Marquez 2005	1. yes 2. EXCELLENT 3 FAIR not clear how handled 4. POOR sample size 5 POOR 6. not ax so POOR 7. EXCELLENT 8. FAIR sample primarily male due to veterans 9 EXCELLENT 10 n/a 11 n/a` COSMIN UPDATE 1. very good 2 very good 3 n/a 4 n/a	poor - excellent	Cronbach's alpha computed for each domain seven of eight had alpha values greater than or equal to 0.89 for both mail and telephone modes. The emotion domain had alpha of 0.76 mail and 0 83 telephone	plus

Internal Consistency					
Outcome measure	**Author**	**Findings**	**COSMIN**	**Study results**	**Terwee Rating**
SIS	Jenkinson, Fitzpatrick, Crocker & Peters 2013	1. yes 2. excellent 51.66% missing data 3 excluded excellent Bias 4good sample size 73 5 excellent higher order factor analysis completed 6. excellent when calculating # items as 8 (i.e. number of domains). 7. excellent 8. Fair percentage of missing data 9. excellent COSMIN UPDATE 1. very good 2 very good 3 na 4 na	fair - excellent	Cronbach's alpha for each domain of SIS ranged from 0.86 - 0.96 suggesting possible item redundancy.	plus with a note that this excludes memory which was greater than 0.95
UL MAS	Johnson & Selfe 2004	1. yes 2. GOOD 3. GOOD 4. POOR 5. EXCELLENT 6 N/A 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 N/A 11 N/A COSMIN UPDATE 1. very good 2 very good	good - excellent	Cronbach's alpha for items 1- 8 0.953 for mRMI 0.949. If sitting balance was removed (which was the lowest item- total correlation 0.545) alpha increased to 0.961. item 6: 0.893, 7 :0.889, 8: 0.854	? plus
UL MAS	Lannin 2004	1. yes 2. EXCELLENT 3 n/a 4 GOOD as reference for needed sample to complete power analysis given 5. EXCELLENT factor analysis performed 6. excellent 7. excellent as states only one dimension underlying the UL-MAS 8. FAIR - not a very representable sample 9. n/a 10 EXCELLENT COSMIN UPDATE 1 very good 2very good 3	fair - excellent very good	internal consistency for the three items of the UL-MAS subscale provided an acceptable Cronbach's alpha of 0.83	plus
UL MAS	Miller, Slade, Pallant, Galea 2010	1. yes 2. GOOD - not described 3. GOOD 4. 80 individuals 140 observation excellent 5. GOOD 6. excellent 7. excellent 8. fair - no detail re blinding of assessors Range of time between two sites when assessing n/a 10 n/a 11 excellent COSMIN UPDATE VERY GOOD	fair - excellent	UL -MAS is a unidimensional scale that measures a single construct - UL motor recovery. With 17 items (removal of wrist deviation) principal component analysis confirmed the uni-dimensionality of the scale - differences in scores between positive and negative loading test items resulted in significant t-tests ($p<0.05$) for only 2.85% of participants falling below the acceptable guideline of 5%. Ul MAS 17 items was shown to have adequate fit to the Rasch model with good internal consistency (Person Separation Index =0.95, Cronbach Alpha =0.93) as compared to original 18 item	plus

Internal Consistency					
Outcome measure	**Author**	**Findings**	**COSMIN**	**Study results**	**Terwee Rating**
Excluded at COSMIN update					
Motricity Index	Kopp, Kunkel, Flor, Platz, Rose, Mauritz, Gresser, McCulloch & Taub 1997	1. yes reflective model 2 GOOD 3 GOOD 4 FAIR and assumed whole sample was used as not clearly stated 5 POOR 6 POOR 7 POOR 8 POOR 9 EXCELLENT Cronbach's alpha calculated	poor - excellent	internal consistency reported as Cronbach's alpha 0.62.	minus

Measurement Error						
Outcome Measure	**Author**	**COSMIN score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Notes**	**Terwee rating**
9HPT	Schwid, Goodman, McDermott, Bever & Cook 2002	poor-excellent	a change of 20% on the 9HPT can be considered the threshold that reliably indicates a true change in function = SDC smallest detectable change -> change beyond measurement error	1. not directly stated but can deduce 0 GOOD 2. GOOD not stated but can deduce 3. POOR n=27 4 EXCELLENT 5 EXCELLENT 6. EXCELLENT 7. GOOD 8. EXCELLENT 9. EXCELLENT 10. EXCELLENT 11 EXCELLENT 12 EXCELLENT (SDA, LoA calculated from what I can ascertain) COSMIN UPDATE 1. adequate 2 very good 3 very good 4 very good 5 na		plus, MIC not defined but argument that agreement is acceptable
BI(C&W)	Green et al 2001	adequate	Mean difference 0.4 (95%CI 0.01-0.90)	was assumable that patients were stable. Given there was only a week between measurements I thought that was pretty reasonable.		plus
GAS	Bovend'Eerdt, Dawes, Izadi & Wade 2011	good excluding sample	Limits of agreement between the measurements by the two assessors - 1.52 +/- 24.54.	1. EXCELLENT 2. GOOD 3 POOR 4. EXCELLENT 5 EXCELLENT 6 EXCELLENT 7. GOOD 8. EXCELLENT 9 GOOD 10. EXCELLENT, EXCELLENT COSMIN UPDATE 1. adequate 2 very good 3 adequate 4 very Good		?
MAL	Chen, Wolf, Zhang, Thompson & Winstein 2012	fair - excellent	MDC MAL amount of use 16.8% quality 15.3%. Actual amount of use test (AAUT) amount 24.2%, quality 14.4%. Difference in measurement variability between AAUT and MAL was much greater for the amount of use than the quality of movement subscale. When investigating therapeutic changes post CIMT only those measured by the MAL exceeded MDC.	1. GOOD 2 GOOD 3. excellent 4. excellent 5. GOOD 6 FAIR 7 GOOD 8 n/a as time interval not stated 9 GOOD 10 FAIR missing data to determine methodology and administration of MAL Raters trained and blinded 11. excellent COSMIN UPDATE 1. adequate 2. excellent 3 adequate 4 very good	? Minimal important change not defined	
excluded at COSMIN review						
10MWT	Bower, McGinley, Miller, Clark 2014	good - excellent	SEM 0.06, MDC 0.16 (16.7%)	1. EXCELLENT 2 GOOD 3 FAIR sample 4 EXELLENT 5 EXCELLENT 6 EXCELLENT oneweek 7. EXCELLENT self-rated global rating ofchange completed with nil participants identifying any perceived change 8 EXCELLENT 9 EXCELLENT study protocol applied in administration of tests etc 10 EXCELLENT 11 EXCELLENT.		plus as high ICC, no argument that agreement is not acceptable

Measurement Error						
Outcome Measure	**Author**	**COSMIN score**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Notes**	**Terwee rating**
10MWT comfortable and fast pace	Combs, Dugan, Passmore, Reisner, Whipker, Yingling & Curtis 2010	poor - excellent sample excluded fair	minimal detectable change for both speeds were calculated using the 3 pretest trials. CWT MDC was 0.08m/s and was exceeded by 9 of 16 (56%) at posttest and 10 of 16 (66%) at retention. FWT MDC was 0.16m/s and achieved by 3 of 16 (19%) at post test and 7 of 16 (44%) at retention.	1. yes EXCELLENT 2 GOOD 3 POOR n=16 4 EXCELLENT 5 FAIR - doubtful administrationswere independent, nil information to support otherwise given 6 EXCELLENT 8 EXCELLENT9 FAIR - unclear if test conditions similar 10 FAIR - rater not blinded 11. EXCELLENT SDC or referred to as minimal clinically meaningful change.		plus
ARAT	Page, Levine & Hade 2012	fair with sample excluded	SDC 22.54	1. GOOD % missing items not described but can be deduced from paper 2. GOOD 3. POORn=29 4. EXCELLENT 5. GOOD doubtful as the rater was blinded to the purpose of the study butlikely knew first score before administering second. 6 EXCELLENT 7 EXCELLENT 8 EXCELLENT 9 GOOD 10 EXCELLENT (appears to be typo as cut off for MMSE greaterthan or equal to 70). 11. EXCELLENT	? No MIC defined	
SIS - ADL/IADL, mobility, participatio n, stroke recovery scale	Combs, Dugan, Passmore, Reisner, Whipker, Yingling & Curtis 2010	poor - excellent	Participants meeting criteria for clinically meaningful change (established as 10 to 15 points) on SIS ADL/IADL (44%, 38%), SIS mobility (31%, 38%) SIS participation (50%, 44%) at post test (post 8 wk Ix) and retention (6mo post Ix). Clinically meaningful change on SIS stroke recovery scale achieved by 10 of 16 (63%) post test (69%) retention.	1. yes EXCELLENT 2 GOOD 3 POOR n=16 4 EXCELLENT 5 GOOD - assumable administrations were independent, nil information to support otherwise given 6 EXCELLENT 8 EXCELLENT 9 GOOD - assumable test conditions similar 10 FAIR - rater not blinded 11. EXCELLENT SDC or referred to as minimal clinically meaningful change.		

Content Validity					
Measure	**Author**	**COSMIN**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee rating**
AQoL	Hawthorne, Richardson & Osborne 1999	inadequate		Box 1a designed to measure HRQoL in regards to illness, independent living, social relationships, physical senses, psychological well-being VERY GOOD 2. very good - theoreticalmodel 3 very good general population 4 doubtful not clearly described but assumed general 5 very good 6 very good 7 very good 8 very good 9 doubtful 10 doubtful - not stated 11 12 14very good 15 very good 16 does not appear patients were askedabout comprehensiveness or comprehensibility rather trialled the test	
ArmA	Ashford, Slade & Turner-Stokes 2013	adequate +, +, +	1. very good 2 very good 3 verygood 4 yes - based on SR and ordinal scale adopted 5 n/a 6 doubtful 7 doubtful 8 doubtful	1a very good 2a very good 3a very good - hemiparetic arm and focal spasticity 4a very good - 'a measure of difficulty in active andpassive function for application following focal therapy interventionand in particular for spasticity interventions' 5a Very good 6a adequate 7a n/a as only quant info (surveys) no pure qual 7 n/a 8 /a 9 n/a 10 n/a 11 n/a 12 n/a 13 adequate sample size 14 pilot study/checked with participants very good 15 very good 16 verygood 17 adequate 18 n/a 18 adequate - phoned, face to face orsent out scale 19 adequate 20 doubtful 21 n/a ...26/ very good	
ArmA	Ashford, Jackson & Turner-Stokes 2015	very good	1 Very good 2 very good 3 very good 4 very good 5 very good 6very good – well justified 7 very good 8 very good 9 n/a 10 very good 11 very good 12 n/a 13 n/a 14 NO	Goals were categorised into passive function, active function, symptoms, cosmesis and impairment. Two passive function items not previously identified were identified. Items derived from patientgoals informed standardised measure development.	plus / minus
ArmA	Ashford, Slade, Turner-Stokes 2013	adequate		20-item instrument (7 passive and 13 active) with a five-point ordinal rating scale for use with people with a hemiparetic upperlimb was systematically developed. Tool development methodology support content validity of this tool.	plus

Content Validity					
Measure	**Author**	**COSMIN**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee rating**
ARAT		DOUBTFUL		PROM development using Carroll 1965 and Lyle 1981; Box 1: 1. very good - construct clearly defined " measures of what the patient is able to in everyday task with upper limb 2 doubtful - nil specific model, framework or theory was identified there was rationale provided but not to COSMIN clear level regarding defining other construct to be measured 'it must be kept clearly inmind that the test we are trying to develop is one which ill have direct relationship to what the patient is able to do in every-day activities. Measurement of individual muscle strength and joint range of motion gives some idea of what the patient may be ableto use his hands for, but the ingenuity and use of muscle substitution seen in chronic hand disabilities makes prediction of upper extremity function from muscle test and range of motion tests inaccurate. Clinical observation and p the upper extremity which prosthetic research has shown that there are certain movements of the upper extremity which are of more value than others' CARROLL. 3 VERY GOOD measure was developed for adults with general hemiplegia (included stroke, trauma, arthritis) 4. VERY GOOD The UEFT "can be used in properly selected patients to determine what functions of the upper extremity are impaired and need treatments it can also be used to measure changes in hand function with advancing disease, surgical or other treatment' CARROLL, ARAT 'LYLE: rapid, yet reliable standardised performance test for use in assessing recovery of upper limb function following cortical damage - monitoring progress through treatment, evaluating treatment clinically or in a research context, 5 VERY GOOD - sample that the development was completed in included target sample, Carroll was mixed but did include neuro and Lyle did also include neuro plus amp and arthritis. 6. DOUBTFUL as per COSMIN this should be assigned ifonly a survey was completed as they recommend personal contact 7. NOT APPLICABLE as interviews not conducted 8. NOT APPLOCABLE as per 7 9. NOT APPLICABLE as per 7 through to 12 13 VERY GOOD initial 200 then 278 in nursing home	

Content Validity					
Measure	**Author**	**COSMIN**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee rating**
ARAT		DOUBTFUL likely inadequate		1b "comprehensibility' and "comprehensiveness' 14. VERY GOODyes a pilot test t determine items was completed 15 VERY GOODwas in representative sample 16 DOUBTFUL not clear if pts wereasked assumed not - skip to 25 VERY GOOD problems found adjustments made and retested (LYLE data) 26 DOUBTFUL NO pts not asked 27.	? Combined GRADE low +
ARAT			BOX 2 evaluating quality of studies on content validity	no studies	
CMSA	Gowland 1990	inadequate		box 1a 1 very good 2 very good ICF 3 very good 4 very good acute recovery staging 5 very good 6 doubtful only quan no qual 7na 8 na 9 na 10 na 11 na 12 na 13. 14 very good 15 very good 16 inadequate	
CMSA	Gowland 1990			BOX 1b pts not asked about comprehensiveness nor comprehensibility therefore Part 2 a,b,c not completed. Professionals not ASKED about relevance or comprehensiveness.	
CMSA	Moreland, Gowland, Van Hullenaar & Huijbregts 1993	inadequate		Box 1a 1 very good 2. very good 3. very good 4. very good 5 very good 6 inadequate 7. n/a 8 n/a 9 n/a 10 n/a 11 n/a 12 n/a 13verygood 14 inadequate 35 adequate theoretical basis provided underlying the content of the CMSA	
CMSA	Moreland, Gowland, Van Hullenaar& Huijbregts 1993			Box 1b pts not asked about comprehensiveness norcomprehensibility	
EQ-5D	EuroQol group	inadequate		Box 1a very good 'measure capable of being used in large scale survey of the community and self completed questionnaire, shouldcompliment other forms of QoL measures 2. origin clear based ontheoretical ideas VERY GOOD 3 very good - community general public 4 very good - evaluative of general community in addition tocollection of a common data set 5 very good 6 adequate - doubtful7 na 8 na 9 na 10 na - 14 very good 15 very good 16 no INADEQUATE	

Content Validity					
Measure	**Author**	**COSMIN**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee rating**
FIM	Keith et al 1987			box 1a 1 very good 2 doubtful 3 very good 4 very good 5 very good 6 doubtful 7 doubtful 8 doubtful 9 doubtful 10 doubtful 11 n/a 12 doubtful 13 n/a 14 very good pilot test 15 very good 16 inadequate 26 doubtful	
FIM	Keith et al 1987			relevance - inadequate (pts not asked) comprehensiveness and comprehensibility doubtful as not assessed at all	
MI	Demeuriss e et al 1980 not found in search	doubtful inadequate		Box 1a - 1. very good - motricity of the UL, LL 2 Doubtful - based on Medical Research Council 3. very good - vascular hemiplegia target population 4. doubtful - context is not detailed down to age, assumed across the acute to subacute rehab phase post vascularhemiplegia 5. very good study development was completed in target population 6 Doubtful - nil qualitative analysis undertaken toidentify relevant items 7 n/a 8 n/a 9 n/a 10 doubtful 11 n/a 12 n/a 13 unknown as items were chosen via "selection from occ therapy records' not stated how many n/a 14 yes pilot test very good 15. very good 16 no Inadequate	
MI	Demeurisse et al 1980	no studies		box 1b pts not asked about comprehensibility norcomprehensiveness	
SIS	Salter	doubtful		2e comprehensiveness 27. adequate or doubtful UNMSURE 28 doubtful 29 doubtful 30 adequate 31 very good 2 researchersinvolved	?
SF-36	Ware & Sherbourne 1992	inadequate		box 1a very good - used criterion full length MOS 2. very good 3.very good - designed for use by 14 years plus self administered via person or phone general population 4 very good 5 very good6 doubtful - assumable as compared to existing measures ot identify items and only quantitative not qual 7 n/a 8n/a 9 na 10 ? INADEQUATE as pts not included in a qual approach	
UL MAS	Carr et al 1985	inadequate		box 1 a 1. very good 2. very good 3 very good target population stroke patients 4 very good stroke rehab across the continuum 5 very good 6 doubtful - nil qual completed 7 na 8 na 9 na 10 na 11 na 12 na 13 ok 14 very good 16 inadequate	
UL MAS	Carr et al 1985			box 1b pts not asked about comprehensibility nor comprehensiveness	

Content Validity					
Measure	**Author**	**COSMIN**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee rating**
Content validity studies after tool development					
EQ-5D	Salter, Moses,Foley & Teasell 2008	only applicable forrelevance and comprehens iveness - by professional snot patients	relevance 22. doubtful - used Fitzpatrick's research to gauge tems 23 doubtful professional were not all from required disciplines 24 doubtful not tested in a number of professionals 25 doubtful 26 very good 2 researchers used in analysis	2 e comprehensiveness asking professionals 27. doubtful 28 doubtful 29 doubtful 30 doubtful 31 very good SIS included 7/9 dimensions - did not include symptoms,satisfaction with care 78% most comprehensive out of al measures.	relevance doubtful ? comprehensiveness doubtful +
EQ-5D	Kuspinar & Mayo 2013	relevance and comprehens iveness by patients	1 adequate 2 very good number of pts 3 doubtful not stated if interviewer/moderators were trained 4. very good 5 very good 6 adequate 7 very good 4 researchers	comprehensiveness 8 adequate 9 very good 10 doubtful 11 very good 12 very good 13 adequate 14 very good EQ-5D contains 4/10 domains (school/work, walking, housework,mood) identified by MS subjects and 2/7 domains in a generic utility measure (one including manual dexterity)	relevance doubtful ? comprehensiveness doubtful -
SA-SIP	Salter, Moses,Foley &Teasell 2008	only applicable forrelevance and comprehens iveness - by professional s not atients	relevance 22. doubtful - used Fitzpatrick's research to gauge items 23 doubtful professional were not all from required disciplines 24 doubtful not tested in a number of professionals 25 doubtful 26 very good 2 researchers used in analysis	2 e comprehensiveness asking professionals 27. doubtful 28 doubtful 29 doubtful 30 doubtful 31 very good SA-SIP-30 included 5/9 dimensions did not include symptoms,personal constructs, satisfaction with care 56%	relevance doubtful ? comprehensiveness doubtful ?

Content Validity					
Measure	**Author**	**COSMIN**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee rating**
SA-SIP	van Straten, deHaan, Limburg, van den Bos 1997	developmentstudy	1. very good 2 very good full SIP 3 very good 4 very good 5 very good 6 very good 7 n/a 8 very good original SIP 9very good 10 very good 11n/a 12 n/a 13 very good n = 329, 14 very good 15 very good 16 inadequate pts not asked about comprehensibility once items removed		inadequate development, relevance + other 2 ?
SF-36	Salter, Moses,Foley &Teasell 2008	only applicable forrelevance and comprehensiveness - by professionalsnot patients	relevance 22. doubtful - used Fitzpatrick's research to gauge items 23 doubtful professional were not all from required disciplines 24 doubtful not tested in a number of professionals 25 doubtful 26 very good 2 researchers used in analysis	2 e comprehensiveness asking professionals 27. doubtful 28 doubtful 29 doubtful 30 doubtful 31 very good SF-36 included 6 of 9 dimension - did not include cog function,personal constructs, satisfaction with care 67%.	relevance doubtful ? comprehensiveness doubtful ?
SIS	Salter, Moses,Foley &Teasell 2008	only applicable forrelevance and comprehensiveness - by professionalsnot patients	relevance 22. doubtful - used Fitzpatrick's research to gauge items 23 doubtful professional were not all from required disciplines 24 doubtful not tested in a number of professionals 25 doubtful 26 very good 2 researchers used in analysis	2 e comprehensiveness asking professionals 27. doubtful 28 doubtful 29 doubtful 30 doubtful 31 very good EQ-5D 6/9 dimension - did not include cog function, personalconstructs, satisfaction with care 67%.	relevance ? Comprehensiveness ? Comprehensibility?

Content Validity					
Measure	**Author**	**COSMIN**	**Finding**	**Reason for score as per COSMIN Guidelines**	**Terwee rating**
SIS	Duncan, Wallace, Min Lai, Johnson, Embretson, Laster 1999		inadequate see next	inadequate see next	inadequate development, relevance + other 2 ? reports SIS developed followinga comprehensive iterative process - unpublished information?? Discusses use of patients and caregivers and standardised instrument development guidelines but unable to view sameso INADEQUATE

Structural Validity					
Outcome measure	**Author**	**Findings**	**COSMIN SCORE**		**Terwee rating**
ARAT	Lyle 1981	inter-item correlations was completed to reduce items with 4 subscales derived from the UEFT toform the ARAt; Grasp, Grip, Pinch, Gross Movements, each in hierarchical order evidenced by fulfilling the Guttman scaling statistical criteria (coefficient of reproducibility 0.9and above coefficient of scalability above 0.6 - Grasp 0.98, 0.94 Grip 0.99, 0.94, Pinch 0.99, 0.98 Gross 0.98, 0.97). computer simulation program demonstrated between 67% and 87%reduction in test length through reliance on hierarchical properties of the Guttman scale.	1. yes 2. EXCELLENT 3. N/A 4. POOR sample size, 5. EXCELLENT 6 not CTT 7 IRT EXCELLENT. COSMIN update results Box 3. structural validity 1. have completed correlational analysis of items but to fit withGutman so not CTT 2. Every good 3. inadequate, 20 people4 very good. INADEQUATE	poor with sample, excellent without sample	plus
ArmA	Ashford, Turner-Stokes, Siegert & Slade 2013	1 yes reflective and concerns unidimensionality	1a adequate Principal component analysis 2 very good 3. unsure how to rate sample size - as paper references PCA and minimal sample size - however this boxadopts both IRT and CTT approaches and has <100 sample size? 4 very good	adequate	plus
ArmA	Ashford, Siegert & Alexandrescu 2016	4 of the 7 items in the ArmA passive subscale initially had disordered thresholds -these were rescored to 4 scoring options - from there the subscale conformed to the Rasch model and could differentiate between at least 2 groups of patients	yes and structural validity 1n/a 2 very good 3 adequatesample 3 very good	adequate	plus
FIM	Chen, Chen, Hreha, Goedert & Barrett 2015	KF-NAP accounted for 11.6% additional variancein patients performance compared to fIM and was distinct from 2 factors containing subsets of fIM and BI - captures different performance of ADL information - FIM, BIU unique to KF-NAP	1. 2. EXCELLENT 3. FAIR 4. EXCELLENT sample size 5 EXCELLENT 6 EXCELLENT	fair - excellent	plus COSMIN UPDATE +
FIM	Heinemann et al 1994	97.9% of the variance was accounted for.	Good as IRT model (rasch) was not adequately described, i.e. what is Rasch.	adequate	plus
FIM	Heinemann, Linacre, Wright, Hamilton & Granger 1993	Supported FIM as an interval measure and supported a motor and cognitive subscale. Differences between subgroups as expected.	good; assumable how missingitems were handled.	adequate	plus

Structural Validity					
Outcome measure	**Author**	**Findings**	**COSMIN SCORE**		**Terwee rating**
FIM	Sharrack, Hughes, Soudain& Dunn 1999	Supported 2 factor structure, 89.4% of totalvariance explained.	based on unidimensionality 1. adequate 2 n/a 3inadequatesample size 65	inadequate	plus
FIM	Stineman et al 1996	Item -total correlations 0.26-0.91 (FIM motor) and 0.60-0.87 (FIM cognitive). Supported the motor and cognitive subscales but also raised that a 3 or 4 factor structure was possible.>50% of variance explained in following subgroups; stroke, non traumatic brain injury, traumatic braininjury, non traumatic spinal cord injury, traumaticspinal cord injury, Gullian Barre, general neurological, lower extremity fracture, joint replacement, other orthopaedic, lower extremity amputation, other amputation, osteoarthritis, rheumatoid arthritis, cardiac, pulmonary, pain, major multiple trauma, multiple trauma with brainand spine, others.	adequate structural validity good, internal consistency good; I went with good for structural validity as technically PCA analysis is viewed separately to CFA or EFA, however, I went good asI didn't want to mark the paperdown completely for using PCA like COSMIN seems to do. I applied a good rating to internal consistency for this reason too as it has the link to factor analysis.	adequate	plus
MAL	Uswatte, Taub, Morris, Light & Tompson 2006	2 items - write on paper (48%) and put in makeup / shaving cream (20%) had a higher or equal proportion of missing data than the priori cut off of 20% and were eliminated. 92% of these items had item-total correlations >0.5	1. yes 2. not stated but assumed 0 GOOD 3. GOOD 4. excellent sample 5 excellent 6. POOR no exploratory or confirmatory analyses performed COSMIN UPDATEconcerns structural validity 1. inadequate 2. very good 30 x 7 = 210 n=222 4. doubtful - very difficult to follow structureof study INADEQUATE	poor - excellent	?
MAL-28	Uswatte, Taub, Morris, Light & Tompson 2006	2 items - write on paper (48%) and put in makeup / shaving cream (20%) had a higher orequal proportion of missing data than the prioricut-off of 20% and were eliminated. 92% of these items had item-total correlations >0.5	1. yes 2. not stated but assumed 0 GOOD 3. GOOD 4. excellent sample 5 excellent 6. POOR no exploratory or confirmatory analyses performed COSMIN UPDATEconcerns structural validity 1. inadequate 2. very good 30 x 7 = 210 n=222 4. doubtful - very difficult to follow structureof study INADEQUATE	poor - excellent	?

Structural Validity					
Outcome measure	**Author**	**Findings**	**COSMIN SCORE**		**Terwee rating**
BI(C&W)	Barer & Murphy 1993	**POOR** Authors conducted Guttman scaling and determined that the Barthel Index does not form a unidimensional assessment tool.	COSMIN UPDATE 1 n/a 2 very good 3. 1870 sample verygood however no reference or clear reasoning to unidimensionality of scale, no information as how or why scale was simplified	inadequate	minus as infit and outfitmean square - removal of get up from chair, participation feel emotionally connected
RMA	Adams, Pickering & Taylor1997 (acute)	RMA gross motor and arm sections are ordered when used for acute stroke based on coefficientsof scalability (CS) (>0.6) and reproducibility (CR)(>0.9). Leg and trunk sections did not. Gross function CS - wk 1 0.81, wk3 0.81, wk6 0.81 CR wk 10.91, wk3 0.96 wk6 0.97; leg and trunk CS wk 1 0.007, wk3 0.21 wk6 0.10 CR 0.80, .84, 0.81 arm function CS 0.84, 0.86, 0.79 CR 0.96, 0.96, 0.97	1. yes 2. good 3. good 4. poor sample size 5. fair - strict inclusion criteria but not reported what this was ? Representative sample 6. poor7 n/a COSMIN UPDATE structural validity 1 n/a 2very good 3 inadequate sample	poor - excellent	plus IRT COSMIN UPDATE gross section + arm section + leg trunk - (50% -) scored - as Guttman
RMA	Adams, Pickering, Ashburn& Lincoln 1997 (2 non acute)	coefficients of reproducibility (CR) and scalability(CS) were performed CR>0.9 and CS > 0.6 indicate scale meets Guttman scaling criteria. Non acute stroke patients over 65 yrs gross section met both CR and CS at 12 mo but failed CS at 6 mo. Leg and trunk was a negative valuefar less than 0.6. The arm section did not meet Guttman scaling on CS but did on CR. Gross fx 6 mo (n=206) CS 0.59 CR 0.91 12 mo n = 83 CS 0.65 CR 0.926. Leg and trunk 6 o CS -0.45, CR 0.69 12 mo CS -0.34, CR 0.71, Arm function 6 mo CS 0.56 CR 0.92 12 mo CS 0.55 CR 0.91. PTs under 65 years gross section 6 mo CS 0.49 CR 0.93 12 mo CS 0.61 CR 0.94 Leg and trunk 6 mo CS -0.28 CR 0.73 12 mo CS -0.34 CR 0.74Arm section 6 mo CS 0.54 CR 0.93 12 mo CS 0.37 CR 0.91. summary only non-acute stroke inboth over and under 65 met Guttman criteria in the gross function section - suggesting correct order of difficulty, leg and trunk did not meet andwere closer to scaling f order was reversed. ARMsection issues with item 10 and 12. arm and leg and trunk may be used as a clinical checklist , RMA not expected to be an appropriate assessment with non-acute stroke pts	1 yes 2. yes EXCELLENT 3 no but can be deduced GOOD 4. GOOD 5. EXCELLENT 6 N/a asIRT approach considered (POOR no exploratory or confirmatory factor analysis performed) 7. Poor COSMIN UPDATE . Structural validity 1n/a 2 very good 3 n = 83 at follow up inadequate sample size 4	poor - excellent	COSMIN UPDATE - as Guttman scaling criteria notmet

Structural Validity					
Outcome measure	**Author**	**Findings**	**COSMIN SCORE**		**Terwee rating**
SF-36	Guilfoyle, et al 2010	The two component model explained less than 75% of the reliable variance in three of the eightdomains, and accounted for 68.6% of the total variance in the data. Findings do not support theuse of The Physical and mental component summary should be interpreted with caution in the TBI population. Scaling assumptions for reporting domain scores in TBI fulfilled however	COSMIN UPDATE 1. STRUCTURAL VALIDITY 1. very good 2 n/a 3 very good 5very good	very good	CTT as factor analysis(OCA completed)no Rasch analysis Plus
SF-36	Hobart, Williams, Moran & Thompson 2002	Regarding SCALE SCORES: All item/own scale correlations corrected for overlap except 2 itemsin the GH scale exceeded 0.40. indicating, for the other 7 scales, the items in each scale measured a common underlying construct, and that the criterion of Ware et al for equivalence of item-total correlations was satisfied. All item/owncorrelations exceeded item/other scale correlations - no scaling failures. Regarding SUMMARY SCORES: PCA scales extracted 2 components with Eigenvalues exceeding unity and the scree plot supported the existence of 2 higher order factors - supporting the hypothesis that a 2D model of health underpins the SF-36 instroke. HOWEVER, these 2 components explained only 58.8% of the total reliable variance in all SF-36 scales and less than 75% of the reliable variable in 5 out of the 8 scales. Therefore, a substantial amount of information from SF-36 scales is lost when summary measures are reported in stroke. CONCLUSION:scaling assumptions were not fully satisfied for either scale or summary scores.	COSMIN UPDATE concernsstructural validity 1. CTT adequate (principal component analysis is part of exploratory factor analysis) 2n/a 3 adequate 4 very good	very good	minus
SF-36	MacKenzie 2002 'Using theSF-36'	The correlations between SF-36 and rotated principal components for the Pennsylvania trauma study followed those obtained for the USpopulation by Ware et al thus supporting the useof the two dimensional factor structure of the SF-36 (i.e. MH and PH) - supports the two valid constructs of physical and mental health	COSMIN UPDATE 1. structural validity 1. very good2 n/a 3 very good 4 very good	very good	plus model fit

Structural Validity					
Outcome measure	**Author**	**Findings**	**COSMIN SCORE**		**Terwee rating**
SIS	Duncan, Bode, Lai, PereraGAIN 2003	Most items in 8 domains measured intended construct and were unidimensional. 3 items misfit the rest if the items in their domains - fit statistics for deleted items (infit mean square) memory: add and subtract no 1.51, mobility: get up from chair 1.50, participation: feel emotionally connected 1.56. Two items (handler money, manage finances) misfit the composite physical domain and were deleted asmeasure cognitive skills while the rest measure motor skills. Empirical ordering of items by difficulty was consistent.	1. yes 2. excellent 3. good 4 excellent 5 excellent 6 not applicable 7 excellent COSMIN UPDATE unidimensionality 1. n/a 2 verygood 3 very good 4 very good	good - excellent	plus as meets bi factor model
SIS	Jenkinson, Fitzpatrick, Crocker & Peters 2013	principle component factor analysis of the 8 items suggested that the 8 domains could be aggregated into a single index (SIS index) with internal reliability using Cronbach's alpha scoringas 0.93, indicating high level of internal reliability, accounting for 68.76% of the variance. SF-SIS was created with PCA of 8 items produced a single factor accounting for 57.25% of the variance with internal consistency reliability of the 8 items was high 0.89.	1. yes 2. excellent 3. excellent 4. excellent 5 fair % of missing responses may have biased results 6.excellent as investigating in a new sample COSMIN UPDATE 1. structural validity 1. adequate 2 n/a 3 adequate 4 very good	very good	plus
SIS	Duncan, Lai, Bode, Perera, DeRosa & GAIN Americas Investigators 2003	SIS 3 created following Rasch analysis and consequent removal of memory -add and subtract numbers, mobility - get up from a chair,participation - feel emotionally connected. Remaining items	1. yes 2. excellent 3. excellent 4. excellent n= 621 5.EXCELLENT 6. n/a 7EXCELLENT	excellent	plus as per references within paper
UL MAS	Khan, Chien & Brauer 2013	confirmed was a unidimensional structure.	1. yes 2 excellent 3 n/a 4 excellent 5 fair - nil informationgiven re recruitment criteria, blinding of results as treating physio 6. n/a 7. excellent COSMIN UPDATE unidimensionality 1. n/a 2very good 3 very good 4 adequate -limited details	fair - excellent adequate	plus

Structural Validity					
Outcome measure	**Author**	**Findings**	**COSMIN SCORE**		**Terwee rating**
UL MAS	Lannin 2004	factor analysis revealed one dimension underlying the UL-Mas, explaining 81% of the total variance The UL-MAS score can provide asingle composite score (not needed to report on individual 3) that could be interpreted as a total score for the UL function in this population. UL- MAS is a valid and reliable independent scale	1. yes 2. EXCELLENT 3. N/A 4. excellent reference given forsample 5.EXCELLENT 6. excellent - confirmatory factor analysis completed. COSMIN UPDATE 1. very good 2 n/a 3 very good 4 very good	very good	?
UL MAS	Miller, Slade, Pallant, Galea2010	removal of item 72 (radial deviation of the wrist) improved the fit and internal consistency as item response bias was found for pts over 65 years ofage who found it considerably harder. • Rasch analysis found that Item 72 – wrist deviation – was found to be systematically easier for patientsunder 65 years of age to perform – so bottom line if your clients are over 65 years of age radialdeviation of the wrist may not be indicative of function. The validity of the hierarchical scoring for item 6 and 7 was confirmed however inconsistences were found within item 8. Implying that the order of difficulty was not ascending and accurate. UL MAS could discriminate/stratify patients with differing levelsof UL abilities Person Separation Index (PSI) 0.96 using the 17 items only as wrist deviation is now not considered an essential movement to beretrained for reaching and grasping.	1 yes 2. GOOD 3 GOOD 4 EXCELLENT 5 EXCELLENT (trained assessors, standardised approach -masked 6 n/a 7 principal component analysis confirmedthe unidimensionality of the scales EXCELLENT COSMIN UPDATE 1. n/a 2 very good 3very good 4 very good	? COSMIN updateplus	? Not supporting summary scores and 100% scale scores only6/8
UL MAS	Pickering, Hubbard, Baker& Parsons 2010	scoring hierarchy was upheld in subset 6, not in subset 7 with results indicating that item 3 was least difficult followed by 1, 4, 2, 5 and 6 in orderof increasing difficulty. In subset 8 hierarchy wasnot upheld , item 1 was least difficult followed by6, then 2 and 5 equal value and 3 and 4 of equalvalue. recommend every item should be scored regardless.	1. yes 2. EXCELLENT 3 EXCELLENT 4 POOR but reference given 5 POOR unclear what actually happened methodologically COSMIN UPDATE 1. n/a 2. very good 3 n = 40 inadequate4 doubtful methodology detail very limited	poor - excellent inadequate	plus
UL MAS	Sabari, Lim, Velozo, Lehman, Kieran & Lai 2005	inconsistencies were found regarding the ordering of the hierarchical scoring system for hand function and advanced hand activities. There was support for the scoring scale for the upper-arm scale. Additional items and the removal of some items were proposed with further testing recommended.	1. yes 2. GOOD 3 GOOD 4 GOOD 5 FAIR 6 n/a 7 EXCELLENT COSMIN UPDATE 1. n/a 2 very good 3 adequate 4 doubtful - gave criteria in stats but did not apply it in the end re:MnSQ	check new criteria -	minus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
10MWT	Donovan, Lord, McNaughton & Weatherall 2008	nil specific - explore relationship between clinic and community-based outcome measures of gait speed	1. 1. GOOD 2. N/A 3. Fair 4 Poor 5 Good 6 GOOD 7 GOOD 8 FAIR 9 FAIR – generalizability only, limited info in assessor. 10. EXCELLENT COSMIN UPDATE 1. very good 2. adequate 3 very good 4 adequate	poor - excellent	LOA (limits of agreement) between 10MWT and 6MWT resulted in wide intervals (-13.7 to 13.4m/min) the 10MWT tended to overestimate the gait speed 6MWT performance of the slower participants (<35m/min) and underestimate the 6MWT performance of the faster stroke participants (>45m/min). The clinic 6MWT compared more favourably than the clinic 10MWT with the street or mall 6MWT (LOA -5.7 to 8.9 and -9.2 to 7.8 respectively).	? COSMIN update -0/2 10MWT didn't Reflect changes In community, didn't agree With 6MWT
10MWT	Hirsch, Williams, Norton & Hammond 2014	hypothesised that participants would demonstrate high test-re-test reliability during walking at self-selected pace as well as when the gait task becomes more challenging (increased speed). Nil specific hypothesis regarding concurrent validity	1. EXCELLENT 2 N/A 3 EXCELLENT -power analysis for sample size calculated needed 10-15 as per calculation and recruited 23. 4 fair 5. GOOD 6. GOOD 7 EXCELLENT 8.EXCELLENT 9 EXCELLENT 10 EXCELLENT COSMIN UPDATE 1. VERY GOOD 2. VERY GOOD 3 VERY GOOD	fair - excellent	the use of self selected pace or fastest ace did not meaningfully affect gait velocity results – it doesn't clinically matter which version of the 10MWT used. There was slightly more variability with the fastest pace than for self selected pace. For SSP the range was -0.133 seconds to 0.157 with 95.7% of the difference between -0.1 and +0.1 seconds. For FP the range was between - 0.152 to +0.189 seconds with 88.9% of the differences between -0.1 and +0.1 seconds.	? No validity hypothesis COSMIN update +1/1 both speeds achieve agreement

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
10MWT	Kuys, Bew, Lynch, Morrison& Brauer2009	research question: which measure of activity limitation on admission to rehabilitationafter stroke best predict walking speed at discharge. COSMI UPDATE hypotheses 10MWT would be a predictorof DC walking speed as per correlations and significance	1. EXCELLENT 2. n/a 3. EXCELLENT 4. FAIR 5. GOOD 6. GOOD 7. GOOD 8 EXCELLENT 9 EXCELLENT 10EXCELLENT COSMIN UPDATE 1. adequate 2 very good 3 verygood	fair - excellent adequate	when significant predictors of discharge walking speed (admission walking speed 10MWT, modified elderly mobility scale score, MAS item 1to 5, FIM motor component, FIM and TUG, wereentered into multiple linear regression, dischargewalking speed was best predicted by 10MWT at admission and MAS item 2 (supine lying to sitting over side of bed) R2 = 0.36. (accuracy of prediction). clinicians could predict using equation: discharge walking speed m/s = 0.33 + 0.47 admission walking speed + 0.05 Item 2 MAS score. OTHER: admission walking speed (10MWT) (n=120) relationship with discharge walking speed (from univariate analysis) 0.32 (<0.001), MAS item 6 (n=105) 0.09 (0.14) item 70.06 (0.23) item 8 0.07(0.36), FIM 0.25 (>0.001).	? + (1/1)
10MWT	Miller, Combs, Van Puymbroeck, Altenburger, Kean, Dierks & Schmid 2013	nil specific hypothesis stated but purpose to examine the correlations between fatigue and pain with identified outcomes relevant to rehab (balance, gait, activity, participation, chronic diseaseself-efficacy and balance self-efficacy COSMIN UPDATE hypotheses generic # 3 for pain and #3 for fatigue	1. EXCELLENT 2. n/a 3. GOOD 4. fair -references to prev correlations and state expectations 5 GOOD 6 GOOD 7Excellent 8 EXCELLENT 9 FAIR 10 EXCELLENT COSMIN UPDATE 1. very good 2 very good 3 adequate	fair - excellent	10MWT correlated weakly with fatigue 0.175 (P.0.128) and with pain -0.044 (P. 0.706). Pts with high fatigue and high pain scored 1.28 +/-0.66 compared to high with low fatigue and low pain 1.47 +/- 0.60 (P. 0.342). There were strong correlations between post stroke pain and fatigue with patients beliefs about their abilitiesindependent of performance on physical measures.	? COSMIN UPDATE + 2/2 didn't correlate with unrelated constructs

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
10MWT	Schmid, Van Puymbroeck, Altenburg er, Dierks,Miller, Damush & Williams 2012	examine the associationsbetween activity and participation and multiple poststroke mobility variables determine which post stroke mobility variables are independently associated with activity and participation COSMIN UPDATE generic hypotheses #1 - 10MWT and activity/participation measure IMPACT.	1. excellent 2 n/a 3. good 4. fair 5good 6 good 7 excellent 8 excellent 9 fair - very high functioning volunteers 10 excellent COSMIN UPDATE 1 vert good 2 very good 3 adequate (sample high functioning, variability in completion of measures)	fair - excellent	correlation between 10MWT and IMPACT activity = -0.309, P<0.05 was significant and a weak correlation with both IMPACT participation = -0.219, and IMPACT total = -0.289. Gait speedwas initially not included in the regression modeldue to not being significantly correlated, but when forced into the activity regression mode didnot remain and was not found to be independently associated with activity. Study found that balance or falls self efficacy is more strongly associated with poststroke activity and participation than physical performance measures of gait. Walking capacity (measured with 6MWT) was more important to getting out and about (10MWT) after stroke for 74% of the participants	? COSMIN UPDATE - 0/1 as <0.5

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
10MWT	Mudge & Stott 2009	" we hypothesised that performance during these common walking conditions may have a stronger relationship to usual walking activity in natural settings than do the timed walking tests. Nil stated but aims; to determine the strength of therelationship between commonly used clinical tests of walking ability and the available StepWatch outputs,and in particular to determine how well clinical walking testspredict ambulatory activity in natural environments as measured by the StepWatch.COSMIN UPDATE generic hypotheses #1	1. Excellent 2. EXCELLENT 3. EXCELLENT 49 but as per powersample size calculation 4. EXCELLENT 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 FAIR selected sample high level, volunteers not generalisable 10EXCELLENT COSMIN UPDATE 1. very good 2 very good 3 doubtful	fair - excellent	10MWT did not correlate strongly with StepWatch outputs excluding highest step rate in one minute 0.71. other correlations were moderate 0.41 - 0.64 and included outputs suchas mean steps, % of time with no steps, numberof steps at low rate and high rate, peak activity index, highest step rate in 60 minutes. In addition the 10MWT did not make any further independent contribution t the variance when completing regression analysis. The 6 minute walk test demonstrated the strongest relationship with the StepWatch and better predictions of walking ability however only 50% of the variability in usual walking ability was explained.	plus COSMIN UPDATE - <75% (0/1) didnot correlatewith activity

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
10MWT - comfortable pace and fast pace	Combs, Dugan, Passmore, Reisner, Whipker, Yingling & Curtis 2010	secondary purpose was to examine the relationship of changes in gait speed to changes in balance, balanceconfidence, and health related quality of life.	1. EXCELLENT yes 3 excluded due to missing data 2. GOOD 3 POOR n=16 4. POOR knew purpose what to examine the relationship but not clear what was expected 5 GOOD 6 GOOD7 EXCELLENT 8 EXCELLENT 9FAIR - researcher not blinded topurpose of the study 10 EXCELLENT	poor - excellent	change in walking speed (comfortable walk speed and fast walk speed) and the relationshipwith QoL using Pearson correlation coefficient (r); the change in CWT from pre to post (1 week before to ix and immediately after 8 week Ix) andchange pre test to retention (1 week before to 6 mo post Ix) with SIS ADL / IADL 0.07 (-0.44 to 0.55), 0.16 (-0.36 to 0.61) SIS mobility 0.25 (-0.28 to 0.66), 0.19 (-0.34 to 0.63) SIS participation 0.14 (-0.38 to 0.59), 0.14 (-0.38 to 0.59) SIS Stroke recovery scale 0.30 (-0.23 to 0.69), 0.52 (significant) (0.03 to 0.81), summary no correlation to low. Fast walk speed same timeframes with SIS ADL/IADL 0.04 (0.46 to 0.53), 0.15 (-0.37 to 0.60) SIS Mobility 0.37 (-0.15 to 0.73), 0.33 (-0.20 to 0.71) SIS Participation 0.03(-0.47 to 0.52), 0.08 (-0.43 to 0.55) SIS stroke recovery scale 0.23(-0.30 to 0.65), 0.29 (-0.25 to 0.68) summary no to low correlation between increased walking speed with QoL.	?
9HPT	Beebe & Lang 2009	to determine how 6 clinical test were related to each other in the first weeks and months after stroke COSMINUPDATE #1 jebsen taylor hand test, #1 ARAT #1 SIS-hand, # 3 grip strength #3pinch strength	1. excellent 2. fair 3. poor 4 fair 5good 6 good 7 excellent 8 excellent 9 fair 10 excellent	Fair - Excellent	spearman clinical test correlations at 1, 3 and 6 months; 9HPT & Grip 1: 0.80 3: 0.78 6: 0.82. 9HPT & Pinch 1: 0.77 3: 0.78 6: 0.61 9HPT &ARAT 1: 0.87 3: 0.93 6: 0.85 9HPT & Jebsen Taylor 1: 0.84 3: 0.97 6: 0.97 9HPT & SIS-Hand 1: 0.66 3: 0.62 6: 0.53 (strong correlation r>0.75)	? No clear hypotheses COSMIN UPDATE Grip -, pinch - (impairment 0.61 - 0.82), Jebsen +, ARAT +,SIS hand + (3/5)60% (hand function / dexterity 0.53 - 0.97)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
9HPT	Benedict, Holtzer, Motl, Foley, Kaur, Hojnacki & Weinstock-Guttman 2011	"performance on executive function tasks would be significantly correlated with motor performance in MS after controlling for demographics, disease characteristic such as diseaseduration and other cognitive domains	1 GOOD not described but canbe deduced nil 2. not applicable 3. EXCELLENT sample size 4 GOOD minimal hypothesis formulated 5 EXCELLENT 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 FAIR mild to moddisability and pts taking medications that could impact oncognition as retrospective study 10 EXCELLENT COSMIN UPDATE adequate	fair - excellent	better cognitive performance was associated with faster times on motor tasks. 9HPT & COWAT -0.27, JLO -0.20 CVLT2 -0.47 BVMTR - 0.45 SDMT -0.65 PASAT -0.41 DKEFS CS -0.43DKFEFS DS -0.35 p<0.01	plus COSMIN update6/8 +
9HPT	Costelloe, O'Rourke, McGuigan , Walsh, Tubridy & Hutchinson 2008	not clearly stated but reference to "we would expectthese correlations to be strong" when discussing convergent validity of 9HPT with the MSIS-29 physical scores	1. Excellent 2. n/a 3 Excellent sample size 4. GOOD 5 GOOD 6GOOD 7FAIR 8 FAIR 9 FAIR - minimal description so not sure ofhow, randomised, blinding 10 EXCELLENT	Fair - Excellent	baseline correlation with 9HPT and MSIS-29 physical -0.54 follow-up correlation -0.61 correlation with 9HPT and MSIS-29 psychological at baseline -0.19 follow up -0.29.MSIS-29 = Multiple Sclerosis Impact Scale. Correlations as er author 0.40-0.60 moderate0.60-0.80 good ?0.80 excellent.	-

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
9HPT	Goodkin, Hertsgaard & Seminary 1988	"we hypothesised that a significant percentage of patients might show objective change in UE function in testing instruments with greater sensitivity than the EDSS.MRD or A1 which are more sensitive to detecting change in lower extremity function"	1. not described GOOD 2. GOOD 3. EXCELLENT sample size 4 GOOD 5 EXCELLENT 6 GOOD 7 FAIR description of comparator constructs 8 GOOD 9FAIR minimal info given re the collection of baseline data to determine any flaw/bias, nil cognitive or other participant info given to assist in generalising results 10 GOOD assumable as no detail re type of correlation performed	fair - excellent	9HPT and box and block test moderately correlated with each other control group - 0.761 prospective group -0.745. however each test was able to detect changes the other did not (detail not provided) for e.g. in the prospective group - the 9HPT detected a change of 10, 20 30% the BBT detected change of the same magnitude in only 50% of those. When the BBT detected change of 10, 20, 30% approximately 50% showed a change on the 9HPT. Further; the BBT detected a change of more than 20% on 9.8% of the trials but not by the 9HPT and the9HPT alone detected thus type of change on 12% of trials. both the 9HPT and BBT detected more change than the EDSS.MRD. recommendation to administer both 9HPT and BBT as they detect different changes.	plus
9HPT	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987 (b)	nil stated - four measures were investigated. COSMIN UPDATE NHPT more sensitive than FAT, pts scoring below cut off on NHPT will also score highlyon FAT (other measure of function (NHPT)	1, EXCELLENT nil missing items,discussed n at baseline equalled n within analyses, 2. N/A 3. GOOD sample size 4. POOR unclear what was expected 5 GOOD 6 GOOD 7 EXCELLENT 8 POOR 9 GOOD	poor-excellent doubtful	Frenchay Arm Test score 5/5 9HPT above cut off (18 seconds) n=17 below cut off 16 FAT score 1-4/5 9HPT 0 were above cut off, 5 werebelow cut-off "0" were 1 FAT score 0/5 9HPT 0above cut off, 0 below cut off 17 "0". These results show 9HPT more sensitive than FAT, only 52% who scored 5/5 on FAT were within normal limits on the 9HPT.	plus (1/1)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
9HPT	Jacob-Lloyd, Dunn, Brain & Lamb 2005	not stated - but aim to establish effective methods ofmeasuring the functional performance of n individual who had experienced a stroke and had been discharged from hospital COSMIN UDATE generic hypotheses #1 UL MI and NHPT at DC and follow up (6 mo later) (2hypotheses	1. EXCELLENT 2. GOOD 3. GOOD 4. FAIR 5 GOOD 6 GOOD 7 FAIR 8 EXCELLENT 9 EXCELLENT 10 spearman rank correlations used but commentedon significance not strength GOOD COSMIN UPDATE 1. very good "upper limb function" 2.very good 3 very good 4	Poor - Excellent	spearman rank correlations; correlation between9HPT and UL MI was significant at discharge r: 0.53 n:22 p:0.01 but not at follow up r: 0.36 n:22p:0.10	? COSMIN UPDATE hypotheses 1 on dc +, follow up hypotheses 2 - (1/2) 50% means -
9HPT	Poole, Nakamoto , McNulty, Montoya, Weill, Dieruf & Skipper 2010	are dexterity and visual perception related to perceived ADL ability COSMIN UPDATE hyp 1 - known group discrimination between different types of MS, hyp 2 #1 ADKL, hyp 3 #3visual perception	1. EXCELLENT 2. EXCELLENT 3. GOOD 4. FAIR 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 FAIR - nil comment made on raters being blind, treating therapist to determine if any bias exists 10. EXCELLENT COSMIN UPDATE 1. VERY GOOD 2 VERY GOOD3 ADEQUATE 5 VERY GOOD 6VERY GOOD	Fair - Excellent	correlations (Pearson product moment correlation); 9HPT (right) with 9HPT left 0.74, with GPT-R (grooved peg test) 0.83, with GPT-L0.71, with MVPT-R -0.05, with EDSS 0.42, with FSI-A (functional status index - assistance 0.22,with FIS - Pain - 0.10, with FSI-Difficulty 0.22, with age -0.14. 9HPT (left) with age -0.00, with GPT-R 0.28, with GPT-L 0.72, with MVPT-R - 0.00, with EDSS 0.63, with FSI-A 0.48, with FSI-P - 0.09, with FSI-D 0.33. IN summary 9HPT does not correlate with motor free visual perceptual tests (MVPT-R) or age, does correlate strongly with GPT (stronger with samehand), and EDSS only. Only left hand NHPT correlated with measures of self reported daily living skills 91% of sample were right hand dominant.	COSMIN UPDATE #1 ? #2 FSI-A R-NHPT -, NHPT L +, FSI - dR NHPT 0.22 -, L NHPT -, FSI P L +AND R NHPT + (#3), Visual perception #3 +, +both R and L so total 6/10so total -

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
AQoL	Hawthorne, Gruen & Kaye 2009	No hypotheses stated. Can assume they expected AQoLscores to be lower for people with TBI compared with the case matched A comparison group. COSMIN UPDATE GENERIC HYPOTHESES #5 AQoL utilities lower in TBIthan controls	1COSMIN UPDATE 5 very good 6 very good	very good	The AQoL was significantly sensitive to TBI status. The TBI cohort obtained scores that were 0.20 utilities lower than those without TBI. The means of each AQoL dimensions were lower in the TBI sample compared with the non-trauma group. Change far exceeded minimum importantdifference of 0.06. more sensitive than SF-36 V2	plus (1/1)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
AQoL	Sturm, Osborne, Dewey, Donnan, Macdonne ll & Thrift 2002	1.'That the magnitude of correlation would be highest between scales measuring similar constructs (providing evidence of convergent validity) and the weakest correlations would be between scales measuring conceptually unrelated constructs (discriminant validity).' 2. that the AQoL scores would be lower (worseHRQoL) in those patients withgreater disability and impairment' 3.'the AQoL utilityscore at 3 months after stroke would predict the outcomes ofdeath, institutionalisation or both 12 months after stroke.' p 2889-2890 COSMIN UPDATE 1. AQoL would be more strongly correlated with LHS than with the BI or NIHSS, plus AQoL at 3 months would predict outcomes of death, institutionalisation or both 12months after stroke TOTAL 2 HYPOTHESES	COSMIN UPDATE 1. very good 2adequate (as evidence outside ofresearch article 3 very good	adequate	SF-36 PCS 0.44, mental summary 0.36, individual dimension >0.5 only SF-36 physical functioning, social functioning and mental health with like parts of AQoL, LHS r = 0.83, BI r = 0.77, anxiety and depression r>0.50 , NIHSS r = - 0.69 study hypotheses greater correlation with LHS than BI and NHS (+), In logistic regression models, the AQoL 3-month score was a significant predictor of the outcomes of death (13 cases), institutionalization (7 cases), or both 12 months after stroke.' (+)	2/2 +

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
ARAT	Beebe & Lang 2009	to determine how 6 clinical test were related to each other in the first weeks and months after stroke COSMINUPDATE #1 jebsen taylor hand test, #1 NHPT #1 SIS-hand, # 3 grip strength #3pinch strength	1. excellent 2. fair 3. poor 4 fair 5 good 6 good 7 excellent 8 excellent 9 fair 10 excellent COSMIN UPDATE 1. very good 2very good 3 very good VERY GOOD	excellent poor - sample size fair-excellent	correlation coefficients calculated using spearman's at 1, 3, 6 months; ARAT & Grip 1m:0.86 (excellent/strong), 3m: 0.73 (moderate),6m: 0.8 (excellent/strong) ARAT & Pinch 1m: 0.79 (excellent/strong), 3m: 0.8 (excellent/strong) 6m: 0.65 (moderate), ARAT &Jebsen 1m: 0.87 (excellent/strong), 3m: 0.95 (excellent/strong), 6m: 0.90 (excellent/strong) ARAT 9HPT 1m: 0.87 (excellent strong), 3m:0.93 excellent/strong), 6m: 0.85 (excellent strong) ARAT & SIS-Hand 1m: 0.73, (moderate)3m: 0.57 (moderate), 6m: 0.69 (moderate)	? COSMIN UPDATE Grip -, pinch -, Jebsen +,NHPT +, SIS hand + (3/5)60%
ARAT	Burridge,Turk, Notley, Pickering& Simpson2009	no hypothesis but question "what components of the upper motor neurone syndrome are likely to affect upper limb activity limitation? COSMIN UPDATE so 8 hypotheses concerning impairment measures sogeneric hypotheses #2	1 yes data for 2 particular measures missing EXCELLENT2 FAIR 3 POOR 17 sample 4 POOR 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 Should the impairments assessment extend beyond wrist only as this is not core or sole contributor to activity performance fair 10 EXCELLENT. COSMIN UPDATE 1. Very good 2 very good 3.adequate ADEQUATE	Poor - Excellent sample excluded fair - excellent	Pearson correlation coefficients examined relationship between variables excluding MAS which used nonparametric Spearman's P correlation. Correlation coefficients indicate positive relationships between the level of upperlimb activity (ARAT) and each of the negative features; Tracking (r = 0.710, p=0.003) AROM (r=0.540, p=0.025), isometric muscle force ((r = 0.515, p=0.034) and r=0.575 p=0.016). ARAT not significantly correlated with any positive features excluding MAS (spearman's p= -0.360)but this failed to achieve significance	? COSMIN UPDATE 4 +, 4 - (50%) 4/8
ARAT	de Weerdt& Harrison1985	that the "motor recovery" as expressed in the Brunnstrom-Fugl-Meyer (B-FM) test is consistent wit the "functional recovery" as measured in theARAT.	1.excellent 2. excellent 3. Good (53) 4. good 5 good 6 good 7 excellent 8 excellent 9 fair researcher not blind to hypothesis 10 excellent UPDATED COSMIN hypothesis 1. change in scores measuring similar construct correlate. 1. very good 2 very good 3. very good VERY GOOD	Excellent - fair	spearman rank 0.91 at 2 weeks and 0.94 at 8 weeks. Difference in scores between the ARATand B-FM: improvement is positively correlatedbut not perfect 28% did not improve on either ofthe tests and 8% deteriorated on one or both tests. ARAT quicker to administer than B-FM.	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
ARAT	Dromerick , Lang, Birkenmeier, Hahn, Sahrmann , Edwards006	evaluate the relationship among impairment, functional limitation and perceived disability assessments COSMIN UPDATE #1 X 3(FIM, MAL, WMFT)	1. missing items not stated (good), 2. nil reference GOOD 3. sample size 39 FAIR 4. hypothesis GOOD 5. direction notstated GOOD 6. magnitude of correlation s not stated GOOD 7. adequate description EXCELLENT 8. EXCELLENT 9. limitations in sample: relevant only to post stroke without moderate or severe cognitive or sensory impairments, may not apply to broader population. trained blinded study personnel FAIR 10. EXCELLENT COSMIN UPDATE 1. VERY GOOD 2 VERY GOOD 3 VERY GOOD	excellentfair	correlation coefficients: total ARAT correlated with FIM motor (r=0.47), MAL QOM (r=0.61) MAL no of activities attempted (r=0.60). Total ARAT highly correlated with WMFT time score (r = -0.65) and WMFT functional ability score (r = 0.95). Pts scoring highly on ARAT and WMFT (functional limitation measures) at 90 days post still had some measurable disability - diminishedQOM on the MAL tasks, indicating ARAT did notcapture full spectrum of motor dysfunction. ARAT may not predict everyday productive ULuse in acute stroke	? COSMIN UPDATE FIM motor(-), WMFT (+) MAL 0.60 - 0.61) (+) 2/3 67% (-)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
ARAT	Fleming, Newham, Roberts-Lewis, Sorinola 2014	no hypothesis but aims; explore in chronic stroke survivors, the potential predictors of self-reported amount of arm use (MAL) andthe potential for increases in the amount of use after task specific training. Also aimed to determine whether the predictors of arm use differed between patients whose dominant and nondominant arms were affected. COSMINUPDATE hypotheses #1 - similar constructs (prediction via score on MAL so looking at high correlations between the same)	1. yes EXCELLENT, 2. yes EXCELLENT, 3 n=30 POOR 4 FAIR 5 GOOD 6 GOOD 7 GOOD8 FAIR 9 physio not blinded poor10.spearmans correlations EXCELLENT. COSMIN UPDATE 1. very good 2 very good 3 adequate ADEQUATE	Poor - Excellent	spearman correlations (R) with baseline amount of use (MAL) & ARAT components at(P<0.05) were positive; Grasp R=0.670 P <0.001, Grip R=0.645 P <0.001, pinch R = 0.609 P<0.001, Gross R= 0.537, P <0.001. baseline ARAT predicted 47% of the variability in baselineMAL AOU (F1, 31, = 27.457; P<0.001). In using equation for regression model, an ARAT scoreof 54 is required to reach a 2.5 on MAL. when separated and assessed based on which handaffected the baseline ARAT score strongly predicted AOU for those with dominant hand affected (R2 = 0.6; F1, 17 = 25.518; P<0.001). Equation for this regression model calculates ARAT score of 46 required for AOU score 2.5. For non dominant affected hand, ARAT gross component score predicted 56.8% of variability in AOU (F1, 12 = 15.806; P = 0.002). regressionmodel shows that even if patients achieve max score on ARAT, 57, they will not score greater than or equal to 2.5 on AOU. The predictive power of the model was increased with FMA wrist component score added. AFTER TST, Changes in ARAT score predicted 30.8% of variability in change MAL AOU (F1, 28 = 12.486;P = 0.001). non dominant hand the change in grasp component of ARAT predicted 58.8% of change in the AOU (F1,11 = 15.674, P= 0.002).	? COSMIN UPDATE +(correlati ons with baseline amount ofuse) 1/1

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
ARAT	Lang, Wagner, Dromerick , Edwards 2006	to examine how ARAT scoresrelate to sensorimotor impairment measures, kinematic measures, disability scores, age and initial stroke severity (direction not stated). COSMIN update = 4 hypotheses generic #1 for FIM, kinematics (similar constructs) and # 2 related but dissimilar for NIHSS and sensorimotor (pain, spasticityand light touch)	1. percentage of missing items stated (5%) EXCELLENT , 2. notclear how missing items handledFAIR, 3. moderate sample size initially 50 but 40 at follow up FAIR, 4. multiple hypothesis stated prior EXCELLENT, 5. direction not stated GOOD, 6. expected magnitude not stated GOOD, 7. adequate description of comparator measures EXCELLENT 8. no info on measurement properties of comparator POOR 9. no other methodological flaws EXCELLENT 10. statistical methods appropriate EXCELLENT COSMIN UPDATE1 very good 2 very good 3 very good VERY GOOD	Poor - Excellent	construct validity examined via bivariate correlational analyses, relationships to other measures via Pearson product moment correlation. Pearson correlation coefficients (0.25 and below "low correlation", 0.26 to 0.50 "fair" 0.51 to 0.75 "good" greater than 0.75 "excellent) between age and ARAT total scoreswere Day 0: -0.16, day 14: -0.44, day 90: -0.29.NIHSS minimally related to ARAT.UE strength correlated with ARAT, the greater the strength the greater the ARAT, spasticity was inversely related to ARAT, greater spasticity lower ARAT r= -0.28 to -0.49, light touch unrelated to ARAT, pain scores showed a trend toward increasinglynegative correlations with ARAT at the 90 day time point . Kinematic measures; reach related to ARAT - faster, efficient and accurate reach were associated with better ARAT scores, graspsimilarly related to ARAT. ARAT and disability measures, FIM motor and ARAT: strength of relationship increased from day 0 to 14 and remained stable at day 90. ARAT and FIM UE score increased from day 0 to 14 and remainedstable at day 90. Conclusion ARAT is valid measure of UE functional limitation and may beused in acute stroke.	? UPDATED COSMIN# 1 x FIM 0.2 to 0.6 +, kinematics + #2 NIHSS - 0.15 to - 0.29 -, strength 0.4 to - 0.6 + light touch 0.2 - , spasticity - 0.3 to -0.5 + pain -0.3 + 7 hypotheses 5/7 71% correct. -
ARAT	Morris, van WijckJoice & Donaghy 2013	hypothesised that UL activityand limitation constructs andanxiety would emerge as significant predictors of HRQOL measured on the Nottingham Health Profile, plus hypotheses # 1 ARAT and UL RMA	1. missing data given 1.2% of data across 6 variables EXCELLENT 2. EXCELLENT 3. GOOD sample 85 4. minimal hypothesis GOOD 5. GOOD 6 EXCELLENT 7 EXCELLENT 8 EXCELLENT 9 a considerable proportion of the variance in totalHRQOL score not explained andmay have influenced findings, limited socio-demographics	excellentfair	ARAT and RMA high collinearity r = 0.8, ARAT and Nottingham Health Profile (NHP) total scorelow correlation r = -0.25 and physical activity score r = -0.39 and was therefore excluded fromthe regression analysis, RMA was used instead due to its correlation with NHP -0.30. The ARAt was negatively associated with the NHP indicating that greater UL dysfunction was associated with poorer HRQOL but overall studyfinding was the UL constructs did not predict HRQOL.	minus asdid not predict HRQOL. (1/2) correlation with likemeasure ARAT and ULRMA +

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
			collected to predict HRQOL, sample not representative of general stroke population FAIR 10 COSMIN UPDATE 1. very good 2 very good 3 doubtful DOUBTFUL			
ARAT	Rabadi & Rabadi 2006	testing interests correlation and correlation with UE selfcare function as per FIM -UPDTAED COSMIN #1 with FIM, #1 with FMA,	good 2 fair 3 excellent 4. fair 5 good 6 good 7 excellent 8 excellent 9 fair (treating therapistdoing FMA, Bias) 10 excellentCOSMIN UPDATE 1. very good very good 3 very good VERY GOOD	Fair - Excellent	spearman rank correlation coefficient comparedadmission and dc total scores the ARAT and FMA motor score correlated highly with one another on admission (0.77 P<0.00) and on DC (p 0.87 P<0.001). Changes in motor scores compared with LOS were similar between ARART (p 0.21 P = 0.7) and FMA motor score (p+ 0.14 P = 0.24). ARAT had a strong floor andceiling effect when compared to FMA motor. When compared to FIM both ARAT and FMAmotor showed a stronger correlation with FIMtotal and FIM -ADL sub scores for admissionthan for discharge scores (FIM & ARAT admission 0.33, dc .21	not doneCOSMIN UPDATE Hp#1 withFIM 0.33, 0.21 (-), #2 0.77 - 0.87 (+)
ARAT	Rand & Eng 2015	nil stated only aims 1) to compare the function and daily use of the upper extremities of individuals withstroke between the time of discharge to home and 12 months after stroke 2) predictthe daily use of the affected UE 12 mo post based on baseline assessment (time of dc)	1. EXCELLENT 2 EXCELLENT 3.FAIR sample n=32 4. FAIR 5 GOOD 6 GOOD 7 N/A 8 N/A 9. EXCELLENT 10. EXCELLENT COSMIN UPDATE 1. Very good 2 very good 3 very good	fair - excellent	when post stroke daily use assessed by accelerometers and MAL - age and gender werethe only demographic variables significantly correlated with 12 mo post stroke daily use. Correlation of ARAT with measure of daily arm use at 12 mo with MAL 14 r = 0.78, p 0.000 and with accelerometer r = 0.58, p 0.000 = significantpredictor of 12 mo post stroke daily use as measured with either MAL or accelerometer.Regression model - ARAT predicting 12 mo postuse when assessed with accelerometer R2 0.470 P 0.001, when measured with MAL R20.776 P<0.001. Summary ARAT at dischargecan predict use at 12 mo post stroke	? (1/1) +DC ARAT predictedUL use at12 mo

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
ARAT	Yozbatiran, Der- Yeghiain & Cramer2008	no specific hypothesis - statesto assess the reliability and validity of the currently presented method of ARAT scoring	1. excellent 2. good 3. poor (sample size) 4.excellent 5. Good 6. good 7. fair 8. excellent 9. excellent 10. excellent UPDATED COSMIN hypothesesgeneric 1. met. Convergent validity 1. very good 2 very goodvery good VERY GOOD	Poor - Excellent sample excludedfair - excellent	validity: ARATand arm Fugl-Meyer scores highlycorrelated (r=.94, P<.01) cannot be responsiveness as there is no clarity re: change score just the ARAT score?	? NEW COSMIN + as 1/1 hypothesis met.
ArmA	Ashford, Turner-Stokes, Siegert & Slade 2013	nil specific hypotheses presented "measures selected to allow comparisonto ArmA to test aspects of validity and reliability'. "comparing the passive and active function subscales of the ArmA with respective components of the LASIS and DASH.	1 very good 2 inadequate no - info on measurement propertiesof comparators 3 very good	inadequate	passive subscale of ArmA with passive items onLASIS Rho 0.50;p = 0.01 (convergent validity) but not with active function Rho 0.02 p=0.9 (divergent validity) "as expected". Active functionsubscale of ArmA correlated with LASIS Rho 0.48; p=0.01 and DASH active items Rho 0.63 p=0.01 but not with passive LASIS items Rho 0.23 p=0.078	hyp 1: passive ArmA withpassive LASIS 0.5 (+), hyp 2:passive ArmA withactive 0.02 (+) hyp 3 active ArmA withactive LASIS (0.48) (-) and hyp 4active ArmA withpassive LASIS 0.23 (+) - 3/4 75% +

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
BI	Ali, Fulton, Quinn, Brady on behalf of VISTA 2013	" we hypothesised that existing primary outcome measures (mRS, BI, NIHSS)may also reflect patient quality of life" COSMIN UPDATE USED SAME HYPOTHESES AS STUDY	1. Excellent 2. GOOD 3 excellent4 fair 5 good 6 good 7 Excellent 8 fair 9 excellent 10 excellent	fair - excellent adequate - marked down as nil referencet any measure ment properties within study	at 3 mo patient responses to QoL had a strongerassociation with mRS (SIS n=2970 P<0.0001, r = -0.71, r2 = 0.52, EQ-5D weighted score n=2987 r=-0.7 r2=0.53) where as proxy responses had a stronger association with BI (SIS n=867 P<0.0001, r = 0.68, r2 = 0.48, EQ- 5D n=837, r=0.78, r2= 0.63) than with NIHSS. BI had more mismatches between good primaryoutcome and poor QoL (EQ-5D 11.3%, SIS 19.2%) than the mRS (EQ-5D 8.5% SIS 10%) but less than NIHSS (29%, 23.9%)and for poorprimary outcome and good QoL less mismatches than mRS(EQ-5D 3.1%), (SIS 4.1%) but more than NIHSS (0%)	minus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
BI	Dennis, Wellwood& Warlow1996	nil given bit "…we present further data concerning their concurrent validity (or perhaps convergent constructvalidity since there is no gold standard)"	1, excellent n=2 missing data 2. good 3 excellent sample size 4 fair 5 good 6 good 7 good 8 good 9 excellent 10 excellent	fair - excellent	dependency" question: "do you/they require helpfrom another person with everyday activities?" sensitivity, specificity, accuracy (%) in predictingwhether pts scored above or below the following cut-offs; BI: <100 71, 98, 79; <85 89, 89, 89, <80 95, 86, 90. 'recovery' question "Do you feel that you/they have made a complete recovery from your/their stroke?" sensitivity, specificity, accuracy (%) in predicting whether pts scored above or below the following cut-off BI: =100 39,88, 71 >90 41, 92, 71. the sensitivity, specificity,accuracy of pts whose post stroke score was equal or better than pre-stroke score BI: 40, 90, 71. Recovery question was no more accurate at discriminating between those who were the same or better on BI and OHS after their stroke and those who were worse than it was at discriminating between those with good and poorfunctional outcome. 30 Pts scoring 100 on BI answered recovery question "no" 19 "yes" BI score of 95 5 answered "no" 5 answered "yes" BIscore of 90, 7 answered "no" 2 "yes" BI score of 85 8 answered "no" 2 "yes". 'dependency" question: pts scoring 100 on BI 49 "no" 1 answered "yes" all pts scoring 1 on BI answeredquestion "yes".	?
BI	Duncan, Lai, Tyler,Perera, Reker & Studenski 2002	"evaluate the validity of patient and proxy responses"	1. excellent 2 n/a 3 excellent 4 poor 5 good 6 good 7 poor 8 poor9 excellent 10 excellent	poor - excellent	BI and SIS mobility domain pt 0.69 proxy 0.70,BI and SIS ADL/IADL pt 0.72 proxy 0.78	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
BI	Filiatraultet al 1991	"an intercorrelation matrix was computed with the use of spearman's rho correlation to estimate level of association between the different pairs of variables" COSMIN UPATE generic hypotheses #1 with BIand FM and Functional Test for Hemiplegic/Paretic UE	COSMIN UPDATE 1. very good 2adequate 3 very good	adequate	correlation between BI and fugl-meyer test/functional test rho=0.60 and 0.61 p<0.01,,FM and function test 0.96, p<0.01 - BI less association with specific UL measures	plus 2/2
BI	Kwon et al 2004	non re: correlation of tools, second was vague about the tools ability to differentiate clinically distinct categories ofdisability similar to the MRS levels. COSMIN UPDATE generic hypotheses #1 for BI, FIM, MRS	COSMIN UPDATE 1. very good 2 very good 3.	very good	BI and m-FIM r_s=0.9479 p <0.0001, BI and MRS r_s=-0.8856 p <0.0001.BI discriminated disabilitybetter in lower levels of disability	plus (2/2)
BI	Wade & Hewer 1987	nil but states - this paper reports an analysis which establishes the frequency andseverity of paralysis after stroke, the validity of assessments used and extentof recovery made. COSMIN UPDATE 2 hypotheses at generic #1 MI and mBI	1. EXCELLENT 2 GOOD 3 EXCELLENT 4 POOR 5 GOOD 6GOOD 7 POOR 8 POOR 9 explicit data re assessors, analysis missing FAIR 10 coefficients is all that s stated GOOD COSMIN UPDATE 1. very good 2 very good 3 doubtfulmissing info	poor-excellent	total MI correlation coefficient with Barthel Indexwere 0.749 initially, 0.774 at 3 weeks and 0.610at 6 months. States that the coefficients for arm and leg scores similar but not given.	? COSMIN UPDATE + (1/1)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
BI	Wellwoodet al 1995	We aimed to examine the performance and the validity of the BI as a measure of disability after stroke. We set out to establish its concurrentvalidity by comparing it to a 'gold standard' measure of disability [3]. The measure wechose for this purpose was the Office of Population Censuses and Surveys (OPCS) Disability instrument [4] which was developed to provide a comprehensive measure of disability for use in the 1985 survey of disability among adults, manyof whom had stroke-related disability. COSMIN UPDATE generic hypotheses #1	1. very good 2 doubtful 3 adequate - very limited detail	doubtful	correlations between BI and OPCS -0.73 p<0.001 (OPCS - Office of Population Censusesand Surveys disability instrument - comprehensive measure of disability; locomotion, reaching and stretching, dexterity, personal care, continence, seeing, hearing , communication, behaviours consciousness, eating drinking, disfigurement, intellectual functioning	plus (1/1)
BI	Lai, Studenski, Duncan & Perera 2002		1. GOOD - reports some variables had missing data but unable ton determine which via tables 2. FAIR 3. GOOD n=81 4.POOR 5. GOOD 6 GOOD 7 n/a 8. n/a 9. FAIR - minimal detailgiven re methodology is not transparent 10. EXCELLENT	poor - excellent	pts scoring greater than or equal to 95 on the BI(deemed recovered or minimal disability) 3 months had mean hand function scores 9 pointslower, social participation 12.8 points lower and ADL/IADL 5 points lower than non stroke community dwelling participants.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
BI(C&W)	Barer & Murphy1993	not reported, looked at association between ADL score at one week and length of stay - broken into patients who died, dc to NH and thosestill in hospital COSMIN UPDATE - very difficult to assign hypotheses - 1. abilityt predict dc home and care (r≥0.50)	inadequate as BI completed informally from a range of sources - information from wardstaff or carers) no info on who collected, bias, data not presented as to who dies versusinto care - unable to determine specific results	inadequate	correlation with dc home and dc into care or diedall greater than 0.50	plus (2/2)
BI(C&W)	Dennis etal 2000	nil reported COSMIN UPDATE generic hypothesesGeneral Health Questionnaire #3 , HADS anxiety and depression #3, MRS #2,	Poor; no hypothesis and no info on /comparator instruments. Alsounsure as to what version of the BI was used - assume Collin and wade as British. COSMIN UPDATE 1. inadequate 2 inadequate 3 very good	inadequate	BI and GHQ =rs -0.4, BI and HAD - anxiety =rs - 0.2 HAD depression rs-0.4 and modified rankinscale =rs-0.8 all P<0.01. BI has stronger relationships of tools assessing disability and physical abilities.	4/4 +
BI(C&W)	Sarker,Rudd, Douiri, Wolfe 2012	"the aims of this present studyare to 1) compare the FAI NEADL and BI in terms of distribution of scores, concurrent validity, reliability, and their agreement. COSMIN UPDATE generichypotheses #1 mBI with NEADL and FAI	Poor, no discussion of constructs/mp of the tools. Also not clear what BI was used, no reference to the tool. Assume Collin and Wade as British StrokeStudy. Results indicate that the BIand FAI have a strong agreement rs=0.80 and BI and NEADL have a strong agreement rs=0.88. BI NEADL and FAI have some overlapping constructs COSMIN UPDATE 1. Very good 2. doubtful - ref to study, very minimal descriptions of evidenceor constructs 3 very good	doubtful	Results indicate that the BI and FAI have a strong agreement rs=0.80 and BI and NEADLhave a strong agreement rs=0.88. BI NEADL and FAI have some overlapping constructs	plus (2/2)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
BI(C&W)	Wilkinsonet al 1997	"The objective of this presentstudy was to consider whether the Barthel Index alone provides sufficient information about the long term outcome of stroke. Thisshould help decide which outcome measures are the most pragmatic and appropriate for assessing thelong term outcome of stroke patients" COSMIN UPDATE generic hypotheses mBI with SF-36 #1, NEADL #!, MRS #1, FAI#1, LHS#1, Caregiverstrain #1, Life satisfaction index #2, anxiety and depression #2	**Fair;** Hypothesis not formulated, Poor information on the mp properties of comparator tools. Also not clear what version of theBI was used, assume Collin and Wade as a 20 point scale used. COSMIN UPDATE 1. very good 2. doubtful 3 very good	doubtful	The Barthel index correlated stronger with assessment tools of disability, such as the LHSrs=0.726 p<0.001; FAI rs=0.826p<0.001; SF-36 physical rs=0.810; social functioning rs=0.481 p<0.001 role physical rs=0.415; vitality rs=0.5 p<0.001; General health rs=0.438 p<0.001 NHPenergy rs=-0.605 p<0.001 pain rs=-0.499 p<0.001 emotion rs=-0.423 p<0.001 and social interaction rs=-0.460 p<0.001 and physical mobility rs=-0.40 p<0.001. BI had less strong relationships with HADS anxiety rs=-0.187 P>0.05l LSI total rs=0.361 p<0.001; LSI acceptance rs=0.307 p<0.01; SF-36 MH rs=0.332 p<0.001; Bodily pain rs=0.356 p<0.001;sleep rs=-0.189 p>0.05	8/8 plus
CMA	Dang, Ramsaran , Street, Syed, Barclay-Goddard, Stratford & Miller 2011	nil hypothesis but aim to estimate the predictive accuracy and clinical usefulness of the CMSA predictive equations for pts with stroke undergoing rehabilitation COSMIN UPDATE hypotheses AI andII shrinkage values closer to <0.10 to meet hypothesesthat predictive models are valid	1. yes EXCELLENT 2. GOOD 3GOOD sample size 4. FAIR unclear what was expected i.e. ifsupported or refuted 5 GOOD 6 GOOD 7 n/a 8 n/a 9 EXCELLENT 10 EXCELLENT COSMIN UPDATE 1 na 2 na doubtful - raters experience, limited detail what constituted specifically what was acceptableresult for prediction equations	fair - excellent doubtful	shrinkage values (<0.10 indicate a reliable model) for Impairment inventory varied from -0.05 to 0.09, AI 0.21, GMFI 0.19, WI 0.24. addition the 95% prediction bands are approximately +/- 1.5 stages for II and +/- 24 point s for AI- too large/large error associated with predictions, to provide a confident prediction or clinically useful information - data does not support Gowlands predictive equationsfor AI but does support II..	minus(1/2) ?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
CMSA	Gowland, Stratford, Ward, Moreland, Torresin, Van Hullenar, Sanford, Barreca, Vanspall & Plews 1993	construct validity - "specific impairments and disabilities would have the highest correlations with similar attributes on other measures and those correlations would be significantly greater than 0.60" concurrent validity " the magnitude of the correlations of the total scores would be significantly greater than 0.60" it was also thought that this would adequately expressa positive correlation. COSMIN UPDATE #1 CMSA impairment subscales with FM subscales #1 CMSA disability with FIM subscales#1 CMSA total with total FIMand FM >0.60 as per study hypotheses	1. Good 2. n/a 3. Fair n=32 for this section 4. excellent 5. excellent 6 excellent 7. FAIR 8 FAIR references only given 9 excellent 10 excellent COSMIN UPDATE 1very good 2 adequate3 very good 4 very good	fair - excellent adequate	1. CMA impairment inventory arm and hand correlated with Fugl-Meyer shoulder, elbow, forearm, wrist and hand 0.95 and CMA impairment inventory total score correlated with Fugl-Meyer total score 0.95 other items correlated well with their respective counterpartsin the FIM and Fugl-Meyer supporting construct validity and total scores correlated well 0.95 ($p<0.001$) for CMA and fugl-Meyer and disability and FIM 0.79 ($p<0.05$) supporting concurrent validity. Both values exceeded the 0.60 minimum acceptable level established a priori.	plus COSMIN UPDATE 4/4 +

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
DAS	Brashear, Gordon, Elovic, Kassicieh, Marciniak,Do, Lee, Jenkins, Turkel, 2002	nil, conducted a trial to assess the effects of one setif injections with botulinum toxin A on measures of disability with respect to self care, limb position, and painas well as on muscle tone.	1. yes 4 in control group EXCELLENT 2. FAIR nil information given 3. GOOD sample size (appraised on individual sub group 64 and 58) 4. FAIR nil hypotheses stated butrefer to previous correlations andstudies looking at impact on functional improvement 5 GOOD6 GOOD 7 GOOD 8 EXCELLENT 9 EXCELLENT 10 EXCELLENT	fair - excellent	6 weeks (chosen as primary endpoint) DAS score for the chosen therapeutic target (either hygiene, dressing, limb position or pain - almost1/3 selected limb position otherwise participantswere more concerned with dressing, limb position and hygiene than they were about pain)correlated with composite score on ashworth scale for muscle tone in wrist and fingers (r=0.61, P<01), the score for the physicians global assessment (r = -0.46, P<0.001) and the score for the patients or caregivers global assessments r = -0.51, P<0.001. DAS capture change in areas important to patient and caregivers and reflective of clinicians assessment of change. may be a valid measureof functional disability in patients with spasticity undergoing botulinum therapy.	?
DAS	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	greater disability is associatedwith worse HRQoL and a greater need for caregiver assistance. COSMIN UPDATE generic hypotheses#2 QoL, caregiver assistance #1	1. EXCELLENT 2. N.A 3. EXCELLENT n = 279 4. GOOD 5. EXCELLENT 6 GOOD 7. EXCELLENT 8 FAIR reference tostudy 9 EXCELLENT 10 GOOD marked down due ot ordinal nature of DAS not accommodated in one way analysis of variance. COSMIN UPDATE 1. very good 2 very good 3 adequate	fair	greater DAS was significantly associated with higher SA-SIP physical dimensions (P<.05). Increasing disability is associated with reductionin HRQoL and caregiver burden (P<.05)	plus 2/2met

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
EQ-5D	Aldermanet al 2001	characteristics of neurological populations may significantly reduce the validity of the EQ- 5D, this will be low because ofreliance on self-report, poor ability to reflect rehab gains through limited range of ratings and inadequate number of items failing to capture complexity of needs. COSMIN UPDATE generic hypotheses #1 EQ-5D with DEX and #1 EQ-5D with BI	1 good 2 fair 3. poor 11 4 fair 5 good 6 good 7 fair 8 fair - reference to study that gives same 9 - 10 excellent COSMIN UPDATE 1. very good 2adequate 3 very good	fair - adequate	Staff completed EQ-5D Index and EQ-5D VAS with BI .48, p<.001, 0.33, p=.02, no relationshipbetween DEX-O and EQ-5D. Individual items of EQ-5D with DEX-0 ranged from -0.02 to 0.29 and with the BI 0.04 (anxiety and depression) to -0.55. Staff completed and participant completed EQ-5D no significant differences, however participants did perceive their health state had improved (t = 1.9, p = .43).	plus COSMIN UPDATE staff - (0/2), pt - (0/2)
EQ-5D	Barton2008b: comparison of EQ- 5D and EQ6D >/=45	In addition to seeing to confirm that, for other patient groups, healthier individuals tend to have higher scores onthe EQ-5D and less healthy individuals higher scores on the SF-6D, we also assess whether these differences as significant and could be considered to constitute a MID, and seek to ascertain where the cross-over point of scores on the EQ-5D and SF-6D might be. hypotheses #1 and #5	COSMIN UPDATE 1. very good 2 adequate 3 very good 5 adequate4 very good + (2/2)	adequate	Correlation between EQ-5D and SF-6D (table VI p. 825) ICC 0.628, Pearson 0.769. (n=51 for thisanalysis).	plus agreement between measuresand correlations) 1

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
EQ-5D	Barton, Sach, Doherty, Avery, Jenkinson , Muir et al 2008a dsicrim ability nowhynes	We hypothesised that HRQLmeasures would be able to discriminate in accordance with differences previously observed in the literature. Namely that HRQL would vary according to age, gender, smoking status, ethnicity, body mass index, education, social class, economic status, housing tenure, income and the presence of health problems... (back pain, hip pain, knee pain, heart disease, stroke, asthma, cancer, diabetes, rheumatoid arthritis and osteoarthritis.) COSMIN UPDATE generic hypotheses #5	COSMIN UPDATE 1 very good 2very good	very good	Researchers specified that a difference of >0.03for the EQ-5D index and >3.0 for the VAS would constitute a MID. People with a stroke hada mean (SD) Index of 0.612 (0.318) which was less (worse) than the rest of the sample (people registered at a general practice) whose index scores were 0.784 (0.234). The VAS mean (SD)for stroke also significant difference (p<0.001) was 65.83 (20.01) compared with others 75.88 (16.87). This indicates there was a MID betweenpeople with stroke and people without. These differences were all statistically significant (p<0.001). The direction and magnitude of the expected differences between people with strokeand the remainder of the sample was not specified.	plus (1/1)
EQ-5D	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	greater disability is associatedwith worse HRQoL and a greater need for caregiver assistance. COSMIN UPDATE generic hypotheses#2 QoL, caregiver assistance #1	1. EXCELLENT 2. N.A 3. EXCELLENT n = 279 4. GOOD 5. EXCELLENT 6 GOOD 7. EXCELLENT 8 FAIR reference toa study 9 EXCELLENT 10 GOOD marked down due ot ordinal nature of DAS not accommodated in one way analysis of variance. COSMIN UPDATE 1. very good 2 very good 3. adequate	fair adequate	Increasing disability in DAS was associated withdiminishing EQ-5D index scores (P < .002) Increasing disability is associated with reductionin HRQoL and caregiver burden (P<.05)	plus 2/2

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
EQ-5D	Dorman 1997 "is euroqual a valid'	From abstract: "The Euroqol measures aspects of quality of life that are highly relevant to stroke patients. It is short and simple and many stroke patients can complete the form without help. However itsvalidity has not been adequately assessed after stroke. We therefore assessed its concurrent and discriminant validity in a group of prospectively studied stroke survivors. COSMIN UPDATE hypotheses generic #1 for convergent validity withFAI, pain and HADS and discriminant validity with generic hypotheses #5	COSMIN UPDATE 1.very good 2doubtful 3 very goof	doubtful	Mod-strong correlations with comparator instruments for mobility, self care, usual activitiesand pain (r= 0.61, 0.64, 0.60 and 0.71 respectively). Moderate to weak correlation between psychological functioning and comparator (r=0.56 and 0.35). Able to distinguish between baseline stroke severity andtype of stroke.	6 hypothes es at #1 5/6 met + discrimant validity + overall 6/7 +
EQ-5D	Dorman, Dennis & Sandercock 1999	since both instruments aim to measure HRQoL there shouldbe a strong correlation between responses on the 2 instruments.	1. very good 2 adequate 3 adequate	fair adequate	Moderate correlations between EQ-5D mobility, self-care and usual activities with SF-36 physicalfunction (r=0.57, 0.65, 0,63 respectively). Moderate correlation (r=0.66) between EQ-5D pain and SF-36 bodily pain. The correlation between patient responses to the psychological functioning domain of the EQ-5D and the mentalhealth domain of the sf-36 was weak (r=.21) but there was a stronger relationship with general health (r=0.44). The EQ-5D VAS was moderately correlated with SF-36 general health r=0.66	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
EQ-5D	Fisk, Brown, Sketris,Metz & Stadnyk 2005	For [construct validity] we compared three health utility measures with clinician ratings of neurological disability and with objective measures of upper and lowerextremity task performance (p59). Lower correlations are expected for instruments that measure different constructs using data collection methods, higher correlations are expected for similar constructs using similar methods, with moderate correlations expected for similar constructs measured by different methods or similar methods that measuredifferent constructs. Using these guidelines, we considered correlations less than 0.3 as weak evidence ofvalidity. correlations of between 0.3 to .59 as moderate and correlations above 0.59 as strong (p 60). COSMIN UPDATE generic #1with NHPT, ambulation total EDSS, HUI III, SF-6D 5 hypotheses	COSMIN UPDATE 1. very good 2adequate 3 very god	adequate	EQ-5D with EDSS -0.66, ambulation index -0.68,NHPT -0.56, HUI III 0.80, SF-6D 0.70	1

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
EQ-5D	Gillard etal 2015	Spasticity has been hypothesised to have a significant negative impact onthe health-related quality of life of stroke survivors. To date, the published evidence to support this negative association is limited.' & 'we studied whether spasticity hasa negative impact on health related quality of life' COSMIN UPDATE generic hypotheses #5	COSMIN update 5. very good 6adequate	adequate	Lower EQ-5D scores were reported by those with spasticity than without at all time points (3 months, 1 year, 2 years) (poor sample size in some analyses). Mean difference in EQ-5D between those with spasticity and those without is -0.07 (CI -0.12 to -0.03). This is equal with theMCID established for patients with other health conditions (Stroke MCID has not been established).	plus (1/1)
EQ-5D	Kohn, Sidovar, Kaur, Zhu& Coleman 2014	The purpose of this study wasto estimate the minimal clinically important difference (MCID), or smallest differencein score PwMS from Nth America perceive as being both beneficial and nontrivial' COSMIN UPDATE correlations with Patient determined disease step (PDDS) measures 0 = no disability to 8 = bed bound #1,MSW12 - measure of mobility#3	COSMIN UPDATE 1. very good 2 very good 3 very good 4 adequate - only 65& returned and80% female	adequate	Moderately strong correlations between the EQ-5D and the PDDS and the MSWS-12 were observed (Spearman's r=-0.56 and -0.59 respectively, p<0.0001 for both.)	EQ-5D with PDSS plus, walking too stronga correlatio n - (1/2) overall -

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
EQ-5D	Kuspinar& Mayo 2013	The global aim of the study isto contribute evidence for thecontent validity of generic utility measures with respect to capturing the relevant domains for people with MS. The specific objective was to estimate the extent to which generic utility measures capture important domains that are affected by MS.' There were no specific hypotheses about correlation,size or direction. COSMIN UPDATE generic hypotheses#1 with SF-6D and PGI	1 very good 2 adequate 3 adequate	adequate	EQ-5D and SF-6D 0.58 Patient generated index(PGI) and EQ-5D 0.53	plus (2/2)
EQ-5D	Kuspinar, Finch, Pickard &Mayo 2014	the discriminant ability of theP-PBMSI index and the EQ-5D was tested against different clinical subgroups	1. yes 2. n/a 3. excellent 4. Poor 5. n/a 6 n/a 7. excellent 8 excellent 9 excellent 10 excellentCOSMIN UPDATE 5 very good very good	poor very good	EQ-5D discriminates between different clinicalsubgroups, functional walking capacity and general health perception.	? Plus(1/1)
EQ-5D	Moore, Wolfson, Alexandrov & Lapierre 2004	purpose of study to determinewhich measure would be more useful in clinical practice COSMIN UPDATE generic hypotheses #1 EQ-5D with PtQOL rating and with pt severity rating and with EDSS, #1 3 hypotheses	1. 3% 2. n/a 3 excellent 4 fair 5n/a 6 n/a 7 excellent 8 excellent 9. fair - excluded even mild cognitive impairment, 10 excellent COSMIN UPDATE 1.very good 2 inadequate 3 very good 4 very good	fair INADEQUATE	Participants felt completing all 3 measures wasmost reflective of their QOL (EQ-5D, SF-36, MSQOL-54 - all correlated similarly). EQ-5D predicting EDSS OR 0.07 (95% CI 0.01 - 0.35). Correlations with "Pts QOL rating" 0.49 ((%%CI0.35, 0.72), with pts severity rating 0.36 (0.18, 0.58), EDSS -0.56 (-0.82, -0.44).	1/3 (-) >0.5 withEDSS only

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
EQ-5D	Nicholl etal 2001	The aim of this study was to assess whether this short generic scale is equivalent tothe longer MS specific QoL scales, or to the longer generic SF-36. COSMIN UPDATE generic hypotheses EQ-5D with SF-54 #1, EQ-5D with FAMS #1	1. Good 2 good 3 good 4 fair 5 good 6 good 7 good 8 poor 9 excellent 10 excellent COSMIN UPDATE 1. very good 2 doubtful3 very good	doubtful	EQ-5D did not correlate as highly with other QOLmeasures (FAMS and SF-54) subscales, was found to be less sensitive.	? - (0/2)
EQ-5D	Xie et al2006	We sought to quantify the national impact of stroke on HRQoL in the non-nstitutionalised population in the United States with four commonly used HRQoL measures. COSMIN UPDATE generic hypotheses #5 - discrimant between groups	COSMIN update 5. very good 2.very good	very good	After controlling for age, gender, race, geographic region, risk factors and comorbidities, Stroke survivors had on averagea 6.9% lower index and 7.2% lower VAS than other civilian, non-institutionalised Americans.	plus (1/1)
FIM	Brown, Therneau, Schultz, Niewczyk & Granger2015	"…to consider a wide range ofclinical elements known at impatient rehab admission to identify those that predict functional gain, length of stay and discharge to home" COSMIN UPDATE hypotheses #1 predicting FIMgain, #1 predicting LOS #1 predicting dc destination all	1. good 2 n/a 3 excellent sample4 poor unclear what was expected 5 good 6 good 7 excellent 8 excellent 9 excellent10 excellent COSMIN UPDATE 1. very good 2 very good 3 verygood	poor very good	FIM motor admission and walking distance greatest effect for prediction outcome FIM gain with a range of 9.1 and 8.9 FIM points. FIM motor admission greatest effect on LOS and probability of dc home. Most clinically relevant 3-variable model was FIM motor sub score, age and walking distance r2 = 0.107.	? + (3/3)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
		hypothesised ES 0.5 or greater				
FIM	Cullen, Krakowski& Taggart2014	logistic analysis was used to explore whether acute neurological function post TBI(GCS) and baseline cognitiveand functional ability at rehabadmission (FIM and DRS) predict whether or not patients return to driving following TBI COSMIN UPDATE hypotheses FIM willpredict return to driving betterthan GCS and GCS	1. excellent 2 excellent 3 good sample size 4 poor - unclear whatwas expected 5 good 6 good 7 excellent 8 excellent 9 excellent 10 excellent COSMIN UPDATE discriminative 1. very good 2 verygood 5very good 6very good	poor very good	Pts who returned to driving had higher FIM scores at rehab admission (2 mo post TBI) (100.09 +/- 16.95 vs 83.38 +/- 23.70 p = 0.006. OR = 1.07, p<0.01 only the FIM score made a significant unique contribution to return to drivingpost TBI - GCS or DRS did not significantly contribute. Sensitivity of FIM to predict RTD 72%, specificity 73%, ROC 0.71.	? COSMIN UPDATE plus (2/2)
FIM	Cuthbert, Harrison-Felix, Corrigan, Bell, Haarbauer Krupa & Miller 2015	"…to assess the prevalence of unemployment 2 years post injury ….as well as the factorsassociated with an increased prevalence of unemployment." COSMIN UPDATE hypotheses #1 FIM predictor of employment at 2 years	1. excellent 2 excellent 3. excellent sample size, 4 poor notclear what was expected 5 good6 good 7 excellent 8 excellent 9 excellent 10 excellent COSMIN UPDATE adequate - missing data not accounted for possible Influence on results	poor adequate	FIM cognitive was not significantly associated with unemployment. Risk of being unemployed as compared with employed at 2 years PR (prevalence ratio) 0.99 (95%CI 0.99-1.00) P =.13risk of being employed part time as compared with full time FIM PR 0.97 (95%CI 0.93 - 1.01) P = .15	? - (0/1) FIMc nota predictor

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
FIM	Egan, Davis, Dubouloz, Kessler & Kubina 2014	there is a reciprocal relationship between poor emotional well-being and participation - low FIM shouldbe associated with high participation (via Reintegration into Normal Living Index) #1	1. excellent 2. excellent 3 GOOD sample size 4. GOOD 5 EXCELLENT 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 FAIR - less representative sample no cognitive impairments and very high level functioning 10 EXCELLENT COSMIN UPDATE doubtful - sample only 35% eligible participated	fair doubtful	when FIM scores are high, health was not related to participation, when FIM low, good health was associated with increased participation. FIM moderated physical wellbeing,not emotional well being - the effect of participation on change in physical wellbeing was marginal	plus marginal results
FIM	Grant, Goldsmith& Anton 2014	no specific hypothesis but study aims to retest previous findings that LOS is influenced by motor functional independence	1. excellent 2 excellent 3 excellent sample size 4 Fair hypotheses 5 good 6 good 7 n/a8 n/a 9 excellent 10 excellent	fair-adequate	admission FIM motor score, age and geographicregion (in Canada) best predicted rehab LOS, explained 20% variability.	? (0/1) as FIM only accounted for 20%of variance in LOS proposingother variables other thanfunction contributeto prediction
FIM	Hall, Hamilton, Gordon & Zasler 1993	COSMIN UPDATE all hypotheses #1 FIMm withDRS, FIM+FAM motor, FIM+FAM cognition, FIMc with DRS, FIM+FAM motor, FIM+FAM cognition		?	FIM Motor correlated with Disability rating scale (DRS r=0.641), FIM and FAM motor (r=0.992) and FIM and FAM cognition (r+0.653); FIM cognition with DRS (r=0.728) FIM and FAM motor (r=0.645) and FIM and FAM cognition(r=0.635).	plus (6/6)
FIM	Heinemann et al 1997	nil specific given - aims are 1)describe the nature of nurse contact hours 2) evaluate the concurrent validity of FIM by examining correlations between nurse-patient contact and FIM scores COSMIN update	inadequate - a lot of variance notcaptured, many professional activities undertaken by nursing staff may have been missed as only admission and discharge data was collected not what occurred in the interim	inadequate	more than half the variance in contact times is not explained by the FIMm measure. Wk 1 measures time spent in and correlation with FIMm medication -0.39, <.001, treatment -0.59, <.001, teaching -0.62, <.001, social 0.13, NS, indirect -.04, NS, other -0.1, NS. last week correlations medications -0.54, <.001, treatment - 0.50, <.001, teaching -0.63, <.001, social 0.26, .003, indirect -.06,	plus 1/1

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
		generic hypotheses #1: 1) FIMm and time correlations >-0.5, 2			NS, other -0.08, NS. Summary majority >-0.50	
FIM	Kuys, Bew, Lynch, Morrison & Brauer 2009	research question: which measure of activity limitation on admission to rehabilitation after stroke best predict walking speed at discharge COSMIN UPDATE hypotheses - FIM would be a predictor of DC walking speed as per correlations and significance	1. EXCELLENT 2. n/a 3. EXCELLENT 4. FAIR 5. GOOD 6. GOOD 7. GOOD 8 EXCELLENT 9 EXCELLENT 10 EXCELLENT COSMIN UPDATE 1. adequate 2 very good 3 very good	fair - excellent adequate	when significant predictors of discharge walking speed (admission walking speed 10MWT, modified elderly mobility scale score, MAS item 1 to 5, FIM motor component, FIM and TUG, were entered into multiple linear regression, discharge walking speed was best predicted by 10MWT at admission and MAS item 2 (supine lying to sitting over side of bed) R2 = 0.36. (accuracy of prediction). clinicians could predict using equation: discharge walking speed m/s = 0.33 + 0.47 admission walking speed + 0.05 Item 2 MAS score. OTHER: admission walking speed (10MWT) (n=120) relationship with discharge walking speed (from univariate analysis) 0.32 (<0.001), MAS item 6 (n=105) 0.09 (0.14) item 7 0.06 (0.23) item 8 0.07(0.36), FIM 0.25 (>0.001).	? Minus (0/1) r = 0.25 did not Predict walking speed
FIM	Madden, Hopman, Bagg, Verner, O'Callaghan 2006	hypothesised that variables describing similar facets of function or quality of life (e.g. the SF-36 PF domain with theFIM mobility and locomotion domain) would have correlations of ≥ 0.25 COSMIN UPDATE generic hypotheses #1	1. very good 2 very good 3 verygood		nil correlations between SF-36 and FIM ≥ 0.5	0/1) -

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
FIM	Oczkowski& Barreca 1993	nil hypothesis "this article further clarifies the use of theFIM as a prognostic indicator of outcome in stroke patients"COSMIN UPDATE hypotheses - FIM predicts DC destination	1. EXCELLENT - 3 died 2. FAIR3 EXCELLENT 4 FAIR 5 GOOD 6 GOOD 7 POOR 8 POOR 9 EXCELLENT 10 EXCELLENT COSMIN UPDATE adequate - no detail re assessors and	poor - excellent	The degree of recovery of postural control as measured by the CMA highly correlated with admission FIM. Admission postural control via CMA correlated highly with discharge location. Postural control at admission (median stage) 495%CI 4-4.5 predicted dc home, stage 3 95%CI2.5-3.5 predicted nursing home and 1.5 95CI% 1.0-2.0 predicted chronic care. multiple logistic regression showed best predictors of location ofdc were admission FIM score, admission postural staging (CMA) and age, the most powerful was admission FIM score.	? Plus(1/1)
FIM	Ouellette, Timple, Kaplan, Rosenberg & Rosario 2015	no hypotheses but "…purpose of this study was to identify if diagnostic specific (S-STREAM) and general rehabilitation outcomes (FIM) measured at admission can predict a discharge destination from inpatient rehabilitation.	1. no GOOD 2. n/a 3. Excellent sample 4 fair (2 part as they then had hypotheses re: sensitivity/specificity of tool) 5 GOOD 6 GOOD 7 excellent 8 excellent 9 fair - limited details 10 excellent COSMIN UPDATE doubtful - limited detail	fair doubtful	Logistic regression identified a cut off score of 29 and above using FIM and S-STREAM prediction scale predicted discharge to community from inpatient ($X^2 = 69.4$, $P<0.001$, AUC = 0.76, sensitivity = 0.76, specificity = 0.64).	? Plus (1/1) FIM predict dc with S-STREAM
FIM	Rabadi & Vincent 2013	nil "we set out to compare the EDSS to the FIM scale as an effective measure of MS- related disability and to determine which of the two scales is more responsive to clinical change following treatment. COSMIN UPDATE hypotheses #2 FIM and EDSS	1. good 2. n.a 3 good sample size 4 poor not clear what was expected 5. good 6 good 7 excellent 8 excellent 9 fair - not a representative sample 10 excellent COSMIN UPDATE 1. very good 2 very good 3 adequate	poor adequate	concurrent validity for assessing MS related disability: EDSS and FIM at initial evaluation rs = -0.69, $p<0.0001$ when controlling for variables known to influence MS-related disability (age, disease duration, gender, race and MS type) rs = -0.51 $p<0.00010$. FIM and Impairment Inventory rs = 0.51 $p<0.00010$	? COSMIN update - as >0.30 to 0.5 change to plus as EDSS global measure of disability not Impairment

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
FIM	Sharrack, Hughes, Soudain & Dunn 1999	hypotheses generic #1 for all except impairment #2		adequate	FIM correlated with BI(C&W) r=0.88 (ability), EDSS r=0.74 (disability) SNRS r=0.69 (impairment) AI r=-0.72 (disability) and CAMBS r=-0.69 to -0.61 respectively (disability)and SF- 36 (r=0.88) (function) (all p<0.001). FIM correlated with work r=-0.59 (p<0.001), housework r=-0.64 (p=0.001), independence (r=-0.44 p=0.001) and disability rank r=-0.96	8/10 (too high with Impairment and too low independence) +
GAS	Brock, Black, Cotton, Kennedy, Wilson & Sutton 2009	nil specific priori but aims stated 1. determine level of goal achievement at 6 mo post rehab and correlate with perceived levels of participation 2. determine the factors associated with higher goal achievement COSMIN UPDATE 5 hypotheses GAS with LHS FIMm MMSE, mood, self efficacy all at #1 as comparing constructs within GAS to other measuresof same constructs	1. yes specific assessments had varied sample size as per detailed in tables EXCELLENT 2.GOOD 3. FAIR sample size 4. FAIR hypotheses 5 GOOD 6 GOOD 7 POOR 8 FAIR 9 FAIR -nil detail re blinding of raters, slight variations to administrationof standardised tools but neededdue to sample cognition and communication 10 UNSURE used median goal achievement score COSMIN UPDATE 1 inadequate 2 doubtful 3 inadequate	fair - excellent inadequate	goal achievement via GAS with London Handicap Scale (LHS) (spearman rho) getting around -0.47, work and leisure -0.48, getting onwith people -0.52, awareness of surroundings -0.47, looking after yourself -0.48. Summary - GAS scores moderately correlated with perceived activity and participation (between - and -0.51, p<0.005). Nil significant correlations between GAS and initial measuresat discharge. 6 mo post median GAS score withFIM motor 0.55, self efficacy (SUPPH- coping) and depression (CES-D) 0.46. Higher goal achievement associated with higher mobility andless depression and better self-efficacy. nil other moderate correlations	? COSMIN UPDATE LHS (0.47-.052) -, FIMm +, MMSE +, mood +, self efficacy + correlatedwith constructs within goals on other measures - too low with participation only 4/5
GAS	Doig, Fleming, Kuipers & Cornwell 2010	determine agreement between participants and theirsignificant others on perceived change in performance (COPM and GAS) COSMIN UPDATE #1GAS and COPM agreement	1. yes EXCELLENT 2 good 3 sample 14 POOR 4. FAIR 5 GOOD 6. GOOD 7 EXCELLENT8 EXCELLENT 9 EXCELENT 10FAIR percentage agreement withno clear hypothesis COSMIN UPDATE 1. very good 2 very good 3 adequate	fair	70% agreement between participants and theirsignificant others on the direction of change inCOPM performance which corresponded with objective GAS rating.	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
GAS	Joyce, Rockwood& Mate- Kole 1994	nil hypotheses but question; is GA scaling a feasible method of program evaluationin post acute brain injury patients and what are the measurement properties of GA scaling in these patients? COSMIN UPDATE hypotheses 1 # with Globalclinical impression and 5 other ADL scales #1	1. good 2 n/a 3. poor sample size, 4. poor - unclear what was expected 5. good 6 good 7 poor 8poor 9 fair - no talk of blinding or training of assessors or additionaldetail re participants to determine generalisability also measures reported to be used in texts don match with names on tables?? 10excellent COSMIN UPDATE 1. very good 2 adequate 3oubtful	poor - excellent	Pearson correlation with global clinical impression 0.8031 (high) , but modest or weaklywith the following; IADL 0.1611, Milwaukee evaluation of daily living skills 0.5026, QL indexSpitzer quality of life index 0.2186, Rappaport disability rating -0.6162, Kohlman evaluation of daily living skills - 0.0039. Paired t test comparing GAS scores at admission and discharge with scores from standard outcome measures; GAS -9.30 (<0.0001), Rappaport disability rating scale 3.74 (0.002), Milwaukee - 4.13 (0.001), instrumental activities of daily living -1.33 (0.20), Kohlman evaluation of daily living - 3.61 (0.003), Spitzer QoL -7.20 (<0.0001). Summary; 3. GA scaling change scores correlated highly with global clinical impression (clinical judgement of efficiency) (r= 0.8061) and weakly with standard performance measures (r= -0.0039 – 0.5026). Paired t test comparing GAS scores at admission and d/c with scores from standardised outcome measures GAS t = -9.30, RDR: 3.74, MEDLS -4.13, IADL = -1.33, KELS =3.61 QL index -7.20.	? COSMIN UPDATE 4/7 -
GAS	Khan, Pallant &Turner-Stokes 2008	nil specific but objectives (1) explore the type and nature of person-centred goals that arecommonly set and achieved during the program (2) compare the responsiveness and relative efficiency of GASwith the FIM and BI as outcome measures for rehab in MS COSMIN UPDATE	1. not stated GOOD 2 N/A 3. POOR sample size 4 POOR knew what they were doing statistically but didn't know what they expected to find 5 GOOD 6 GOOD 7 GOOD 8 EXCELLENT 9excellent 10 FAIR - SRM and Effect sizes COSMIN UPDATE 1.very good 2 very good 3 very good	poor - excellent	spearman rank correlations between measuresGAS T-score at d/c; with GAS change score 0.95, with BI change score -0.25, with FIM change score -0.16, CGI -0.86. GAS change score with BI change score -0.15, with FIM change score -0.06 with CGI -0.77. Over half ofthe goals chosen were in areas not included in the FIM or BI.	? GAS and BI (-.25) - , FIM -0.16(-) CGI (0.86) + (2/3) -

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
		hypotheses GAS and FIM #1,GAS and BI #1 clinical global impression scale #1 as per generic hypotheses.				
GAS	Malec1999	nil specific stated but "…thefollowing psychometric properties will be reviewed appropriate correlational analyses with other outcome measures" other references made to expected outcomesin text	1. yes 54/88 only completed Mayo Portland assessment EXCELLENT 2 GOOD 3. GOODsample size 4 FAIR 5 GOOD 6 GOOD 7 EXCELLENT 8FAIR some information on some but not all of the comparative measures 9 EXCELLENT 10 EXCELLENT COSMIN UPDATE1 very good 2 inadequate 3 verygood	fair - excellent	Pearson and Spearman correlations (all GAS refer to post programme GAS T-score) ; pre programme ILS (Independent Living Scale) withpost programme GAS T-score 0.06, 0.06, post programme ILS with GAS 0.40, 0.39 preprogramme VOS (vocational outcome scale)with post programme GAS T-score -0.23, -0.25, post programme VOS 0.34, 0.33, pre programme MPAI (mayo-Portland adaptability inventory staff version) with post programme GAS T-score -0.52, -0.54, post programme MPAI with GAS -0.69, -0.69. Summary; GAS generally correlate with other outcome measuresin a rehabilitation setting at a moderate level, about as well as those outcome measures correlate with each other.	? COSMIN UPDATE 3 hypotheses MPAI #1 +keep separate as its verygood and others are inadequate and (1/250% -) ILS - VOS + (2/3) -

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
GAS	Malec, Smigielski& DePompol o 1991	nil clear hypotheses; "we examined the clinical usefulness of GAS by comparing final GAS scores to other outcome measures. Also examined the predictive ability of early (2mo) GAS scores and initial scored on Portland Adaptability Inventory (PAI) in relation to outcome measures. COSMINUPDATE GAS prediction #1, with PAI #2, work outcome # 2 LOS #3	1. yes n=7 dropouts initially not included, then n=2 removed to avoid redundancy due to initial and f/u identical scores EXCELLENT 2. EXCELLENT 3. POOR n=14 4. FAIR - expectations found in their clinicalpractice 5. GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 GOOD discuss significance of correlation outcome rather than strength or direction. COSMIN UPDATE 1 very good 2 very good 3 very good	fair - excellent	Pearson correlations GAS T-score 2 mo with; GAS final 0.66 (p> 0.01) with LOS -0.08, work outcome 0.31, PAI re admission -0.15, PAI final - 0.35. Final GAS with LOS -0.13, work outcome 0.52 (p> 0.05) PAI preadmission -0.39 PAI final 0.62 (p>0.01). In summary only mod strong correlations 2mo GAS final GAS, final GAS with work outcome and final PAI. in multiple regression analysis a significant degree of variance on work outcome was accounted for byboth LOS (p<0.02) and final GAS T-score (p<0.05), a significant amount of variance on final GAS T-score was accounted for by both LOS(p<0.02) and final PAI (p<0.01). SUMMARY: 2 mo GAS score predicted final GAS 0.66, Final GAS and PAI r = 0.62, p>.01 work outcome r = 0.52, p>.05, no correlation with LOS -0.13	? COSMIN UPDATE GAS prediction +, PAI +,LOS + work outcome +, (+ 4/4)
GAS	Turner-Stokes, Baguley, De Graaff,Katrak, Davies, McCrory & Hughes 2010	nil but research questions -does GAS provide added value as a responsive indicator over other measures, (2) how does it relate to other measures COSMIN UPDATE 7 hypotheses GAS with MAS#2, carer burden #1, patient disability #1, pain #3, mood #3, AQoL #1, global benefit #1	1. EXCELLENT n= 6, 2. GOOD not described but can de deucedwere omitted 3. GOOD sample size, 4 FAIR 5 GOOD 6 GOOD 7FAIR 8 FAIR 9 EXCELLENT 10 EXCELLENT COSMIN UPDATE 1. doubtful, doubtful, very good	fair - excellent doubtful	spearman rho correlations GAS outcome T-score with MAS 0.35, global benefit patient-report 0.46, global benefit - investigator report0.41, HADS anxiety 0.05, HADS depression 0.06, pain at rest 0.03, pain on movement -0.03,AQoL 0.07, patient disability score 0.19, carer burden score 0.14. there was a significant relationship between change from baseline in GAS score and reduction in spasticity in the BoNT-A treatment arm (rho 0.28 p 0.04) but notin placebo (rho 0.04, p= 0.80).	? COSMIN UPDATE MAS +, global benefit -, mood +, pain +, AQoL -, pt disability -, carer burden -,(3/7) overall -

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
MAL	Dromerick , Lang, Birkenmeier, Hahn, Sahrmann , Edwards 2006	evaluate the relationship among impairment, functional limitation and perceived disability assessments COSMIN UPDATE #1 X 3(FIM, MAL, WMFT)	1. missing items not stated (good), 2. nil reference GOOD 3. sample size 39 FAIR 4. hypothesis GOOD 5. direction notstated GOOD 6. magnitude of correlation s not stated GOOD 7. adequate description EXCELLENT 8. EXCELLENT 9. limitations in sample: relevant only to post stroke without moderate or severe cognitive or sensory impairments, may not apply to broader population. trained blinded study personnel FAIR 10. EXCELLENT COSMIN UPDATE 1. VERY GOOD 2 VERY GOOD 3 VERY GOOD	excellentfair	correlation coefficients at 90 days; total ARAT correlated with MAL QOM (r=0.61). WMFT functional ability and MAL-QOM r = 0.65, MAL- AOU r = 0.40 WMFT time and MAL QOM r = - 0.43, and MAL AOU r = 0.007. Pts scoring highly on ARAT and WMFT (functional limitationmeasures) at 90 days post still had some measurable disability - diminished QOM on the MAL tasks, indicating ARAT did not capture full spectrum of motor dysfunction and/or improved QOM was not sufficient for substantial everydayUL use.	? COSMIN UPDATE WMFT (-)ARAT (+) 1/2 50% (-)
MAL	Harris & Eng 2007	1. there would be a significantrelationship among variables of upper limb impairment activity and participation 2. given that participation can beinfluenced by a number f factors upper limb impairment variables would explain a large portion of activity and participation in people with chronic stroke COSMIN UPDATE 7 HYPOTHESES - 5 AT #2 as impairments and 2 at #1 as activity and participation measures	1. GOOD not stated but assumed0 2.GOOD no 3 EXCELLENT 4 GOOD 5 EXCELLENT 6 GOOD 7 EXCELLENT 8 EXCELLENT EXCELLENT EXCELLENT COSMIN UPDATE 1. very good 2very good 3 very good	good - excellent	spearman rank correlations between impairment,activity and participation variables; MAL & MAS - 0.71 ($p<0.01$) MAL & UL strength 0.84 ($p<0.01$)MAL & grip strength 0.61 ($p<0.01$) MAL & sensation -0.43 ($p<0.01$) MAL & Brief pain inventory -0.06 MAL & Chedoke Arm and HandActivity Inventory 0.82 ($p<0.01$) MAL & Reintegration to normal living index 0.23 ($p<0.05$). MAL correlates moderate to high withimpairments excluding pain and very low with participation measure. in the regression model upper limb strength was the only retained variable accounting for 78% of the variance of the MAL scores.	plus COSMIN UPDATE (-)impairment hypotheses #2 generic MAS (-) strength (-), sensation(+), grip (-), pain(+), activity (+), participation

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
MAL 28	Uswatte, Taub, Morris, Light & Thompson 2006	no hypotheses - "we present item, content, reliability, and validity analyses of this instrument "COSMIN UPDATE generic hypotheses1 with like constructs so participant and carer MAL-28AOU and QOM with SIS handand with accelerometry 8 hypotheses	1. GOOD 2 GOOD 3 EXCELLENT 4 POOR 5 GOOD 6GOOD 7 EXCELLENT 8 EXCELLENT 9. 10 EXCELLENT	poor - excellent inadequate due to challenges following paper what related to MAL-30 and MAL28	Convergent validity - with accelerometry and SIS hand function scale with Participant QOM0.52 p<0.01, 0.72, p<0.01 AOU 0.47, p<0.01, 0.68, p<0.01 Caregiver QOM 0.61, p<0.01, 0.40, p<0.01 AOU 0.57, p<0.01, 0.35 p<0.01. Divergent validity with less impaired arm accelerometer and SIS mobility Scale ParticipantQOM 0.14, 0.14 AOU 0.14, 0.14, Caregiver QOM 0.23 p<0.001, 0.07, AOU 0.25 p<0.001, 0.10	? COSMIN updatepatient3/4 (+) caregiver 2/4 (-)
mFrenchay Arm Test	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987	nil stated specifically but aim was to test whether they actually used their affected arms	1. GOOD not described but deduced 2. GOOD 3 POOR n=14 4 POOR unclear what was expected 5 GOOD 6 GOOD 7 n/a 8 n/a 9 FAIR minimal data as many individual studies presented eg raters, inclusion exclusion criteria, generalisability 10.		1. Clinician scored pts as 5/5 (14 pts chronic stroke) – pts then completed 5 “normal” bilateral tasks (not standardised assessment items) to check if were actually using affected hand – 12 pts used both hands for all 5 tasks, 1 used only his dominant hand unaffected hand, only 4 of 14 felt their arm had fully recovered, 5 felt there were few things they could not do, 5 felt they still had a major handicap. Authors concluded valid test of arm function with pts scoring 5/5 using their arm even if they feel it is not normal 2. When a patient scored 5/5 on FAT (n=33) only 17(52%) were within normal limits on 9HPT or 16 (48%) outside normal limits (completed within 18 seconds = normal) - 9HPT more sensitive. Also 6 (18%) on finger tapping and 5(15%) grip strength were outside normal limits - more sensitive and could detect further change/impaired function in the presence of a "normal" FAT score. For patients who scored 0 on FAT, 7 of those 17 were able to record on grip strength - grip strength more responsive in early stages of recovery.	

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
mFrenchay Arm Test	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987 sub study 1	nil stated COSMIN UPDATE	1, EXCELLENT 2. EXCELLENT 3. Poor sample size n=14 4. POOR unclear what was expected 5 GOOD 6 GOOD 7 n/a 8 n/a 9 GOOD 10 GOOD	poor - excellent	12 pts used both hands for all 5 tasks (included 5 tasks normally requiring bilateral arm use daily not a standardised assessment), 1 used only his dominant hand unaffected hand, only 4 of 14 felt their arm had fully recovered, 5 felt there were few things they could not do, 5 felt they still had a major handicap. authors concluded valid test of arm function with pts scoring 5/5 using their arm even if they feel it is not normal.	?
mFrenchay Arm Test	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987 sub study 2	nil stated COSMIN UPDATEpts scoring 5/5 on FAT will also score highly on other measures of hand function (NHPT)	1,GOOD appeared nil missing but random statement " using interpolated scores when assessments were missing's, thevalues used being calculated by assuming a steady change between the two nearest known values" 2. EXCELLENT 3. GOOD sample size 4. POOR unclear what was expected 5 GOOD 6 GOOD 7 EXCELLENT 8 POOR 9 GOOD 10 GOOD	poor - excellent doubtful	when a patient scored 5/5 on FAT (n=33) only 17(52%) were within normal limits on 9HPT or 16(48%) outside normal limits (completed within 18seconds = normal) - 9HPT more sensitive. Also 6 (18%) on finger tapping and 5(15%) grip strength were outside normal limits - more sensitive and could detect further change/impaired function in the presence of a "normal" FAT score. For patients who scored 0 on FAT, 7 of those 17 were able to record on grip strength - grip strength more responsive in early stages of recovery.	? COSMIN update -(0/1)
MI	Bohannon 1999	no hypothesis- primary purpose to examine the criterion validity of UE MI scores and secondary to confirm the construct validity of the scores. COSMIN UPDATE generic hypotheses#2 correlations with related constructs 0.3 - 0.5	1. GOOD 2 n/a 3 POOR n=10 5poor no hypothesis and unclearwhat was expected 5 good 6 good 7 excellent 8 poor no information re measurement properties 9 fair - very minimal information regarding methodology to identify and potential bias or flaws 10 excellent COSMIN UPDATE 1.very good 2 inadequate 3doubtful	poor - excellent	hand grasp dynamometry (HGD), elbow flexiondynamometry (EFD), shoulder abduction dynamometry (SAD) pinch grasp motricity (HGM) elbow flexion motricity (EFM) shoulder abduction motricity (SAM) Pearson correlationsvaried from 0.74 to 0.94.	? COSMIN UPDATE - (0/1) as correlatedvery strongly with measure of grip strength (impairment based not function)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
MI	Collin &Wade 1990	nil stated COSMN UPDATEgeneric hypotheses #1 correlation with similar constructs RMA UL and MI UL	1. GOOD 2 n/a 3 poor 4 poor 5 Good 6 GOOD 7 FAIR 8 FAIR 9 EXCELLENT 10 EXCELLENT COSMIN UPDATE 1. very good 2 adequate 3 very good	poor - excellent	Rank order comparisons (Rho) of RMA arm v MIarm 6 weeks post (n=27) 0.76 (p<0.001) 12 weeks (n=25) 0.73 (p<0.001) 18 weeks (n =14) 0.74 (p<0.01). Strong correlations	? COSMIN UPDATE hypothesis #1 +
MI	Jacob-Lloyd, Dunn, Brain & Lamb 2005	not stated - but aim to establish effective methods ofmeasuring the functional performance of n individual who had experienced a stroke and had been discharged from hospital COSMIN UDATE generic hypotheses #1 UL MI and NHPT at DC and follow up (6 mo later) (2 hypotheses	1. EXCELLENT 2. GOOD 3. GOOD 4. FAIR 5 GOOD 6 GOOD 7 GOOD 8 EXCELLENT 9EXCELLENT 10 spearman rank correlations used but commentedon significance not strength GOOD COSMIN UPDATE 1. verygood "upper limb function" 2. verygood 3 very good 4	poor - excellent	correlation between UL MI and 9HPT was significant at discharge (r_s = 0.53 n = 22 p = 0.01) but not at follow up (r_s = 0.36 n = 22 p = 0.10)	? COSMIN UPDATE hypothesis 1 on dc +, follow up hypothesis 2 - (1/2) 50% means -
MI	Stone, Patel & Greenwood 1993	nil specific stated - study carried out to determine the prognosis of patients presenting with visual neglectat two and three days after stroke - COSMIN UPDATE use hypotheses MI would be a predictor of independence post stroke when neglect is present	1. EXCELLENT 2. GOOD 3 FAIR4 POOR 5 GOOD 6 GOOD 7 EXCELLENT 8EXCELLENT 9 FAIR -don't really know what happened in the interim in regards to therapy etc 10 EXCELLENT COSMIN UPDATE 1. very good 2 very good 3doubtful	poor - excellent	linear logistic regression showed that combiningthe MI with Visual neglect recovery index and patients age scores at 2 - 3 days post stroke were significant predictors of independence (mod Barthel Index) at 3 and 6 months. Regression equations correctly predicted 78% of outcomes with a sensitivity and specificity for "independence" of 84% and 90% respectively and "moderate/severe" 89% and 80%. CHECK	? COSMIN UPDATE plus as MI was a predictor of independence

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
MI	Wade & Hewer 1987	nil but states - this paper reports an analysis which establishes the frequency andseverity of paralysis after stroke, the validity of assessments used and extentof recovery made. COSMIN UPDATE 2 hypotheses at generic #1 MI and mBI	1. EXCELLENT 2 GOOD 3 EXCELLENT 4 POOR 5 GOOD 6GOOD 7 POOR 8 POOR 9 explicit data re assessors, analysis missing FAIR 10 coefficients is all that s stated GOOD COSMIN UPDATE 1. very good 2 very good 3 doubtfulmissing info	poor-excellent	total MI correlation coefficient with Barthel Indexwere 0.749 initially, 0.774 at 3 weeks and 0.610at 6 months. States that the coefficients for armand leg scores similar but not given. Patients with severe paralysis (0-31 on MI) were often unconscious, those with sever arm paralysis, 40% had died by 3 weeks, 55% by 6 months. Ofthe survivors 73% still had severe paralysis at 3weeks 55% at 6 months 2 % were normal at 3 weeks and 10% by 6 months - note nil information re therapy/medical involvement.	? COSMIN UPDATE + (1/1)
NHPT	Morris, van WijckJoice & Donaghy 2013	hypothesised that UL activityand limitation constructs andanxiety would emerge as significant predictors of HRQOL measured on the Nottingham Health Profile	1. missing data given 1.2% of data across 6 variables EXCELLENT 2. EXCELLENT 3. GOOD sample 85 4. minimal hypothesis GOOD 5. GOOD 6 EXCELLENT 7 EXCELLENT 8 EXCELLENT 9 a considerable proportion of the variance in totalHRQOL score not explained andmay have influenced findings, limited socio-demographics collected to predict HRQOL, sample not representative of general stroke population FAIR 10 COSMIN UPDATE 1. very good 2 very good 3 doubtful DOUBTFUL	fair - excellent	NHPT and NHP r = -0.08 no correlation with quality of life not a predictor of overall HRQOL	minus asdid not predict HRQOL. (0/1)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
OHS	Rigby, Gubitz, Eskes, Reidy, Christian, Grover & Phillips 2009	nil specific - to identify patientfactors that contribute to higher levels of caregiver burden COSMIN UPDATE hypotheses #1 OHS predictorof caregiver burden	1. GOOD 2. N/A 3. EXCELLENT 4 POOR 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 two different scales used to measure global handicap at discharge and 12 months Limitscomparisons FAIR 10. GOOD COSMIN UPDATE doubtful as per above	poor-excellent doubtful	OHS score at discharge (score of 3-5) was not apredictor of caregiver burden as measured by Relatives Stress Scale (RSS) P<0.0001 and Bakas Caregiver Outcomes Scale (BCOS) P 0.059	? - (0/1)
OHS	Simon, Kumar &Kendrick 2008	nil specific - explore support provided to stroke survivors and their live-in informal carers, the factors associatedwith it and carers views about support provided COSMIN UPDATE hypotheses - OGHS predictor of services requires, 2) time required (assistance time)	1. not stated GOOD 2, GOOD 3.EXCELENT gave evidence of power calculations for sample size needed 4 Fair - gave furtherbackground information end evidence to deduce what was expected 5 GOOD 6 GOOD 7 N/A 8 N/A 9 EXCELLENT 10GOOD COSMIN UPDATE adequate	fair - excellent adequate	6 weeks post discharge OHS (level of handicap)was the strongest predictor of number of services 13% of variance F = 9.53 d.f.1=1, d.f.2=64, P=0.001, and amount of time provided26.5% variance F=26.39, d.f.1=1, d.f.2=64, P<0.001. The more handicapped the survivor the more support was provided (B=0.43, P<0.001) and (B=0.37, P=0.001) for amount of time provided. B= estimated regression coefficients. At final interview mean 15.5 monthspost, forward stepwise regression OHS (revealed level of handicap) to be the only determinant of number of services accounting for 14% variance, and was most important predictor accounting for 28.2% variance for timeallocated. In both cases the higher levels of handicap resulted in ore care provision (B=0.38,P=0.006 and B=0.52, P<0.001 respectively).	? + (2/2)
RMA	Collin &Wade 1990	nil stated	1. GOOD 2 n/a 3 poor 4 poor 5 Good 6 GOOD 7 FAIR 8 FAIR 9EXCELLENT 10 EXCELLENT	fair - excellent	Rank order comparisons (Rho) of RMA arm v MIarm 6 weeks post (n=27) 0.76 (p<0.001) 12 weeks (n=25) 0.73 (p<0.001) 18 weeks (n =14) 0.74 (p<0.01). Strong correlations - concurrent validity. RMA leg v MI leg 6 wk 0.81 p< 0.001, 12wk 0.81 p< 0.001, wk 18 0.75 p< 0.01, RMA GF v TCT 6 wk 0.70 p< 0.001, 12 wk 0.72 p< 0.001,18 wk 0.79 p<0.001.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
RMA	Jones1998	H1 & H2 both physiotherapists and patientswould predict functional change accurately within a 6 week period. H3 physiotherapists would showa higher level of accuracy than patients COSMIN UPDATE SAME HYPOTHESES HOWEVER for H1, H2 the correlationwould need to be> 0.50	1. GOOD not stated but can deduce 0 2 n/a 3 Poor sample size 4 excellent 5 excellent 6 good 7 n/a 8 n/a 9 excellent 10 excellent used ICC and Bland Altman for level of agreement and accuracy of predictions. COSMI UPDATE 9b 5. very good 6 very good	poor - excellent	physio UL section predicted mean 3.86, median 2.00 SD 4.45 achieved mean 4.03 median 1.00SD 4.65. LL and truck and gross also available.Patient predicted mean 4.38 median 4.00 SD 4.53 achieved mean 4.03 median 1.00 SD 4.65. ICC between physios predicted and achieved follow up scores = 0.965 and patients predicted and achieved follow up scores 0.908. Bland Altman measurement of agreement between physio predicted and follow up scores 96.6 and patents 79.3. all hypotheses were accepted - physio and patient predictions demonstrated high and significant agreement with the achievedRMA scores at 12 weeks, physios predictions demonstrated marginally higher levels of agreement than patients across the 3 sections.	plus 3/3100%
RMA	Sackley1990	nil hypotheses, not even stated as aim but as "in this study the pts ability to stand symmetrically was measuredand at the same time assessments were made of motor function and the abilityto perform activities of daily living. The relationship between these measures were then evaluated" COSMIN UPDATE generic hypotheses for complete RMA 1) with ADL scale #1, 2) Balance coefficient #2 UL only 1) with ADL #1,	1. GOOD not stated but can deduce 0 2. N/A 3. good n=90 forrelationship 4. Poor unclear whatwas expected 5 good 6 good 7 excellent 8 excellent 9 fair - minimal detail re potential bias with recruitment, who was completing assessments 10 excellent COSMIN UPDATE 1 Very good 2 adequate 3adequate	poor - excellent adequate	spearman rho correlations - arm function & BC2r= -0.45 P<0.001 Arm function & ADL r=0.51 P<0.001 total motor function and ADL r=0.68 P<0.001 Total motor function and BC2 r= -0.45 P<0.001. (BC2 = balance coefficient which gavethe difference from the midpoint of weight distribution).	? COSMIN UPDATE total RMAwith balance r = -0.45 (+) ADL r = 0.68 (+) total (2/2) + UL RMA with balance r = -0.45 (+) withADL r = 0.51 (+) total 2/2

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
		2) with balance coefficient #2				
RMA - UL	Morris, van WijckJoice & Donaghy 2013	hypothesised that UL activity and limitation constructs and anxiety would emerge as significant predictors of HRQOL measured on the Nottingham Health Profile # 1ARAT and UL RMA	1. missing data given 1.2% of data across 6 variables EXCELLENT 2. EXCELLENT 3. GOOD sample 85 4. minimal hypothesis GOOD 5. GOOD 6 EXCELLENT 7 EXCELLENT 8 EXCELLENT 9 a considerable proportion of the variance in totalHRQOL score not explained andmay have influenced findings, limited socio-demographics collected to predict HRQOL, sample not representative of general stroke population FAIR 10 COSMIN UPDATE 1. very good 2 very good 3 doubtful DOUBTFUL	excellentfair	UL section of the RMA was used as an impairment orientated measure of UL motor control. UL RMA and activity limitation (ARAT) were negatively associate with the Physical Activity and total NHP score (Nottingham HealthProfile) indicating that greater UL dysfunction was associated with poorer HRQOL. correlationsUL-RMA with 6 months NHP total score -0.30, energy levels -0.21, pain -0.14, emotional reactions 0.02, sleep -0.20, social isolation -0.19,physical activities -0.47. But the UL -RMA demonstrated higher collinearity with NHP and physical activities than the ARAT. A multivariate analysis found that UL RMA did not predict overall HRQOL.	minus for predicting HRQoL (1/2) correlation with likemeasure ARAT and UL RMA +
SA-SIP	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	greater disability is associatedwith worse HRQoL and a greater need for caregiver assistance. COSMIN UPDATE #1 SA-SIP and DAS	1. EXCELLENT 2. N.A 3. EXCELLENT n = 279 4. GOOD 5. EXCELLENT 6 GOOD 7. EXCELLENT 8 fair reference tostudy 9 EXCELLENT 10 GOOD marked down due ot ordinal nature of DAS not accommodated in one way analysis of variance. COSMIN UPDATE adequate	fair	in pts with UL stroke spasticity greater disabilityscores in hygiene, dressing, limb posture and pain domains were significantly (P<.05) associated with higher overall SA-SIP dysfunction scores. Greater DAS were associated (P<.05) with higher SA-SIP physicaldimension scores in all domains. Increasing disability is associated with reduction in HRQoLand caregiver burden (P<.05)	plus COSMIN UPDATE 1/1 +

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SA-SIP	Edwards, Hahn, Baum & Dromerick 2006	"…we investigated measures of life satisfaction and participation in meaningful activities in patients 6 monthsafter mild stroke "COSMIN UPDATE 1. generic hypotheses #1 predictor of life satisfaction #1 discrimant validity with FIM	1. EXCELLENT 2 N/A 3 EXCELLENT 4 FAIR 5. GOOD 6 GOOD 7 EXCELLENT 8 POOR 9 EXCELLENT 10 EXCELLENT 1. very good 2 adequate 3 very good	fair - excellent	1. low NIHSS scores at hospital admission correlated with 6mo SA-SIP r=0.11 (r=Pearson correlation) (low). 2. Individual NIHSS items didnot significantly predict stroke disability as measured by SA-SIP (P≤ 0.08) in patients with mild sensorimotor impairments. Multiple linear regression analysis predicting life satisfaction using Reintegration to Normal Living (RNL) Scale as the dependent variable and age, sex, race, stroke severity as measured by admissionNIHSS score, total FIM, stroke-related disabilitymeasured by SA-SIP, SF-12, activity card sort and any post acute rehab treatment as independent variables, the equation yielded R2 =0.63)P<0.0001) with SA-SIP accounting for 53% of variance, ACS 9% SF-12 4%. Age, race,sex, admission NIHSS, FIM and rehab treatment were not significant predictors of RNL.	? COSMIN UPDATE 2/2 + predictiveof life satisfaction and more sensitive than the FIM 6 months

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SF-36	Anderson 1996	Construct validity was assessed by examining the extent to which the SF-36 scores varied according to predefined hypotheses. We hypothesized that the scores should vary in a predictable manner among patients with physical disability, mental ill health and varying levels of social activities according to the measures outlined below.Physical disability using the BI, Mental ill health GHQ-28, social activities using the APPCOSMIN UPDATE no clear hypotheses despite stating in article, generic hypotheses 1) SF-36 with Adelaide ActivitiesProfile #1, 2) SF-36 can discriminate between those with varying level of independence as per mBI	COSMIN UPDATE 1 very good 2 adequate 5 very good 6 adequate - reporting vague and difficult to follow methodology and statistical analyses.	adequate	Construct validity was demonstrated by clear difference across all eight SF-36 scales for patient with identified health problems (Firstly those with physical function dependence and secondly for those with mental ill-health). Meanscompared to make comparison between groups-mann whitney u test (found to be significant). COSMIN UPDATE SF-36 discriminated betweenthose independent or dependent as determined via mBI scores.	#1 - <0.50 correlation #2 discrimant + (1/2) overall -

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SF-36	Dorman etal 1999	In investigating the EQ5D andSF36, it was hypothesized that 'there should be a strong correlation between responses on the 2 instruments. A poor correlation might suggest poor validity of 1 or both of the measures.' COSMIN UPDATE generic hypotheses #1	1. very good 2 adequate 3 adequate	adequate	P2150 we found a close relationship between the domains that assessed physical functioning, social functioning, bodily pain and overall health-related quality of life. Correlation between patient' responses to the mental health domain of the sf-36 and the psychological functioning domain of the EQ5D was poor. Strong correlation between patients' responses to thesedomains and supports the view that both these domains are measuring the same underlying trait. Table 4	plus excluding mental health other domains correlatedstrongly with relevant domain in other measure
SF-36	Duncan1997	Not explicit. Wanted to compare three patient groups(mild stroke, TIA, those at riskof stroke) and compare a 'broad array 'of measures. Reported reduced sensitivityof BI COSMIN UPDATE generic hypotheses #5 discriminant validity with SF-36 and BI	COSMIN UPDTAE 5. very good 6. doubtful as statistical significance with no hypotheses,also collection of BI via medical records	doubtful	For participants scoring 100 on the Barthel, the stroke group was significantly more impaired than the asymptomatic (risk of stroke) group in every dimension of the MOS-36 except pain. (statistical significance ranged from 0.34 to .001.	plus (1/1)(did not look ta correlation s as outside initial focus of study andmetal health focus)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SF-36	Findler2001	If the SF-36 is a valid measure of health problems in persons with TBI, scores on its subscales should be correlated with scores on theSCL and HPL. Correlations also predicted between the SF-36 scales and the BDI-II since both health problems and depression are commonamong persons with TBI. It was expected that the strongest correlations would be found between subscales measuring similar constructs. It was expected that participants in the TBI groupswould have significantly lowerSF-36 scores (i.e. more health complaints) than the comparison group and that members of the moderate/severe TBI group would have lower scores thanthe mild TBI group. COSMIN UPDATE generic hypotheses#1 fcorrelations with TIRR symptom checklist SF-36 withHealth Problems List 3@, SF-36 with Depression #2	COSMIN UPDATE 1. very good 2 adequate 3 adequate 4 very good + (2/2)		There was no information on the measurement properties of the comparator instruments. Med- strong correlations were reported between SF- 36 items and scales measuring similar constructs in TBI groups. For between group comparisons, 'The TBI groups obtained significantly lower SF-36 scores than the comparison group, and the mild TBI group scored lower than the moderate± severe group. For the most part, the differences between the TBI groups disappeared when BDI-II scores were controlled for. These findings suggest that the SF-36 is a reliable and valid measure for usewith persons with TBI'. *abstract*	? (for convergent) + (for between groups)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SF-36	Freeman, Hobart, Langdon et al 2000	hypotheses #1 (similar) GHQ,FIM, EDSS, LHS hypotheses #3 (different constructs) FIM, GHQ, hypotheses #5 meaningful changes betweensubgroups - hypothesis that patients categorised into severe group would report lower scores on both of the summary scales than those inmild group, additional study hypotheses 1. pts requiring carer assistance will report lower scores in the physical dimension than hose independent in daily life 2. ptswith relapsing remitting MS will report higher scores in thePCS than those with secondary progressive 3. pts scoring greater than or equal to 5.0 points on the GHQ (indicating emotional distress)will report lower scores on theSF-36 MSC	COSMIN UPDATE 1. very good 2. very good 3 very good 4 adequate as commenting on statistical significance for some hypotheses results - not optimal	adequate	**Numbers do not related to hypotheses (except 4):** 1) Intercorrelations between the SF- 36 dimensions: related dimensions were more strongly associated than less related dimensions. 2) Associations b/n SF-36 dimensions and instruments measuring related constructs were strongest between those measuring similar concepts 3) statistically significant differences between the patient subgroups (mild, mod, severe) occurred in threedimensions (social function, physical function and physical role limitations) and the physical summary score. 4) a. patients requiring carer assistance reported lower scores in the physicalrole limitations dimension than those who are independent ($p<0.0001$, mean scores=13.5 and 43.7 respectively); 4) b. patients with relapsing- remitting MS reported higher scores in the physical summary scale than those with secondary progressive MS ($p<0.0001$, mean scores=35.7 and 27.4 respectively); 4)c. patientsscoring >5.0 points on the GHQ reported lower scores on the mental summary scale than those scoring <5.0 points ($p<0.0001$, mean scores=40.4 and 52.2 respectively).	plus (7/7)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SF-36	Freeman, Langdon, Hobart & Thompson 1996	nil so will use hypotheses #1 -FIM, GHG, EDSS,	COSMIN UPDATE 1. very good 2very good 3 very good		Overall, Fair. 1) Compared to the general population, the patients with mod-severe physical disability report poorer levels of health in every dimension of SF-36. Compared with four condition specific groups - low back pain, menorrhagia, suspected peptic ulcer and varicose veins - poorer levels of health were reported by people with MS in all dimensions other than that of pain. 2) Spearman's rank correlation coefficient demonstrated a moderate negative correlation between the total score on the GHQ and both the emotional well being dimension (r=-5.173; p<0.001) and the role limitations due to emotional health problems dimension (r=-0.4323;p<0.002) of the SF-36. Rethe FIM, only a 'moderate' positive correlation between the FIM motor domain and the physicalfunctioning dimension of SF-36 (r-0.5565; p<0.001) and no correlation between the dimension of role limitations due to physical problems (r=-0.0319; p>0.826).	2/3 (-)
SF-36	Hagen, Bugge & Alexander 2003	study hypotheses 1. there would be positive bivariate correlations between all SF- 36 sub scores and the BI, MMSE, and CNS at each timepoint #1 for BI and CNS and #2 with MMSE	COSMIN UPDATE 1. Inadequate 2. adequate 3 very good	inadequate	The strongest correlations were found between SF-36 scores and the Barthel Index and CNS ashypothesized, however, they were lower than expected. Highest for Physical functioning, and social functioning. SF-36 responses from subjects that were not fully conscious at stroke onset or incontinent in first 7 days with differences observed between the groups (mannwitney u test = not significant)	1/3 (-) did not correlate >0.5 with expected dimensions and BI or CNS, as expected low correlations with MMSE

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SF-36	Hermann, Vickrey, Hays, Cramer, Devinsky, Meador, Perrine, Myers & Ellison 1996	First to compare the SR HRQoL of patients with epilepsy to another chronic neurological condition (MS and to a chronic disorder not primarily affecting the CNS (diabetes) using a generic HRQoL instrument the SF-36. COSMIN UPDATE generic hypotheses #5 MS with epilepsy, MS with diabetes	5 very good 6 very good	very good	Patients with MS scored significantly lower (worse HRQoL) than both the epilepsy and diabetes groups on the following scales: PF, RLP, E and SF. Patients with epilepsy and MSdid not differ from one another (all p values > 0.1) and both scored significantly lower (worse)compared to the diabetes group on the emotional wellbeing and role limitations- emotional subscale	1/2 50% -
SF-36	MacKenzie 2002 'Using theSF-36'	It was hypothesised that whencompared with either the dimension-specific or summary scores of the SF- 36, the cognitive function scale would discriminate better among patients with and without head injury of varying severity. In accordance with the hypothesis we expected that the F-ratios for the SF-36 scales to be small and not significantly different from 1, whereas the F-ratio for the cognitive scale would be larger and statistically significant. COSMIN UPDATE - Use study hypotheses	COSMIN UPDATE 1. very good 2. adequate (used in previous studies) 3 very good 5 very good	adequate	Using the SF-36 alone, there was little variationin scores of PCS and MCS observed between different levels of severity of HI. (F= 0.1-2). In contrast, t cognitive function component demonstrated greater variation with F-statisticsof 5.6 - 8.4 depending on with or without orthopaedic injury. Indicates use of COG supplement to the SF_36 when evaluating outcome from multiple trauma involving head injury.	(0/1) - didnot differentiate requires cognitive supplement in TBI

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
		which relate to generic hypotheses #5				
SF-36	Madden, Hopman, Bagg, Verner, O'Callaghan 2006	hypothesised that variables describing similar facets of function or quality of life (e.g. the SF-36 PF domain with the FIM mobility and locomotion domain) would have correlations of ≥ 0.25 COSMIN UPDATE generic hypotheses #1	1. very good 2 very good 3 very good		nil correlations between SF-36 and FIM ≥ 0.5	(0/1) -
SF-36	Moore, Wolfson, Alexandrov & Lapierre 2004	purpose of study to determine which measure would be more useful in clinical practice COSMIN UPDATE generic hypotheses #1 SF-36 PC with Pt QOL rating and with pt severity rating and with EDSS, SF-36 MC with the same 3 all hypotheses #1 6 hypotheses	1. very good 2 inadequate 3 very good 4 very good	inadequate	SF-36 physical and pt QOL rating 0.47, severity rating 0.38 and EDSS -0.69. SF-36 mental with QOL rating 0.29, severity rating 0.18 EDSS - 0.06. SF-36 physical Odds Ratio 0.86 95%CI 0.81 – 0.91, SF 36 mental 1.02 95%CI 0.98 – 1.06 crosses null value 1.	1/6 (-)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SF-36	Moore, Wolfson, Alexandrov & Lapierre 2004	purpose of study to determine which measure would be more useful in clinical practise	1. 3% 2. n/a 3 excellent 4 fair 5 n/a 6 n/a 7 excellent 8 excellent 9. fair - excluded even mild cognitive impairment, 10 excellent	fair	Participants felt completing all 3 measures was most reflective of their QOL (EQ-5D, SF-36, MSQOL-54 all correlated similarly). SF-36 predicting EDSS OR; SF-36 Physical 0.86 (0.81 - 0.91), mental 1.02 (0.98 - 1.06). SF-36 Physical, Mental Correlations with "Pts QOL rating" 0.47 (95%CI 0.30, 0.72), 0.29 (0.09, 0.50) with pts severity rating 0.36 (0.18, 0.58), 0.38 (0.18, 0.62), 0.18 (-0.03, 0.40), EDSS -0.69 (-1.00, - 0.64), -0.06 (-0.26, 0.14).	?
SF-36	Pittock 2004	No hypotheses stated… You might assume they expected QoL to be worse in people with MS and that the SF-36 and the EDSS would be moderately correlated. COSMIN UPDATE hypotheses #2 similar and #4 discriminant between groups - MS and healthy	COSMIN UPDATE 1. very good 2 very good 3 very good	very good	Physical functioning, role physical, general health, vitality and the physical component score were considered clinically worse in MS population compared with American general population. Other scores not clinically significant. PF, RP, GH, SF and vitality were significantly correlated with the EDSS score.	5 of 8 (62.5% so a -) items clinically worse between MS and US Population 1/2 (-)
SF-36	Robinson, Zhao, Kim& Revicki 2009	"to explore relationships between clinical measures and HRQoL scores, and to investigate baseline, cross sectional differences in several HRQoL questionnaires for subsequent validation as MID - these two were hypothesis generating rather than testing" COSMIN UPDATE	1. excellent 2. excellent 3. excellent 4 poor hypothesis wasreported to be a hypothesis generating study not testing 5 good 6 good 7 fair 8 fair 9 excellent 10 excellent COSMINUPDATE 1. very good 2 adequate 3 very good	poor	Baseline SF-36 correlated with MSFC 0.16 – 0.51 (MH), (PF), EDSS -0.13 (MCS) to -0.68 (PF), age no correlation (MCS) to -0.45 (PF) and disease duration -0.02 (RE) to -0.40 (PF), fatigue severity scale -0.31 to -0.72. . Did not correlate with lesion count, pre-baseline relapserate	? COSMIN UPDATE SF-36 and MSFC <0.5 so (-) (0/1)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
		HYPOTHESES #1 SF-36 withMSFC,				
SF-36	Vickrey1995	No hypothesis stated although you could assumethey expected QoL to be worse in people with MS compared with the general population. COSMIN UPDATE hypothesis #5	COSMIN UPDATE 1. very good 2adequate	adequate	MS patients scored 48 points lower than the general population on both the physical functionand role limitations due to physical problems scales. Social function scores were 25 points lower for MS patients relative to the general USpopulation; energy/fatigue, health perceptions and role limitations due to emotional problems scores were approx. 20 points lower.	plus
SF-36	Vickrey, Hays, Genovese , Myers,Ellison 1997	We hypothesised that patients reporting less severeMS symptoms (four categories), less disability in terms of ambulation (four categories), fewer days of missed work or school due to health in the prior month (3 categories: 0 days, 1-15 days,and 16-30 days) and higher overall quality of life (tertiles) would report better HRQOL than other respondents. 4 hypotheses, plus supplementing SF-36 with disease specific measures would contribute unique information (total 8 hypotheses)	good COSMIN UPDATE COSMIN UPDATE 1. very good 2very good 3 very good	good,very good	As hypothesised, patients reporting fewer days of missed work or school due to health in the prior month (3 categories: 0 days, 1-15 days, and 16-30 days) and higher overall quality of life (tertiles) had higher SF-36 scores than other respondents (Tables 3 &4). Those reporting less severe MS symptoms (four categories) mostly performed according to the hypothesis (except role limitations-emotional and pain). SF-36 was less able to distinguish b/n groups based on ambulation status (four categories) with expected results for physical function, emotional well-being (mental health?), social function and current health (general health). COSMIN UPDATE when SF-36 is used as a generic measure other disease specific measures shouldalso be included. second lot of hypotheses - in all 4 areas, disease specific measures had a higher relative validity than SF-36.	4/8 plus (disease specific measures contributed more unique information than SF-36 (-)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SF-36	Williams1999 measuring	The aims of this research are to use the SS-QOL to identify the predictors of poststroke HRQOLO in patients with mildt o moderate ischemic stroke and to compare the ability of the SS-QOL with that of a generic HRQOL scale for predicting overall HRQOL poststroke. The authors question the validity of the SF-36 (and other generic measures) to truly capture theQoL of people with stroke given they do not assess common stroke impairments such as hand dexterity, communication or vision. So they want to make a comparison between these two types of measures. It could be assumed that they predict that SSQOL is better at predicting QOL post stroke than the SF-36.COSMIN UPDATE HYPOTHESES SF-36 predicted HRQol betterthan SS-QOL	inadequate as per results statement plus no hypotheses and statistical significance wasstats approach (less than ideal)	inadequate	The sample was dichotomised by the use of a simple question about QOL to patients asking whether their QOL was worse or the same as pre stroke. This question is not developed or described in anyway. So it is assumed that thestroke patients can give a valid and reliable answer to this question and no evidence is provided to support this assumption. It is this question which the entire analysis of the SF-36depends on. Hence my rating of poor. results state SF-36 scores not associated with overallHRQoL rating	0/1 -

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SF-36	Guilfoyle, et al 2010	New outcome measures should demonstrate appropriate relationships to established instruments. To determine the external validity to SF-36 scales, the mean domain scores were compared across GOSE categories using multiple one-way analysis of variance (ANOVA, degrees of freedom=5) COSMIN UPDATE generic hypotheses #1	1 very good 2 very good (not for current article but elsewhere) 3 doubtful - minimal methodological information provided to be able to rate for bias etc	doubtful	Multiple ANOVA of the SF-36 domain scale scores over the GOSE categories represented in the sample (i.e., 3–8) showed that mean scores increased with higher GOSE categories, which was significant for all eight domains ($F=19.7–48.8$, $df=5$, all $p<10^{-15}$; Fig. 3).	plus (1/1)
SF-36	Riazi et al 2003	No hypotheses stated… Youmight assume they expected QoL to be worse in people with MS than general population and that people with worse mobility would have lower QoL, COSMIN UPDATE GENERIC HYPOTHESES 5	9b 5 very good very good		Participants with MS had lower mean scores on all dimensions of the SF-36 compared with the UK norms after controlling for sociodemographicvariables ($p<0.001$). Differences in scores were larger for the two physical domains of the SF-36,small for the mental health dimension and substantial for the others (as you might expect LC). Less physically disabled people with MS had significantly higher scores ($p<0.05$) on all SF-36 dimensions than those who used support when walking.	plus (1/1)

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SIS	Duncan, Lai, Tyler, Perera, Reker & Studenski 2002	"evaluate the validity of patient and proxy responses" COSMIN UPDATE use of generic hypotheses all #1, SIS strength - MI, SIS mob - BI, SIS ADL/IADL - BI, SIS ADL/IADL - BI 8 hypotheses (4pt, 4proxy ratings)	1. excellent 2 n/a 3 excellent 4 poor 5 good 6 good 7 poor 8 poor 9 excellent 10 excellent COSMIN UPDATE 1. very good 2 adequate 3 very good	poor - excellent adequate	Pearson correlations between selected SS domains and other assessments Folstein MMSE & SIS memory pt 0.42 proxy 0.37, MODERATE, BI & SIS ADL/IADL pt 0.72 proxy 0.78 STRONG, BI & SIS mobility pt 0.69 proxy 0.70 STRONG Lawton IADL & SIS ADL/IADL pt 0.77 proxy 0.78 STRONG Motricity and SIS strength 0.67 proxy 0.69 STRONG	? COSMIN UPDATE 4/4 for pt Responses 4/4 proxy + (100%)
SIS	Eriksson, Baum, Wolf & Connor 2013	"we sought to determine the extent to which perceptions of participation in everyday occupations were affected in a sample predominantly mild stroke" no hypotheses COSMIN UPDATE generic hypotheses used SIS participation - perceived recovery #1, SIS participation Retained activities (ACS) #1, SIS – participation-community reintegration (RNL) #1 SIS participation - NIHSS #2 (4 hypotheses)	1. excellent 2 good 3 excellent sample size 4 fair 5 good 6 good 7 excellent 8 excellent 9 fair mild impairments - reduced generalisability 10 excellent COSMIN UPDATE 1. very good 2very good 3 very good	fair	predictors of perceived participation were perceived recovery on SIS, percentage of activities retained on the Activity Card Sort andReintegration to Normal Living Index.NIHSS score and age were not significant predictors. SIS participation - NIHSS r = -.25, RNL r = .71,ACS r = .67, perceived recovery r = .53.	? COSMIN UPDATE NIHSS -,RNL +, ACS +, perceived recovery + 4/5

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SIS	Lai, Studenski, Duncan & Perera 2002	aim to compare the disability and QoL as measured by the SIS of stroke pts deemed recovered by a score of greater than or equal to 95 onthe BI to 2 non stroke community dwelling pts	1. GOOD - reports some variables had missing data but unable to determine which via tables 2. FAIR 3. GOOD n=81 4.POOR 5. GOOD 6 GOOD 7 n/a 8. n/a 9. FAIR - minimal detail given re methodology is not transparent 10. EXCELLENT	poor - excellent doubtful	pts scoring greater than or equal to 95 on the BI(deemed recovered or minimal disability) 3 months had mean hand function scores 9 pointslower, social participation 12.8 points lower and ADL/IADL 5 points lower than non-stroke community dwelling participants.	? Update1/1 + (study hypothes es)
SIS	Wolf &Koster 2013	" the persons own perceived recovery, as measured by theSIS, may be a more significant predictor of the amount of retained HDL activities after stroke" High demand leisure (HDL) COSMIN UPDATE generic hypotheses #1	1. excellent 2 good 3 good 4 good 5 good 6 good 7 excellent 8excellent 9 fair - not a very representative sample, mild stroke and volunteers, minimal information in data collections as from registry 10 excellent COSMIN UPDATE 1. very good 2very good 3 doubtful	fair	logistic regression indicated that SIS total perceived recovery and SIS strength domain were the only statistically significant factors determining percent of retained HDL following mild stroke, SIS total OR1.027, 95%CI 1.0-1.056P0.05, SIS strength domain OR 1.033, 95%CI 1.007-1.061 P0.01. BI total score did not show statistical significance.	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SIS	Kwon, Duncan, Studenski, Perera, Lai & Reker 2006	purpose was to assess the utility of the SIS in a community dwelling stroke survivors with more realistic administration methods such as telephone and mail survey.We focus on the construct validity of telephone SIS administration. COSMIN UPDATE generic hypotheses #1 for SIS - FIM motor, FIM - cognitive #3 SIS - SF-36V #1, discriminative ability compared to FIM #1 SF-36V #1 (4 hypotheses)	1. EXCELLENT 2. EXCELLENT 3. Good sample size n=95 4. POOR - not clear what was expected no hypotheses 5. GOOD 6 GOOD 7 GOOD - no information presented on FIM only presented for SF-36V 8 GOOD 9 EXCELLENT 10 EXCELLENT COSMIN UPDATE 1. very good 2 very good 3. adequate - 4 week differencebetween administration of measures 4 very good	poor - excellent	convergent validity (Pearson correlations all p<0.001)- SIS strength with FIM-motor 0.404, SF36V physical functioning 0.477, SF-36RP role physical 0.533, SF-36V general health 0.460, SF-36V PCS physical component summary score 0.520. SIS Memory with FIM-Cognitive component 0.501, SF-36V general health 0.378, SIS Emotion with SF-36V role emotion 0.504, SF36V mental health 0.713, SF-36V general health0.460, SF-36V mental component summary score 0.692. SIS communication with FIM- cognitive 0.637, SF-36V general health 0.362, SIS ADL/IADL with FIM-motor 0.858, SF36V; physical functioning 0.732, role physical 0.711, general health 0.503, physical component summary score 0.586, SIS mobility with FIM motor 0.738, SF-36V; physical function 0.755, role 0.724, general health 0.574, physical component summary 0.632 SIS hand function with FIM motor 0.659, SF-36V physical function 0.682, role 0.631, general health 0.470, physical component summary 0.628, SIS social participation with FIM motor 0.588, FIM cognitive0.549, SF-36V; physical function 0.667, role 0.750, emotion 0.583, social functioning 0.655, mental health 0.601, vitality 0.593, general health 0.531, physical component summary 0.539, mental component summary 0.618, SIS PHYSICAL with FIm motor 0.773, SF-36V physical 0.768, role 0.750, vitality 0.529, generalhealth 0.576, physical component summary 0.687. DISCRIMINANT VALIDITY Kruskal-Wallistest SIS -16 H 19.17 p 0.0003, SIS-PHYSICAL H18.39 p 0.0004, SIS-ADL H 18.79 p 0.0003 compared to FIM motor H 17.83. SIS discriminated 3 pairs of disability levels	? COSMIN UPDATE hypotheses 1 SIS -FIM (FIM motor with SIS strength0.4, ADL 0.8, mobility 0.7, hand function 0.66, social participation .59, physical 0.77) +, FIM cognitive (social participation) +, SF-36 +, discriminates better than FIM +, SF-36 + all via telephone SUMMARY + 4/4

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
UL MAS	Khan, Chien & Brauer 2013	nil specific but purpose was tocompare the discriminatory ability of the Rasch-based and conventional summative scoring using mMAS COSMIN UPDATE hypotheses 1 Rasch based better scoring (more precise)than summative (no correlations defined)	1. good 2 good 3 excellent 4 fair 5 good 6 good 7 n/a 8 n/a 9 fair no detail re potential bias in treating physio and blinding etc 10 excellent COSMIN UPDATE discriminative 5. very good 6 verygood	fair - excellent	Rasch based scoring was more precise in differentiating patient groups (post stroke) by discharge destination than the conventional summative scoring in measuring upper limb function at admission and discharge - 15% precision at admission (RP, 1.15; 95%CI: 1.01,1.40) and 11% at discharge (RP 1.11; 95%CI: 1.02, 1.23) both are statistically significant. - particularly evident at the extreme end of the scale; 20% precision at admission (RP, 1.20; (95% CI: 1.08, 1.42) and a 19% precision at discharge (RP, 1.19; 95% CI: 1.08, 1.37). .	plus (1/1)
UL MAS	Kuys, Bew, Lynch, Morrison& Brauer 2009	research question: which measure of activity limitation on admission to rehabilitationafter stroke best predict walking speed at discharge COSMI UPDATE hypotheses UL MAS would be a predictorof DC walking speed as per correlations and significance	1. EXCELLENT 2. FAIR 3. EXCELLENT 4. FAIR 5. GOOD 6. GOOD 7. GOOD 8 EXCELLENT 9 EXCELLENT 10EXCELLENT COSMIN UPDATE 1. adequate 2 very good 3 verygood	fair excellent adequate	when significant predictors of discharge walking speed (admission walking speed 10MWT, modified elderly mobility scale score, MAS item 1to 5, FIM motor component, FIM and TUG, wereentered into multiple linear regression, dischargewalking speed was best predicted by 10MWT at admission and MAS item 2 (supine lying to sitting over side of bed) R2 = 0.36. (accuracy of prediction). clinicians could predict using equation: discharge walking speed m/s = 0.33 + 0.47 admission walking speed + 0.05 Item 2 MAS score. OTHER: admission walking speed (10MWT) (n=120) relationship with discharge walking speed (from univariate analysis) 0.32 (<0.001), MAS item 6 (n=105) 0.09 (0.14) item 70.06 (0.23) item 8 0.07(0.36), FIM 0.25 (>0.001).	minus (0/1) UL MAS did not contribute

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
UL MAS	Loewen & Anderson 1990	aim to identify the objective indicators of motor, functional, walking and arm recovery at discharge from hospital in patients admitted acutely afterstroke COSMIN UPDATE hypotheses 1 UL MAS 1 mo with dc UL MAS #1, one weekwith dc UL MAS#1, 1 week with DC Item 8 #1, 1 month with DC item 8 #1, 1 weekUL MAS with DC BI #1, 1 month UL MAS with DC BI #1	1. EXCELLENT 2 EXCELLENT 3FAIR 4 FAIR 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 EXCELLENT COSMIN UPDATE 1. very good 2 adequate 3 very good	poor - excellent	predictors of stroke outcome by spearman correlations; UL-MAS at 1 week with mMAS 0.80, at 1 month with mMAS 0.90. Predicting upper arm function at discharge using upper armfunction initial 0.70, 1 week 0.84, one month 0.91. and predicting UL-MAS: upper arm function at 1week 0.81, 1 month 0.87. When using UL-MAS to predict UL-MAS at discharge: 1 week 0.86, 1 month 0.94. Summary 1 week results were better predictors than initial scores. Upper arm function (Item 6) and UL-MAS did notreach 0.70 to predict Barthel Scores. the regression equation predicted UL-MAS at discharge from the 1 month results r2 0.95. predicting UL-MAS at dc = 0.63 + 1.02 (UL-MAS1 month result (only applicable to those with > 1 month stay). In summary use 1 month scores to predict outcome at discharge (r=0.94) , better 1 month scores predict better UL motor recovery (mean length of stay 59 days). UL-MAS at 1 week and 1 mo did not reach 0.70 with BI at d/c	? COSMIN UPATE 6 hypotheses UL MAS 1 week andDC item 8 - (<0.70), UL MAS 1 week withDC UL MAS + (0.86), UL MAS 1 month with item8 DC - (<0.70), UL 1 month with dc ULMAS + (0.94), UL MAS 1 week with DC BI - (<0.70), 1 month UL MAS withDC BI - (<0.70) (2/6) overall -
UL-MAS	Miller, Slade, Pallant,Galea 2010	nil stated but programme wasapplied to estimate the capacity of the UL-MAS to distinguish between or stratify groups of stroke participants with differing UL abilities	1. GOOD 2.GOOD 3. GOOD n = 90 observations 4. FAIR 5 GOOD6 GOOD 7 n/a 8 n/a 9 EXCELLENT 10. EXCELENT	fair - excellent	Person Separation Index 0.96 - excellent abilityto stratify participants in acute/subacute strokeparticipants with differing UL motor recovery.	?
Excluded at COSMIN update						

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
10MWT	Baer &Smith 2001	1. nil specific but aimed to examine whether different subclassifications of stroke had discrete patterns of achievement of walking andwhether the gait velocity achieved differed dependingon subclassification.	1. excellent 2. good 3. excellent 4fair 5 good 6 good 7 n/a 8 n/a 9 excellent 10 excellent	fair - excellent	of the 185 participants, 164 88.7% achieved a 10mWT in a median of 8 days IQR3-25 days. For those with PACI 71 or 96% achieved 10mWT median 6 days, LACI 48 100% achieved 10mWT in median 8 days, POCI 35 or 97.2% achieved in median 9 days and TACI 10 or 37%achieved in 81.5 days. LACI, PACI or POCI achieved most rapidly. a Kruskal-Wallis test on data for days taken to achieve a 10m walk (H=22.586, N=164, df=3) found the results to be significant.	? No hypothesiss
10MWT	Bower, McGinley, Miller, Clark 2014	hypothesised that 1) dynamic Wii Balance Board (WBB) variables would correlate more strongly with clinical tests of dynamic balance than static WBB variables 2)better performance on the WBB assessment would be associated with improved performance on the clinical tests	1. XCELLENT 2 GOOD 3. FAIR sample size 4. EXCELLENT 5. EXCELLENT 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 EXCELLENT	good - excellent	correlations between 10MWT and WBB tests 0.08 - 0.47. Static standing with eyes open and closed, static weight bearing asymmetry, dynamic sit to stand, dynamic mediolateral weight shifting. Hypothesis correct with dynamic MLWS (moderate strength) 0.47. Eyes open total centre of pressure and 10MWT -0.44 hypotheses not clearly supported. overall poor to moderate correlations between dynamic WBB and 10MWT.	minus
10MWT	Smith & Baer 1999	nil hypothesis - investigatedwhether a simple standardised clinical test ofmobility could be used to provide a detailed representation of recovery profiles for clinically identifiable subgroups of stroke	1. excellent 2. excellent 3 excellent 4 fair 5 good 6 good 7n/a 8 n/a 9 fair - nil data re training of staff 10 excellent unsure just calculated percentages	fair - excellent	77.7% achieved 10mwt in median time of 9 days. PACI, LACI or POCI achieved faster andhad shorter hospital stay. 10mWT can be included to predict timeframes for recovery.	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
10MWT	Vernon, Paterson, Bower, McGinley, Miller, Pua & Clark 2015	the addition of Kinect TUG variables combined with total TUG time would strengthen prediction models for performance on other clinical tests	1. EXCELLENT 2 EXCELLENT 3. FAIR sample size 30 4. GOOD 5 EXCELLENT 6 GOOD 7 EXCELENT 8 EXCELLENT 9 EXCELLENT 10 EXCELLENT	good - excellent	10MWT had strongest association with Kinect- TUG variables $R^2 = 0.86$	plus
10MWT	Wolf, Catlin, Gage, Gurucharri , Robertson& Stephen1999	Is the E-FAP (Emory Functional Ambulation Profile)correlated with previously validated measures of walking speed and balance (e.g. 10MWT).	1. GOOD (appears to be 0 but not stated) 2. NA 3. GOOD 4. FAIR - intro alludes to expectations 5 GOOD 6 GOOD7 EXCELLENT 8 fair 9 fair - sampling, high level ? Training ofraters 10 EXCELLENT	excellent poor	Pearson product moment correlation coefficient 10MWT with Berg Balance without stroke r=0.052 P= 0.7946 with stroke r=0.627 P0.0004 Functional reach test without r=0.307 P=0.1123 with stroke r=0.349 P=0.0690 with E-FAP withoutr=-0.759 P=0.0001 with strokes with assistive devices r=-0.708 P=0.0001 with stroke without calculating assistive devices r=-.783),p=0.0001. Summary E-FAP and 10mWT negatively correlated slow times on E-FAP correlated with slow gait speeds.	?
9HPT	Alusi, Worthington, Glickman, Findley & Bain 2000	no specific hypothesis but does state in article - This study evaluates the constructvalidity, intrarater, and inter rater reliability of this scale when used in three different ways to assess upper limb tremor in MS	1. EXCELLENT 2. GOOD - not clear but can deduce have been excluded from the analysis 3. POOR sample size 4. GOOD construct hypothesis 5. GOOD 6.GOOD 7 EXCELLENT 8 GOOD 9 EXCELLENT 10 EXCELLENT	Fair - Excellent	spearman's correlation coefficient; 9HPT with three tremor assessments; 1) 9HPT & On Posture -0.62, 2) 9HPT & Spirals (dominant hand) -0.74 3) spirals non-dominant hand -0.87 4) handwriting -0.78. 9HPT provide useful objective measure of UL function in patients withMS and tremor	plus
9HPT	Fisk, Brown, Sketris,Metz & Stadnyk 2005	no hypothesis looking for evidence of construct validity	1. 3 did not complete the HRQoLutility measures EXCELLENT 2. EXCELLENT 3. EXCELLEN n=187 4. FAIR 5 GOOD 6 GOOD7 EXCELLENT 8 FAIR 9 excellent 10. EXCELLENT	fair - excellent	Spearman correlations between 9HPT and; SF- 6D -0.41 (moderate), EQ-5D -0.56 (moderate) HUI Mark III -0.65 (strong) evidence of constructvalidity	plus - moderate and above correlations

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
9HPT	Marrie & Goldman2011	we expected moderate correlations between Tremor and Coordination Scale and hand function. Also used 9HPT to assess criterion validity of TACS - nil gold standard so hypothesis testing used	1. not stated but can deduce nil missing items GOOD 2. GOOD 3.EXCELLENT as power to detect correlation was performed requiring n=40 4 GOOD 5 GOOD6 EXCELLENT 7 EXCELLENT 8 GOOD 9 study is part of a larger study so more details would be listed there but nil specific reference to that blinding of rater, training FAIR 10. EXCELLENT	Fair - Excellent	9HPT correlates moderately with TACS (tremorand coordination scale) in pts with MS r: -0.51 95% CI -0.70 to -0.29). 9HPT weakly correlateswith age, BMI and cerebellar functional systemscore	plus
9HPT	Rossier & Wade 2002	nil specific hypotheses but aim to establish the validity and reliability of the GNDS - reference to anticipated equal scores with other measures but not clearly stated for all measures	1. EXCELLENT 0.39% 2 EXCELLENT 3 FAIR sample size4 FAIR 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 FAIR volunteers Generalisability 10 XCELLENT	fair - excellent	Pearson correlation between Guys Neurological Disability Scale (GNDS) and 9HPT for people with MS - face to face grp n=22 0.74, postal grp (GNDS posted) n=23 0.70 combined group n=430.71.	?
9HPT	Sunderland, Tinson, Bradley & Hewer 1989	not clearly stated; the study investigated the relationship between grip strength, spasticity and functional recovery to discover whetherin fact it (grip strength) may be a valuable marker of recovery in the typical stroke patient.	1. EXCELLENT yes 7 lost to follow-up 2. GOOD - not clear butcan see that n=31 were included in the analyses 3. FAIR sample size, 4. FAIR 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 FAIR - limited information regarding statistical analysis, nil info re raters, blind unable to determine if any biases	Fair - Excellent	the 9HPT has a weak correlation with percentage grip r:0.71 at the initial assessment and at 6 month follow-up r:0.79. 9HPT was the weakest out of motricity index, frenchay Arm Test, Motor club assessment. The 9HPT resultsat 1 month when used to predict performance of Frenchay Arm Test wrongly classified 27%, wasthe lowest ability to predict out of Motricity Index,Percentage grip, Motor Club	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
AQoL	Turner-Stokes, Baguley, De Graaff,Katrak, Davies, McCrory & Hughes 2010	nil but research questions -does GAS provide added value as a responsive indicator over other measures, (2) how does it relate to other measures	1. EXCELLENT n= 6, 2. GOOD not described but can de deucedwere omitted 3. GOOD sample size, 4 FAIR 5 GOOD 6 GOOD 7FAIR 8 FAIR 9 EXCELLENT 10 EXCELLENT	fair	GAS outcome T-score correlation with AQoL 0.07	?
ARAT	Barden, Baguley, Nott, Chapparo 2014 (a)	nil specific - aimed to evaluate UL performance changes in adults with UMN syndrome who received BTX-A injections for UL muscle spasticity in an outpatient clinical setting by (1) evaluating change in UL performance following BTX-A injections as measured by DCD and current clinical measures. (2) mapping observed changes to the Body Function and Structure and Activity domains of the ICF	1. Not stated GOOD 2. N/A 3. POOR sample size 4. FAIR 5. GOOD 6 GOOD 7. EXCELLENT 8. FAIR only dynamometry information on measurementproperties 9EXCELLENT 10 EXCELLENT	poor - excellent sample fair-excellent	Spearman rank order correlations ARAT & DCD components; Voluntary gip work % 0.07, maximum force kg, 0.50, minimum force kg 0.10,contraction duration s -0.54, relaxation duration s -0.56, max Fvel kgs-1 0.63, min Fvel kgs-1 -0.56. summary ARAT moderate to good relationship with increased speed of force generation and shorter contraction and relaxation duration	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
ARAT	Barden, Nott, Heard, Chapparo& Baguley2012	to explore the relationship between DCD and functional measures of UL performance(ARAT) in adults with ABI	1. excellent 2. excellent 3. fair 4.fair 5. good 6. good 7 excellent 8fair 9. excellent (some methodological issues but not related to ARAT section)	excellentfair	fair (0.25 to 0.50) to good (0.50 to .075) statistically significant correlations between isometric force, force velocity, isometric grip work, and the total ARAT score. Relationship between computerized hand dynamometry motor elements and ARAT total score (Spearman Rank order correlation); negative UMN features impacting on grasp; max isometricforce p: 0.55; voluntary isometric grip work (%) p: 0.59, max F(velocity) p: 0.73; contraction duration p: -0.60. Negative UMN syndromes measures had a positive relationship with the total ARAT score signifying improved ability to grasp is associated with a higher total ARAT score. Positive UMN syndromes impacting on release; min isometric force p: -0.34; involuntary isometric grip work p: 0.59; minimum F (velocity)p: -0.69; relaxation duration p: -0.55 Positive UMN measures had a negative relationship with total ARAT scores, indicating that increasing difficulty in releasing the dynamometer was linked to deteriorating ARAT performance. ARATperformance could predict performance of hand grasp and release in pts post ABI with UL spasticity	?
ARAT	Barreca, Stratford, Lambert, Masters& Streiner 2005	theory was that measures designed to assess a similar attribute should correlate more highly with CMSA arm-hand sum and with the ARAT than with the CMSA shoulderpain score. than	1. missing items not given but can be deduced GOOD 2 GOOD can be deduced 3. FAIR 4. GOOD 5 EXCELLENT 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 EXCELLENT	Fair - Excellent	ARAT correlated highly with CAHAI 0.93 (0.87-0.96) at initial assessment and 0.93 (0.87-0.96)at follow up, ARAT and CMSA arm-hand sum correlated highly at initial Ax 0.87 (0.76-0.93) and at follow up 0.92 (0.85-0.96). ARAT and CMSA shoulder pain low correlation at initial 0.52 (0.24-0.72) and follow up 0.40(0.10-0.64). ARAT measures what is intended - UL function.	PLUS

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
ARAT	Barreca, Stratford, Masters, Lambert, Griffiths 2006	not stated including the ARAT	1. % of missing items not given can see n=105 at initial and n=100 at follow up GOOD 2. notclear how handled FAIR 3. excellent 4. unsure as used ARAT as comparator so hypothesis excluding the ARAT and referring to 2 CAHAI versiondocumented FAIR 5.EXCELLENT 6 GOOD 7.EXCELLENT 8 EXCELLENT 9EXCELLENT 10 EXCELLENT	Fair - Excellent	The convergent cross-sectional construct validityPearson correlation coefficients showed no difference in the magnitude of the coefficients forthe 2 versions at either the initial or follow-up assessment with the ARAT (Z= 0.00, P1 = 0.500. Pearson correlation coefficient (95%CI) initial ax: ARAT & CAHAI-13 0.93 (0.90-0.95), ARAT & CAHAI-9 0.93 (0.95-0.95), follow up ax:ARAT & CAHAI-13 0.95 (0.93-0.97), ARAT & CAHAI-9 0.95 (0.93-0.97).	?
ARAT	Barreca, Stratford, Masters, Lambert, Griffiths, McBay 2006	1. to estimate the test-retest reliability, cross sectional validity, and longitudinal validity of the three shortened versions of the CAHAI.	1. not stated but can deduce is 0GOOD 2. GOOD 3 FAIR 4. not stated POOR 5 GOOD 6 GOOD7 EXCELLENT 8 EXCELLENT 9EXCEELENT 10	Poor - Excellent	Pearson correlation coefficient (() are one sided95% confidence limit) was calculated for convergent cross sectional validity. Initial assessment ARAT & CAHAI-13 0.93 (0.88), ARAT & CAHAI-9 0.94 (0.90), ARAT & CAHAI-80.95 (0.91) ARAT & CAHAI-7 0.95 (0.91) ARAT & CMSA 0.87 (0.78) at follow up ARAT & CAHAI -13 0.93 (0.88) ARAT & CAHAI-9 0.94 (0.90) ARAT & CAHAI-8 0.94 (0.90) ARAT & CAHAI-70.94 (0.90) ARAT & CMSA 0.92 (0.86). In summary high correlations between all versions of CAHAI and ARAT and CMSA. Convergentconstruct validity	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
ARAT	Blennerhassett, Avery & Carey 2010	investigate the concurrent validity of the HFS as a measure of hand function in people with stroke (against ARAT)	1, % missing items given EXCELLENT, 2 no FAIR, 3 POOR sample size, 4. FAIR no specific hypothesis formulated 5direction of correlations not stated in hypotheses GOOD, 6 expected magnitude not stated GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT	Poor - Excellent - samplefar - excellent	stats used: Lin's concordance (for continuous variables) and Spearman Rho (mean data must have been converted to ranks ?this is represented by % change score) which provide an index of the concurrent validity of the hFS as a measure of hand function. Performance scoresfor HFS and ARAT were found to be in strong agreement for both baseline (Rho_c = 0.96, 95%CI: 0.90-0.98) and follow up (Rho_c = 0.95, 95%CI: 0.87-0.98). Strong correlations observed between to measures at baseline (Rho = 0.91, P<0.001) and follow up (Rho = 0.89, P<0.001). performance scores for HFS and ARAT at followup were also found to be moderately to strongly related to those at baseline for both HFS (Rho= 0.74, P<0.001) and ARAT (Rho=0.9., P<0.001). IN SUMMARY A MODERATE AGFREEMENT BETWEEN CHANGE SCORES FOR ARAT ANDHFS	?
ARAT	Celik, O'Malley, Boake, Levin, Yozbatiran & Reistetter 2010	to report correlations of fourrobotic measures to widely used clinical measures	1.not stated but can be deduced GOOD 2. GOOD 3. POOR n=9, 4 POOR 5. GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 GOOD only included mildly impaired participants EXCELLENT 10 EXCELLENT	Poor - Excellent	ARAT correlates highly with trajectory error and smoothness of movement as measured via robotics ARAT and TE -0.83 ARAT and SM 0.51but does not correlate with hits per minute and mean tangential speed (all are measures for a target hitting task that involved repetitive reaching movements).	? (no hypotheses) or plus andminus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
ARAT	Edwards, Lang, Wagner, Birkenmeier & Dromerick 2012	nil specific hypothesis but stated what is the relationshipbetween the WMFT and ARAT (concurrent validity)	1 yes EXCELLENT 2 GOOD 3 GOOD 4 POOR no hypothesis,question stated but nothing regarding correlations and differences 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT EXCELLENT 10 EXCELLENT - results talks about statistically significand and not directional or magnitude so perhaps mark down despite correct stat analysis	Poor - Excellent	Pearson product moment correlation was used to determine concurrent validity at 3 time points.ARAT total score and WMFT FA function score:Day 0: 0.745, day 14: 0.827 day 90: 0.863 WMFT FA time score and ARAT day 0: - 0.641,Day 14: -0.825, day 90: -0.772 WMFT FA grip score and ARAT day 0: 0.702 day 14: 0.631 day90: 0.553. coefficients with an absolute value of >0.40 were statistically significant at the P<0.01level and coefficients >0.30 were statistically significant at the P<0.05.	?
ARAT	McDonnell , Hillier, Ridding & Miles 2006	…to compare a range of grip- lift parameters in the affected and unaffected upper limbs in a heterogeneous sample of stroke patients and to correlate them with two widely-used indices of motor function.	1. excellent 2 Good 3 poor n= 174 FAIR 5 GOOD 6 GOOD 7 GOOD 8 FAIR 9 Flaws in design 10 EXCELLENT	excellent poor	ARAT and FMA correlated highly/significantly (p=0.75), P<0.001) ARAT correlated more highlythan the FMA with grip strength (p=0.73 P<0.001) and tapping speed (p=0.61 P<0.001). Univariate analysis , P<significant negative correlation between ARAT and pre load duration (p=-0.72 P<0.001) and positive correlation with the max dGF/dt and dLF/dt correlation coefficient(p= 0.83, P<0.001) correlations were not influenced by whether affected hand was dominant or not. Combination of preload duration and max correlation coefficient explained 60% of variance of the ARAT and 38%of FMA. Adding grip strength and tapping speed increased this to 71% of the ARAT and 59% of FMA. IN SUMMARY objective parameters the grip lift task correlate more strongly with the ARAT than the FMA, further validating the use ofthe ARAT as a clinical test of hand function.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
ARAT	Notley, Turk, Pickering, Simpson & Burridge2007	to examine the relationship between the quality of wrist movement during a tracking atarget on screen task, comparing three different indices of accuracy of tracking and upper limb function as measured by the ARAT	1. EXCELLENT 2 GOOD 3 POOR 4 FAIR 5 GOOD 6 GOOD7 EXCELLENT 8 POOR 9 FAIR Poor representative sample 10EXCELLENT	Poor - Excellent	Pearson correlation coefficients used for relationship. Mean RMS error & ARAT r= -0.441,P=0.202 (not strongly correlated and did not achieve statistical significance), cross correlation & ARAT r=0.799 P = 0.006, signal to noise ratio & ARAT r = 0.829 p = 0.003. Significant correlation between other two % ARAT with direction of relationship consistent with greater accuracy in tracking being correlated with better function as measured by ARAT.	?
ARAT	O'Dell, Kim, Rivera, Fieo, Christos, Polistena, Fitzgerald& Gorga 2013	AMAT -9 will hold significantrelationships with the ARAT	1. yes EXCELLENT 2. not statedbut can be deduced FAIR 3 FAIR4 GOOD, 5 EXCELENT 6 GOOD7 EXCELLENT 8 EXCELLENT 9 FAIR - poor representation of population, assessor not blinded,measures not randomly administered 10 GOOD	Fair - Excellent	ARAT correlates with the AMAT-9 0.79 (spearman rank correlation coefficient).	plus
ARAT	Page, Hade & Persch2015	the w/h UE FM would displayhigh concurrent validity with the ARAT (a value of less than or equal to 0.70 represented satisfactory association between measures	1. GOOD as not clear 2. N/a 3.FAIR sample size 4. GOOD 5. EXCELLENT 6 EXCELLENT 7.EXCELLENT 8. EXCELLENT 9 EXCELLENT 10 EXCELLENT	fair - excellent	concurrent validity (spearman rank correlationcoefficient r)- w/h UE FM a measure of UE impairment with ARAT a measure of functionallimitation pre test 1 0.74 (P<0.001) pre test 2 0.67 (P<0.001) in sample with only palpable movement in wrist flexors	plus
ARAT	Page, Levine, Hade 2012	nil specific hypothesis stated,objective is stated; …reportsthe concurrent validity of w/hUE FM with an established, stroke specific measure of	1. GOOD missing items not described can deduce from article nil, 2 GOOD, 3. POOR sample size 29 4. FAIR - vaguebut possible to deduce expectation 5 GOOD 6 GOOD 7EXCELLENT 8 EXCELLENT 10EXCELLENT	Fair - Excellent	concurrent validity (established using spearman rank correlation coefficient. ARAT & w/h UE FM concurrent validity 0.72 (P<0.001) indicating highcorrelation (a value above 0.70 represents a high association between measures)	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
		distal paretic UE movement, ARAT				
ARAT	Stinear, Barber, Petoe, Anwar &Byblow 2012	aim was to test and refine thePREP algorithm by comparing the patient stratification predicted by the algorithm with the stratification produced by an unbiased cluster analysis of upper limb function. (to test and refine an algorithm for predicting potential recovery of upper limb function after stroke)	1. EXCELLENT 2 FAIR 3 FAIR 4.POOR 5. GOOD 6 GOOD 7 GOOD 8 FAIR 9 FPOOR no info re assessors, blinding, bias, training 10 EXCELLENT	Poor - Excellent	The SAFE score at 72hrs (r=0.79, P<0.001)ARAT score at 2 weeks (r=0.85, P<0.001) and Fugl-Meyer score at 2 weeks(r=0.86, P<0.001) all positively correlated with ARAT score at 12 weeks. ARAT score at 12weeks could be predicted by SAFE score and fractional anisotropy asymmetry index measuresat 2 weeks when combined in a stepwise manner according to the PREP algorithm.	?
ARAT	Urbin, Waddell & Lang 2015	ratio and paretic UE metricswould exhibit a strong association with the ARAT score.	1. GOOD 2. FAIR 3 FAIR samplesize 4 GOOD 5 EXCELLENT 6 EXCELLENT 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 EXCELLENT	fair - excellent	the ARAT correlated with acceleration metricsderived from outside of treatment ranged from 0.73 - 0.85, P<.001 (significant and strongcorrelations)	plus
BI	Khan, Pallant &Turner-Stokes 2008	nil specific but objectives (1) explore the type and nature of person-centred goals that arecommonly set and achieved during the program (2) compare the responsiveness and relative efficiency of GASwith the FIM and BI as outcome measures for rehab in MS	1. not stated GOOD 2 N/A 3. POOR sample size 4 fair "expected correlations in results but not stated prior so cannot be higher than fair 5 GOOD 6 GOOD 7 GOOD 8 EXCELLENT 9excellent 10 FAIR - SRM and Effect sizes	fair - excellent	spearman rank correlations between measuresGAS T-score at d/c; with GAS change score 0.95, with BI change score -0.25, with FIM change score -0.16, CGI -0.86. GAS change score with BI change score -0.15, with FIM change score -0.06 with CGI -0.77. Over half ofthe goals chosen were in areas not included in the FIM or BI.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
BI	Maujean, Davis, Kendall, Casey & Loxton 2014	"it was expected that there would be a low correlation between the daily living selfefficacy scale and the BI	1. excellent 2 n/a 3 good samplesize 4 good 5 excellent 6 good 7excellent 8 excellent 9 excellent 10 excellent	good	BI and DLSE r = 0.28 low	plus
BI	Rossier & Wade 2002	we anticipated a reasonablyclose relationship between GNDS and BI	1. EXCELLENT 0.39% 2 EXCELLENT 3 FAIR sample size4 FAIR 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 FAIR volunteers Generalisability 10 EXCELLENT	fair - excellent	BI with GNDS face to face -0.51, postal -0.52,total -0.76, BI with EDSS face to face -0.76, postal -0.73, total -0.86	plus
BI	Sprigg, Selby, Fox, Berge, Whynes, Philip & Bath 2013	we sought to examine patients with very poor HRQoL and aim to describepts with very low HRQoL scores, including baseline factors and functional outcome to determine characteristics associated with a health status worse than death	1. excellent 2 excellent 3 excellent 4 poor 5 good 6 good 7excellent 8 poor 9 excellent 10 excellent	poor	Health utility score of EQ-5D was correlated withBI r = 0.84; P<0.001, EQ-5D VAS r = 0.58; P<0.001.	?
BI	Wolf &Koster 2013 BI comparator not included	" the persons own perceived recovery, as measured by theSIS, may be a more significant predictor of the amount of retained HDL activities after stroke" High demand leisure (HDL)	1. excellent 2 good 3 good 4 good 5 good 6 good 7 excellent 8excellent 9 fair - not a very representative sample, mild stroke and volunteers, minimal information in data collections asfrom registry 10 excellent	fair	logistic regression indicated that SIS total perceived recovery and SIS strength domain were the only statistically significant factors determining percent of retained HDL following mild stroke, SIS total OR1.027, 95%CI 1.0-1.056P0.05, SIS strength domain OR 1.033, 95%CI 1.007-1.061 P0.01. BI total score did not show statistical significance.	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
BI	Aldermanet al 2001	characteristics of neurological populations may significantly reduce the validity of the EQ- 5D, this will be low because ofreliance on self-report, poor ability to reflect rehab gains through limited range of ratings and inadequate number of items failing to capture complexity of needs.	1 good 2 fair 3. poor 11 4 fair 5 good 6 good 7 fair 8 fair - reference to study that gives same 9 - 10 excellent	fair	Staff completed EQ-5D Index and EQ-5D VAS with BI .48, p<.001, 0.33, p=.02. Individual itemsof EQ-5D with the BI 0.04 (anxiety and depression) to -0.55.	plus
CMA	Barreca, Stratford, Masters, Lambert, Griffiths, McBay 2006	1. to estimate the test-retest reliability, cross sectional validity, and longitudinal validity of the three shortened versions of the CAHAI.	1. not stated but can deduce is 0GOOD 2. GOOD 3 FAIR 4. not stated POOR 5 GOOD 6 GOOD7 EXCELLENT 8 EXCELLENT 9EXCEELENT 10 EXCELLENT	Poor - Excellent	Pearson correlation coefficient (() are one sided95% confidence limit) was calculated for convergent cross sectional validity. Initial assessment CMSA & CAHAI-13 0.81 (0.68), CMSA & CAHAI-9 0.84 (0.73), CMSA & CAHAI-8 0.84 (0.73) CMSA & CAHAI-7 0.85 (0.75) ARAT & CMSA 0.87 (0.78) (note ARAT and CAHAI version correlations 0.93-.95) at follow upCMSA & CAHAI - 13 0.89 (0.81) CMSA & CAHAI9 0.91 (0.84) CMSA & CAHAI-8 0.91 (0.84) CMSA & CAHAI-7 0.91 (0.84) ARAT & CMSA 0.92 (0.86). In summary high correlations between all versions of CAHAI and ARAT and CMSA. ARAT slight higher across all versions. Convergent construct validity	?
CMA	Ellis, Sukal, DeMott & Dewald 2008	kinematic variables from theArm Coordination Training Device 3D (ACT) are relatedto existing standardised clinical assessments of arm movement following stroke	1. good- not specified assumed 02 n/a 3 n=11 poor 4 good 5 good6 good 7 excellent 8 excellent 9 minimal generalisability high functioning FAIR 10 excellent	poor-excellent	CMAa correlated (Spearman Rank Correlation)with 100% (reaching at limb weight) 0.72 and 175% (reaching while transporting an object) 0.74. Other 0% 0.44, 25% 0.76, 50% 0.56, 75% 0.74, 125% 0.64, 150% 0.68, 200% 0.38. CMAh results not given as not positive nor significantrelationship -no correlation. Note also no correlation with MAS - spasticity	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
CMA	Manns, Tomczak, Jelani, Cress & Haennel 2009	nil hypothesis but purpose 1.determine feasibility of CS-PFP10 for measurement of physical function in stroke survivors, 2. explore associations among physical functional performance, ambulatory activity and VO2 peak	1. GOOD 2. n/a 3 poor sample size n=10 as controls did not complete CMA 4. poor 5 good 6good 7 excellent 8 excellent 9 excellent - nil other than samplesize 10 excellent	poor - excellent	Pearson correlation with CMA & continuous scale physical functional performance 10 item test 0.667 (0.40 significant) CMA with Vo2 peak0.300 (0.433) with steps a day 0.500 (0.170). Summary CMA negatively associated with CS-PFP10 but not associated with VO2peak or ambulatory activity	
CMA	Oczkowski& Barreca 1993	nil hypothesis "this article further clarifies the use of theFIM as a prognostic indicatorof outcome in stroke patients"	1. EXCELLENT - 3 died 2. FAIR3 EXCELLENT 4 FAIR 5 GOOD6 GOOD 7 POOR 8 POOR 9 EXCELLENT 10 EXCELLENT	poor - excellent	The degree of recovery of postural control as measured by the CMA highly correlated with admission FIM. Admission postural control via CMA correlated highly with discharge location. Postural control at admission (median stage) 495%CI 4-4.5 predicted dc home, stage 3 95%CI2.5-3.5 predicted nursing home and 1.5 95CI% 1.0-2.0 predicted chronic care. multiple logistic regression showed best predictors of location ofdc were admission FIM score, admission postural staging (CMA) and age, the most powerful was admission FIM score.	?
CMSA	Semrau, Herter, Scott, Dukelow2015	examine the relationship of robotic measures to existing clinical measures across thefirst 6 months post stroke.	1. Excellent 2. Excellent 3. good sample size, 4 fair - unclear whatwas expected 5 good 6 good 7 excellent 8 fair 9 reference to butno properties stated) 10 excellent 11 excellent	fair - excellent	robotics assesses; position sense, kinaesthesia, motor function and simultaneous bilateral motor function (arms). Assessments were across the following time points post stroke T1 1 week, T2 6weeks, T3 12 weeks T4 26 weeks. Majority had mild stroke(76), 35 had moderate and 2 severe. Correlations between 4 robotic measures acrossthe 4 time points (1, 6, 12, 26 wk) = -0.21 - -0.79.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
CMSA	Barreca, Stratford, Lambert, Masters& Streiner 2005	theorised that CAHAI shouldcorrelate more highly with CMSA arm-hand sum and with the ARAT than with the CMSA shoulder pain score.	1. missing items not given but can be deduced GOOD 2 GOOD can be deduced 3. FAIR 4. GOOD 5 EXCELLENT 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 EXCELLENT	Fair - Excellent	CMA arm hand score correlated highly with CAHAI at initial 0.81 (95% CI 0.66-0.90) follow up 0.89 (95% CI 0.80-0.94). CMA arm hand with ARAT initial 0.87 (95% CI 0.76-0.93) follow up 0.92 (0.85-0.96). CMA shoulder pain did not correlate highly with CAHAI and ARAT 0.39 - 0.55 across initial and follow up. CMA sections measure what is intended supports convergentand discriminant validity.	PLUS
CMSA	Coderre,Zeid, Dukelow, Demmer, Moore, Demers, Bretzke, Herter, Glasgow, Norman, Bagg & Scott 2010	"hypothesised that many stroke patients would show deficits in one or more of the attributes of sensorimotor performance, furthermore wealso expected to find performance asymmetries across their two upper limbs". Nil specific to clinical measures	1. excellent 0 2. n/a 3. good sample size 4. poor nothing in relation to clinical measures, not in data analysis either i.e. lookingat correlations just that the clinical measures were used as areference for results of robotics 5.good 6 good 7 excellent 8 fair (reliability) 9 fair - high functioning, the robotics was effectively inly using completed attempts - best effort, no information regarding assessors 10. excellent	poor - excellent	left affected participants had a significant association between CMSAa and sensorimotor attributes P < .05. right affected participants hada significant association between most sensor motor attributes and CMSA a (P < .05) except UL postural control and feedforward control (P > .05). left affected participants more than right, scored perfect 7 on CMSA arm but still scored with a sensorimotor impairment within feed- forward control (n=5) , feedback control (n=5), and total movement metrics (n=5). (out of n=26).	
CMSA	Levin, Desrosiers, Beauche min, Bergeron& Rochette 2004	nil stated just alluded to testing of concurrent validity	1. Good 2. n/a 3 poor sample size 4. fair 5 good 6 good 7 excellent 8 excellent 9 poor - sample very selective, not generalisable, unknown how non participants varied 10 excellent	fair - excellent	CMA hand section correlated with reaching performance (RPS) close target 0.95 far target0.93, and CMA arm section RPS close target 0.92 far target 0.90. demonstrated concurrentvalidity	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
EQ-5D	Ali, Fulton, Quinn, Brady on behalf of VISTA 2013 EQ-5D comparator	" we hypothesised that existing primary outcome measures (mRS, BI, NIHSS)may also reflect patient quality of life"	1. Excellent 2. GOOD 3 excellent4 fair 5 good 6 good 7 excellent 8fair 9 excellent 10 excellent	fair - excellent	at 3 mo patient responses to QoL had a strongerassociation with mRS (SIS n=2970 P<0.0001, r = -0.71, r2 = 0.52, EQ-5D weighted score n=2987 r=-0.7 r2=0.53) where as proxy responses had a stronger association with BI (SIS n=867 P<0.0001, r = 0.68, r2 = 0.48, EQ- 5D n=837, r=0.78, r2= 0.63) than with NIHSS. BI had more mismatches between good primaryoutcome and poor QoL (EQ-5D 11.3%, SIS 19.2%) than the mRS (EQ-5D 8.5% SIS 10%) but less than NIHSS (29%, 23.9%)and for poorprimary outcome and good QoL less mismatches than mRS(EQ-5D 3.1%), (SIS 4.1%) but more than NIHSS (0%)	plus
EQ-5D	Jenkinson Fitzpatrick , Crocker& Peters2013	nil stated - study aim to validate SIS in UK setting anddevelop SIS short form and SIS index.	1. excellent 2. excellent 3. good sample size 4. fair 5 good 6 good7 excellent 8 excellent 9 fair % ofmissing responses may have biased results 10 excellent	fair - excellent (without sample influencing	the short-form SIS (SF-SIS)_ correlated with SIS 0.98 P<0.001, SIS index and SF-SIS indexcorrelated identically with the EQ-5D 0.83 P<0.001.	?
EQ-5D	Lunde2013	"our main objective is to assess whether the EQ-5D and 15D can be used interchangeably in cost utilityanalysis and agreement between the two measures	1. excellent 2 n/a 3 excellent 4 excellent 5 good 6 good 7 excellent 8 poor 9 excellent 10 good	poor	EQ-5D and 15D r = 0.80, outside scope of studybut findings indicate the two cannot b used interchangeably in cost analysis.	plus
EQ-5D	Sprigg, Selby, Fox, Berge, Whynes, Philip & Bath 2013	we sought to examine patients with very poor HRQoL and aim to describepts with very low HRQoL scores, including baseline factors and functional outcome to determine characteristics associated with a	1. excellent 2 excellent 3 excellent 4 poor 5 good 6 good 7poor 8 poor 9 excellent 10 excellent	poor	Health utility score of EQ-5D was correlated withBI r = 0.84; P<0.001, EQ-5D VAS r = 0.58; P<0.001.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
		health status worse than death				
FIM	Chen, Chen, Hreha, Goedert & Barrett 2015	"we examined whether neglect severity at admissionpredicted DC FIM scores	1. excellent 2 fair 3 excellent 4 poor 5 good 6 good 7 excellent 8excellent 9 excellent	poor - excellent	greater neglect severity (KF-NAP score) at admission predicted lower FIM DC scores (b +/-SE, -.003 +/-.015; 95%CI -.063 to -.003; ß= -.23, P = .033)	?
FIM	Corrigan, Smith- Knapp & Granger 1997	COSMIN UPDATE study hypotheses 1) FIM will predictamount of direct assistance a person requires; prediction will be improved with measures of neurobehavioral impairment; predictive ability of FIM will exceed SF-36. 2) FIM will predict amount of supervision a person requires; prediction will be improved with measures of neurobehavioral impairment; predictive ability of FIM will exceed SF-36 3) FIM will predict amount of combined direct assistance and supervision required; prediction will be improved with measures of neurobehavioral impairment; predictive ability of FIM will	COSMIN UPDATE marked it down due to vague hypothesis. So while they highlight they expect the FIM to predict direct assistance it provides no clarity about accuracy nor does it clarify how much the FIM should exceedother tools. For this reason I felt you couldn't confidently say that the FIM was better than other assessments.	doubtful	FIM predictive of minutes of assistance ($p<0.0001$), supervision ($p<0.0063$) and needother type of assistance ($p<0.0032$). Accuracy of FIM prediction more superior thanSF-36.	plus 9/9

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
		exceed SF-36 TOTAL 9 HYPOTHESES				
FIM	Cusick, Lannin, Hanssen& Allaous 2014	nil hypothesis bit question "what is the concurrent validity of the Western Neuro Sensory Stimulation Profile with the FIM?	1. excellent 2 excellent 3 fair sample 4. poor 5 good 6 good 7excellent 8 excellent 9 excellent10 excellent	fair	admission FIM and WNSSP total scores rs = - 0.146, P = 0.0424 (weak and significant), discharge FIM and WNSSP not significant rs = 0.382, P = 0.41, two tailed) FIM not ideal in slow to recover pts to measure	?
FIM	Edwards, Hahn, Baum & Dromerick 2006	"…we investigated measuresof life satisfaction and participation in meaningful activities in patients 6 months after mild stroke"	1. EXCELLENT 2 N/A 3 EXCELLENT 4 FAIR 5. GOOD 6 GOOD 7 EXCELLENT 8 POOR 9 EXCELLENT 10 EXCELLENT	fair - excellent	FIM was not a significant predictor of Reintegration into Normal Living.	?
FIM	Herrman, Black, Lawrence, Szekely & Szalai 1998	nil specific - to prospectivelyfollow a large sample to determine the frequency, severity, and course of depressive symptoms, their clinical correlates and their effect on functional outcome	1. not stated GOOD 2 FAIR 3 EXCELLENT 4 POOR 5 GOOD 6GOOD 7 EXCELLENT 8 GOOD 9FAIR Blinded assessors 10 EXCELLENT	poor - excellent	FIM significantly correlated with depression scales at 3 mo (except sphincter control and communication) -.27 - -.43 p = .0001, at one year social cognition remained significantly correlated with depression measures -.14 to - .34 (P.001)	?
FIM	Joseph, Pandit, Aziz et al2013	"aim to identify hospital admission factors that predictfunctional improvement after rehab based on FIM scores intrauma patients"	1. good 2 fair 3 excellent 4 fair 5good 6 good 7 n/a8n/a 9 fair - minimal m methodological information 10 excellent	fair	Head Abbreviated Injury Score (ß = -2.3; P = 0.004) and hospital LOS (ß = -0.27; P=0.01) were only independent predictors of functionalimprovement (FIM), age was not predictive.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
FIM	Khan, Pallant &Turner-Stokes 2008	nil specific but objectives (1) explore the type and nature of person-centred goals that arecommonly set and achieved during the program (2) compare the responsiveness and relative efficiency of GASwith the FIM and BI as outcome measures for rehab in MS	1. not stated GOOD 2 N/A 3. POOR sample size 4 fair "expected correlations in results but not stated prior so cannot be higher than fair 5 GOOD 6 GOOD 7 GOOD 8 EXCELLENT 9excellent 10 FAIR - SRM and Effect sizes	fair - excellent	spearman rank correlations between measuresGAS T-score at d/c; with GAS change score 0.95, with BI change score -0.25, with FIM change score -0.16, CGI -0.86. GAS change score with BI change score -0.15, with FIM change score -0.06 with CGI -0.77. Over half ofthe goals chosen were in areas not included in the FIM or BI.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
FIM	Kwon, Duncan, Studenski, Perera, Lai & Reker 2006	purpose was to assess the utility of the SIS in a community dwelling stroke survivors with more realistic administration methods such as telephone and mail survey.We focus on the construct validity of telephone SIS administration.	1. EXCELLENT 2. EXCELLENT 3. Good sample size n=95 4. POOR - not clear what was expected no hypotheses 5. GOOD 6 GOOD 7 GOOD - no information presented on FIM only presented for SF-36V 8 GOOD 9 EXCELLENT 10 EXCELLENT	poor - excellent	convergent validity (Pearson correlations all p<0.001)- SIS strength with FIM-motor 0.404, SF36V physical functioning 0.477, SF-36RP role physical 0.533, SF-36V general health 0.460, SF-36V PCS physical component summary score 0.520. SIS Memory with FIM-Cognitive component 0.501, SF-36V general health 0.378, SIS Emotion with SF-36V role emotion 0.504, SF36V mental health 0.713, SF-36V general health0.460, SF-36V mental component summary score 0.692. SIS communication with FIM- cognitive 0.637, SF-36V general health 0.362, SIS ADL/IADL with FIM-motor 0.858, SF36V; physical functioning 0.732, role physical 0.711, general health 0.503, physical component summary score 0.586, SIS mobility with FIM motor 0.738, SF-36V; physical function 0.755, role 0.724, general health 0.574, physical component summary 0.632 SIS hand function with FIM motor 0.659, SF-36V physical function 0.682, role 0.631, general health 0.470, physical component summary 0.628, SIS social participation with FIM motor 0.588, FIM cognitive0.549, SF-36V; physical function 0.667, role 0.750, emotion 0.583, social functioning 0.655, mental health 0.601, vitality 0.593, general health 0.531, physical component summary 0.539, mental component summary 0.618, SIS PHYSICAL with FIm motor 0.773, SF-36V physical 0.768, role 0.750, vitality 0.529, generalhealth 0.576, physical component summary 0.687. DISCRIMINANT VALIDITY Kruskal-Wallistest SIS -16 H 19.17 p 0.0003, SIS-PHYSICAL H18.39 p 0.0004, SIS-ADL H 18.79 p 0.0003 compared to FIM motor H 17.83. SIS discriminated 3 pairs of disability levels	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
FIM	McNett, Amato, Gianakis, Grimm, Philippbar, Belle & Moran 2014	"examine correlations between 24 hour and 72hour FOUR (Full Outline of Unresponsiveness) and GCSand functional and cognitive outcomes - assuming the truecorrelation between FOUR and FIM to be at least 0.75"	1. EXCELLENT 2 GOOD 3. n=33Fair 4. FAIR 5 GOOD 6 GOOD 7EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 EXCELLENT	fair	total FOUR and GCS scores at 72 hrs significantly correlated (0.34 p = 0.05, AUC 0.640 (0.439 - 0.841; 0.39 p = 0.02), AUC 0.688 (0.496 - 0.879) while 24hr were not (0.27 p = 0.01, AUC 0.625 (0.425 - 0.826) 0.28 p = 0.12, AUC 0.602 (0.392 - 0.812) low correlation.	?
FIM	Perrin, Niemeier, Mougeot et al 2015	"…that duration of PTA wouldbe the strongest predictor of patient functional status at discharge"	1.excellent 2 excellent 3 excellent sample size 4 good 5 excellent 6 excellent 7 excellent 8 good - most had evidence 9 excellent 10excellent	good	hierarchical multiple regression showed PTA was a significant predictor of dc FIM ß = -0.46, P = .001) while TTC (time to follow commands)and GCS (Glascow coma scale) were not. Correlations: PTA length and FIM -0.29, GCS 0.17, TTC -0.07.	plus
FIM	Semrau, Herter, Scott, Dukelow2015	examine the relationship of robotic measures to existing clinical measures across thefirst 6 months post stroke.	1. Excellent 2. Excellent 3. good sample size, 4 fair - unclear whatwas expected 5 good 6 good 7 excellent 8 fair 9 reference to butno properties stated) 10 excellent 11 excellent	fair - excellent	robotics assesses; position sense, kinaesthesia,motor function and simultaneous bilateral motor function (arms). Assessments were across the following time points post stroke T1 1 week, T2 6weeks, T3 12 weeks T4 26 weeks. Majority had mild stroke(76), 35 had moderate and 2 severe. Week 1 FIM did not predict early, late or incomplete recovery. Correlations between 4 robotic measures across 4 timepoints -0.41- - 0.61.	?
FIM	Tyryshkin, Coderre, Glasgow, Herter, Bagg, Dukelow & Scott 2014	robotics can provide a novel approach for quantifying sensory, motor and cognitive impairments associated with neurological disorders	1. GOOD 2. n/a 3. EXCELLENT 4 FAIR 5 GOOD 6 GOOD 7 FAIR8 FAIR 9 GOOD 10 EXCELLENT	fair	Majority of robotics parameters significantly correlated with motor and total FIM. Correlationswere low to moderate -0.07 to 0.62.	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
FIM	Wu, Burgard &Radel 2014	"we hypothesised differencesin levels of independence in ADL's and primary motor impairment of the more affected upper extremity between patients with and without apraxia, we also expected the amount of change on outcomes to be different"	1. EXCELLENT 2 N/A 3 sample size 15 poor 4. GOOD 5 GOOD 6GOOD 7 n/a 8 n/a 9 POOR hypothesis des not match method10 POOR no sensitivity analysis quality of stats low	poor	significant between group differences in admission FIM with and without apraxia (U = 6, z = -2.329, P = .019) Changes in FIM with and without apraxia were similar. Mean FIM DC scores for people with apraxia was similar to mean admission FIM scores without apraxia (U = 18.5, z = -0.552, P = .581).	plus
FIM	Rabadi & Rabadi 2006	testing interetest correlationand correlation with UE selfcare function as per FIM	1. good 2 fair 3 excellent 4. fair 5good 6 good 7 excellent 8 excellent 9 fair (treating therapistdoing FMA ? Bias) 10 excellent	fair - excellent	Admission FIM correlations with ARAT, FMA motor score 0.33 P<.001, 0.54, P<.001 and discharge scores 0.21 P<.01, 0.29 P<.01 stronger correlations at admission than discharge.	?
GAS	Barden, Baguley, Nott, Chapparo 2014 (a)	nil specific - aimed to evaluate UL performance changes in adults with UMN syndrome who received BTX-A injections for UL muscle spasticity in an outpatient clinical setting by (1) evaluating change in UL performance following BTX-A injections as measured by DCD and current clinical measures. (2) mapping observed changes to the Body Function and Structure and Activity domains of the ICF	1. Not stated GOOD 2. N/A 3. POOR sample size 4. FAIR %. GOOD 6 GOOD 7 EXCELLENT 8. FAIR only dynamometry information on measurement properties 9 EXCELLENT Stated raters not blinded but this is true for what occurs in clinical setting so did not reduce rating 10 EXCELLENT	fair - sample not included	There was no statistically significant relationship between GAS and DCD - dynamometry - nil data provided	

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
GAS	Barden, Baguley, Nott, Chapparo 2014 (b)	a measure of fine motor function, particularly in terms of lateral and pincer grips, will show greater relevance to achieving independence in everyday tasks OSMIN UPDATE hypotheses #1 GAS and pinch DCD	1. GOOD 2. n/a 3. poor sample size 28 4. fair - possible to deduce 5 GOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 EXCELENT COSMIN UPDATE 1. very good 2 very good 3 very good	fair	GAS and pinch DCD no significant relationship. 18 active and 4 passive goals were achieved from a total of 34 active and 8 passive.	0
GAS	Malec2001	positive changes would occuron each of the outcome measures at a level consistent with outcomes of other post-acute rehabilitation programs	1. excellent 2 good 3 good 4 excellent 5 good 6 good 7 excellent 8 excellent excluding ILS 9 excellent 10 excellent	excellent	spearman correlations of GAS T-Score at discharge with the Independent Living Status (ILS) 0.35 p<0.001, with 1 year follow up ILS 0.30 p<0.01, with discharge Vocational Independence Scale VIS 0.26 p<0.01, with 1 year follow up VIS 0.28 p<0.01 with discharge MPAI-22 Mayo Portland Adaptability Inventory - 0.55 p<0.0001, LOW correlations with all. 5. Low correlations with outcome measures at discharge and 1 year follow up r = 0.26 - -0.55)for patents graduating comprehensive day treatment program	?
Global Ax Scale	Brashear, Gordon, Elovic, Kassicieh, Marciniak,Do, Lee, Jenkins, Turkel, 2002	nil, conducted a trial to assess the effects of one setif injections with botulinum toxin A on measures of disability with respect to self care, limb position, and painas well as on muscle tone.	1. yes 4 in control group EXCELLENT 2. FAIR nil information given 3. GOOD sample size (appraised on individual sub group 64 and 58) 4. FAIR as per DAS reasoning 5GOOD 6 GOOD 7 GOOD 8 EXCELLENT 9 EXCELLENT 10EXCELLENT	fair - excellent	6 week (identified as primary end point) Globalax scale (physician rated) with DAS r = -0.46, P<0.001, patients or caregivers global assessment with DAS r = -0.51, P<0.001.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
LASIS	Ashford, Turner-Stokes, Siegert & Slade 2013	nil specific hypotheses	1. excellent 2 good 3 good sample size 4 fair no hypothesesbut able to determine what was expected 5 good 6 good 7 excellent 8 excellent 9 fair - no info provided re raters, blinding minimal detail 10 excellent	fair - excellent	LASIS passive ArmA passive Rho 0.5; p = 0.01(convergent validity) LASIS active ArmA active Rho Rho 0.48; p = 0.01 (convergent validity) LASIS passive ArmA active Rho 0.23; p = 0.078(divergent validity) LASIS active ArmA passive Rho 0.02; p = 0.9 (divergent validity)	?
LASIS	Barden, Baguley, Nott, Chapparo 2014 (a)	nil specific - aimed to evaluate UL performance changes in adults with UMN syndrome who received BTX-A injections for UL muscle spasticity in an outpatient clinical setting by (1) evaluating change in UL performance following BTX-A injections as measured by DCD and current clinical measures. (2) mapping observed changes to the Body Function and Structure and Activity domains of the ICF	1. Not stated GOOD 2. N/A 3. POOR sample size 4.FAIR %. GOOD 6 GOOD 7 EXCELLENT 8. FAIR only dynamometry information on measurement properties 9EXCELLENT Stated raters not blinded but this is true for what occurs in clinical setting so did not reduce rating 10 EXCELLENT	fair - sample not included	DCD components and carer burden scale - 0.01 - 0.62, Patient disability scale 0.08 - 0.48. (Spearman Rank order correlation). 0.48 - 0.62, p<0.05 correlation with relaxation and contraction duration components	

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
LASIS	Turner-Stokes, Baguley, De Graaff, Katrak, Davies, McCrory & Hughes 2010	nil but research questions - does GAS provide added value as a responsive indicator over other measures, (2) how does it relate to other measures	1. EXCELLENT n= 6, 2. GOOD not described but can de deuced were omitted 3. GOOD sample size, 4 FAIR 5 GOOD 6 GOOD 7 FAIR 8 FAIR 9 EXCELLENT 10 EXCELLENT	FAIR	no significant associations between GAS and PDS, CBS (LASIS)	?
MAL	Atler, Malcolm &Griefe 2015	nil specific but aimed to question what is the relationship among upperextremity motor function, activity and participation following CIMT. COSMIN UPDATE 2 hypotheses generic #1 - participation measure #2 impairment measure	1. EXCELLENT 2 NA 3 POOR sample 12 4 FAIR through references can see that measures currently being used are not indicating participation outcomes so not expecting correlations 5 GOOD 6 GOOD 7 GOOD older adults no stated if stroke related 9. FAIR difficult to locate all correlation results clearly 10 EXCELLENT COSMIN UPDATE 1. very good 2 very good 3 adequate	poor - excellent (Fair with sample excl)	MAL - How well score correlated with % of timespent with others (a participation indicator) R^2 =0.52, P<0.05 otherwise minimal correlations between activity and participation measures.	? COSMIN UPDATE impairment / motor function measures participation (+), not finished as decision to excludebased on part of Ix study
MAL	Borstad & Nichols-Larsen 2016	nil specific stated " objective was to determine the feasibility of administering theBKT (brief kinaesthesia test) and begin to validate with persons with mild to moderatepost stroke hemiparesis examine the relationship between the BKT scores and other valid sensory and motor measures).	1. EXCELLENT 2. N/A 3. n=12 POOR 4. FAIR identified constructs f each measure to validate new measure expectingcorrelation with construct unclearwhat was expected 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT nil other flaws identified 10 EXCELLENT	fair - excellent	brief kinaesthesia test correlated with MAL howwell r=0.76, p=0.007, How much r = 0.84, p=0.001.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
MAL	Celik, O'Malley, Boake, Levin, Yozbatiran & Reistetter 2010	to report correlations of fourrobotic measures to widely used clinical measures	1.not stated but can be deduced GOOD 2. GOOD 3. POOR n=9, 4 POOR 5. GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT only included mildly impaired participants EXCELLENT 10 EXCELLENT	Poor - Excellent	MAL had weak to moderate correlations with trajectory error -0.49, smoothness of movement 0.57, hits per minute 0.46 and mean tangential speed 0.21 as measured via robotics (all are measures for a target hitting task that involved repetitive reaching movements). Functional use not captured through robotics based on reaching movement.	plus andminus
MAL	Mark, Woods, Mennemei er, Abbas & Taub 2006	they hypothesised that sustained attention to task, episodic memory and executive function cognitiveprocesses are central for treatment compliance and retention if training benefits.	1. EXCELLENT 2. FAIR 3. POOR 4. FAIR 5 GOOD 6 GOOD 7POOR 8 POOR 9 FAIR 10 EXCELLENT	poor - excellent	spearman correlation between UE MAL AOU & MMSE -0.01 p0.96, SART -0.17 p0.60, WMS LMI -0.15 p 0.71, WMS LM II -0.10 p 0.71, WMS VR I -0.44 p 0.11, WMS VR II -0.37 p 0.19, animals -0.14 p 0.66, Trails B -0.17 p 0.56. Moderate (0.3-0.5) correlation with WMS VR I and WMS VR II only (verbal and visual anterograde memory). MAL QOM with MMSE -0.14 p 0.59, SART 0.36 p0.26, WMS LM I -0.26p 0.32, WMS LM II -0.18 p 0.50, WMS VR I -0.37 p 0.17, WMS VR II -0.26 p 0.34, animals - 0.31 p 0.30, Trails B 0.22 p 0.45. moderate correlations only with SART, WMS VR I, animals(sustained attention, memory). Small sample size does not indicate large impact of cognition on CI therapy outcomes. Nil of the cognitive tests were significantly correlated with UE MAL improvement (r= -0.01 – 0.44).	minus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
MAL	Harris & Eng 2006	individuals with their dominanthand affected by stroke would experience less impairment, greater performance in ADL's and higher ratings of participation compared to those with their non dominanthand affected	1. GOOD 2 n/a 3 GOOD samplesize 4. EXCELLENT 5 EXCELLENT 6 GOOD 7 EXCELLENT 8 EXCEELENT 9 FAIR - volunteers, minimal generalizability???? 10 EXCELLENT	fair - excellent	MANOVA for the function model (including CAHAI, MAL, Reintegration to Normal Living Index) showed no interaction or main effect of dominance, it did impact on impairment but did not translate into better performance in ADL's. alldependent variables were significantly affected by severity (as measured by FMA). Summary - dominance did not have an affect or advantage on scores of function over those with non dominant hand affected in chronic stroke	0
mFrenchay Arm Test	Sunderland, Tinson, Bradley & Hewer 1989	not clearly stated; the study investigated the relationship between grip strength, spasticity and functional recovery to discover whetherin fact it (grip strength) may be a valuable marker of recovery in the typical stroke patient.	1. EXCELLENT yes 7 lost to follow-up 2. GOOD - not clear butcan see that n=31 were included in the analyses 3. FAIR sample size, 4. FAIR 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 FAIR - limited information regarding statistical analysis, nil info re raters ? blind unable to determine if any biases	Fair - Excellent	The Frenchay Arm Test correlated with percentage grip r: 0.86 at initial assessment. At the final assessment 6 months the correlation was r: 0.90. The initial assessment (completed atone month post stroke) wrongly classified 13% ofcases when predicting functional outcome at 6 months.	?
MI	Smith-Arena, Edelstein, Rabadi, 2006	nil specifically stated; attempt to prospectively identify which neurological impairment(s) during an acute rehab hospital admission would predict the likelihood of a successful driving evaluation at d/c for stroke patients	1. excellent n=6 incomplete date 2. fair 3 FAIR sample size 4 FAIR 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 EXCELLENT	poor - excellent	logistic regression analysis of variables known to influence in-clinic driving evaluation UL MI regression coefficient -0.005, P value 0.810 OR (odds ratio) 0.995, 95% CI 0.958-1.034. LL MI regression coefficient -0.093 P value 0.060, OR 0.911 95% CI 0.827-1.004.Those who passed the in-clinic driver evaluation had, at admission higher MI scores	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
Motricity Index	Kopp, Kunkel, Flor, Platz, Rose,Mauritz, Gresser, McCulloch & Taub 1997	nil stated - "concurrent validity was assessed by comparing scores on the first Arm Motor Ability test (AMAT) with the arm scale of the MI	1 GOOD not stated but assumed 0 2. not stated GOOD 3. Fair sample size 4. FAIR unclear what was expected 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 FAIR - high functioning patients limited raw MI data given 10 EXCELLENT	fair - excellent	spearman's correlations between the Motricity- Index Arm Score and the first administration of the AMAT were: performance time; r = 0.45 functional ability r = 0.61 quality of movement r = 0.60 (all significant beyond the 0.01 level). Article also mentioned AMAT may be more sensitive than MI arm scale the AMAT scores differed between the 2 groups where the MI arm scale did not.	?
Motricity Index	Sunderland, Tinson, Bradley & Hewer 1989	not clearly stated; the study investigated the relationship between grip strength, spasticity and functional recovery to discover whether in fact it (grip strength) may be a valuable marker of recovery in the typical stroke patient.	1. EXCELLENT yes 7 lost to follow-up 2. GOOD - 3. FAIR sample size, 4. FAIR 5 GOOD 6 GOOD 7 EXCELLENT 8 EXCELLENT 9 EXCELLENT 10 FAIR - limited information regarding statistical analysis, nil info re raters blind unable to determine if any biases	fair - excellent	the UL-MI has a high correlation with percentage grip r:0.87 at the initial assessment (highest out of frenchay arm test, motor club assessment & 9HPTand at 6 month follow-up r:0.83 (weakest except for 9HPT). For patients with spasticity increases in grip paralleled with increasing function as measured by MI and motor club assessment. The MI results at 1 month gave a perfect 100% prediction of the Frenchay Arm Test at 6 months.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
OHS	Dennis, Wellwood& Warlow1996	nil given bit "…we present further data concerning their concurrent validity (or perhaps convergent constructvalidity since there is no gold standard)"	1, excellent n=2 missing data 2. good 3 excellent sample size 4 fair 5 good 6 good 7 good 8 good 9 excellent 10 excellent	fair - excellent	dependency" question: "do you/they require helpfrom another person with everyday activities?" sensitivity, specificity, accuracy (%) in predictingwhether pts scored above or below the followingcut-offs; OHS: >2, 72, 98, 81, >3 98, 78, 85. 'recovery' question "Do you feel that you/they have made a complete recovery from your/their stroke?" sensitivity, specificity, accuracy (%) in predicting whether pts scored above or below the following cut-offs OHS: =0 50, 79, 79, < 2 52,86, 79. the sensitivity, specificity, accuracy of ptswhose post stroke score was equal or better than pre-stroke score OHS: 39, 85, 73. Recoveryquestion was no more accurate at discriminating between those who were the same or better on BI and OHS after their stroke and those who were worse than it was at discriminating betweenthose with good and poor functional outcome. OHS score of 1 , 1 person answered "no" to recovery 1 "yes" OHS score of 1 12 "no" 13" yes" OHS score 2 "17 "no" 7 "yes" . 'dependency' question all pts scoring 5 answered"yes" those scoring 4, 34 said "yes" 1 "no" scoreof 3 27 answered "no" 21 answered "yes". Score 1 25 "no" 0 "yes" score 1 2 "no" 0 "yes".	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
OHS	Gubitz, Reidy, Christian& Phillips2012	"we quantified the physical, cognitive and psychosocial aspects of fatigue using the FIS and identified risk factorsfor post stroke fatigue at oneyear	1. EXCELLENT 2 FAIR 3 EXCELLENT 4 POOR 5 GOOD 6GOOD 7 n/a 8 n/a 9 FAIR ? Whyused mRS instead of repeating OHS 10 EXCELLENT	poor - excellent	younger age was the only consistent significant predictor of fatigue pre-OHS did not predict; fatigue freq -0.002, duration 0.12, disability 0.15symptom rank 0.03 for same factors above discharge OHS was a poor predictor 0.04, -0.10, -0.01, 0.01. patients who had fatigue at 1 year post; a higher pre stroke disability (as measuredby OHS) predicted higher levels of cognitive (0.29 p=0.03) psychosocial (0.33, p=0.02), and physical (0.34 p=0.01) fatigue as measured by OHS. But to summarise the only consistent predictor of fatigue in terms of freq, duration, disability or rank as a symptom was younger agewhich predicted a higher frequency and durationof fatigue episodes at one year. In regards to psychosocial, cognitive and physical, younger age and pre stroke disability and younger age predicted higher levels of those fatigue types.	?
OHS	Herrman, Black, Lawrence, Szekely & Szalai 1998	nil specific - to prospectivelyfollow a large sample to determine the frequency, severity, and course of depressive symptoms, their clinical correlates and their effect on functional outcome	1. not stated GOOD 2 FAIR 3 EXCELLENT 4 POOR 5 GOOD 6GOOD 7 EXCELLENT 8 GOOD 9FAIR Blinded assessors 10 EXCELLENT	poor - excellent	product moment correlations between depression rating scales SDS (Zung self rating depression scale) and MADRS (Montgomery Asberg Depression Rating Scale) a 3 months and 1 year; 3 mo: OHS and 3mo SDS 0.41 (0.0001) and with 3mo MADRS 0.40 (0.0001). 1 year OHS with 3mo SDS 0.35 (0.0001) and with1 year SDS 0.36 (0.0001) and with 3mo MADRS 0.29 (0.0001) and with 1 year MADRS 0.29 (0.001). multiple regression analyses performedto determine predictors of depression, 3 monthsOHS was found to be a significant predictor andat 1 year. SUMMARY OHS (functional handicap) correlated with depression at 3 months (r=0.41 P<0.0001) and 1 year (r=0.35 P<0.0001)	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
OHS	Pittock, Meldrum, Dhuill, Hardiman& Moroney 2003	no hypothesis but aim to investigate the prognostic ability of the OPS within 48 hours of admission to hospital (OPS-1) in predicting outcome at 6 mo and 2 years in acute ischaemic stroke and compare it with the 2 week OPS (OPS-2)	1. excellent 2 good 3 66 at 2 years GOOD 4 fair 5 good 6 good7 excellent 8 excellent 9 fair - no talk of blinding of assessors random allocation 10 excellent	fair - excellent	non-parametric spearman correlations of OPS-1(completed at 48hrs post) and OPS-2 (completed at 2 weeks post) were used to predict OHS at 6 mo 0.59, 0.66 and 2 years 0.55, 0.59 (is this moderate correlation/predictiveability)	
OHS	SCOPE Collaborat ions 7 IST 2007	no specific; we aimed to validate in hyper acute strokepatients, a previously described six sample variable model	1. excellent 2 Good 3 excellent sample size 4 fair 5 good 6 good7 n/a 8 n/a 9 excellent 10 excellent	fair - excellent	Discrimination = ROC curve 0.82 for independent survival at 6 months (indicated by ascore on OHS < 3). Meaning model was good at distinguishing between those with and withoutthe outcome of interest. There were more observed good outcomes than predicted with themodel however - model may be good for stratifying pts in hyper acute stroke trials but probably not accurate enough for decision making in individual patients. Model = age at onset, living alone at time of stroke, independent pre stroke, normal Glasgow Coma Scale verbal score, ability to lift arms and ability to walk	?
RMA	Pittock, Meldrum, Dhuill, Hardiman& Moroney 2003	no hypothesis but aim to investigate the prognostic ability of the OPS within 48 hours of admission to hospital (OPS-1) in predicting outcome at 6 mo and 2 years in acute ischaemic stroke and compare it with the 2 week OPS (OPS-2)	1. excellent 2 good 3 66 at 2 years GOOD 4 fair 5 good 6 good7 excellent 8 excellent 9 fair - no talk of blinding of assessors random allocation 10 excellent	fair - excellent	non-parametric spearman correlations of OPS-1and OPS-2 were used to predict RMA at 6 mo - total-0.75, -0.74, arm -0.75, -0.74, leg -0.62, - 0.64 and gross-0.66, -0.67 and 2 years total -0.56, -0.61, arm -0.54, -0.59, leg -0.55, -0.59 and gross-0.44, -0.48 (is this moderate correlation/predictive ability)	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
RMA	Taylor, Ashburn & Ward 1994	no hypothesis - aim: to determine whether asymmetrical sitting posture (to the affected side) and motor function following hemispheric stroke are related to the side of lesion orthe presence of unilateral neglect COSMIN UPDATE generic hypotheses #5 RMA associated with neglect RMA gross, arm, leg 3 hypotheses	1. excellent 2. n/a 3. fair samplesize 4 fair hypotheses 5 good 6 good 7 excellent 8 excellent 9.10. Groups varied in numbers significantly 9 v 28 inadequate statistics only discussed significance COSMINUPDATE inadequate	fair - excellent	group B 8 with neglect 1 without all 9 leant to affected side had worse RMA particularly grosssection at 3 weeks and 6 compared to group a -5 with neglect 23 without and all able to achievemidline sitting. But arm and leg trunk were not significantly different. Was a relationship between unilateral neglect and gross motor section of RMA.	? COSMNIN UPDATE (1/3) -

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
RMA - UL	Meldrum, Pittock, Hardiman, Dhuill, O'Regan & Moroney 2004	"to investigate whether the OPS scored within 48 hrs ofadmission post ischaemic stroke was a predictor of upper limb function at 6 months and to evaluate function of the UL in an unselected group of stroke patients for a period if two years post stroke"	poor - excellent	1. good 2 good 3 excellent 4. poor no indicationof what was expected5 good 6 good 7 excellent8 excellent9 fair no info on raters, blinding etc 10 excellent	spearman rank correlations (r) OPS scored at 48hrs was most highly correlated with RAS at 6 mo and 2years (-0.73, -0.71). OPS and sensation were most significant predictors of RAS at 6 mo. RAS at 6 mo with OPS (48hrs) - 0.73, age -0.38, sensation -0.66, RAS 48hs 0.63,grip 48hrs 0.69. RAS at 2 years with OPS -0.71, age -0.43, sensation -0.53, RAS 48 hrs 0.63, grip 48 hrs 0.55. ordinal logistic regression analysis incorporated all variables (mentioned above) - when considered o their own were significantly associated with RAS at 6 mo and 2 yrs. when considering all variables, OPS and sensation (48hrs) were the most significant predictor of UL outcome at 6 mo with factors such as class of stroke and RAS at 48hrs no longer significant. incorporating al variables resulted in a model that gave a accurate 6mo prediction of death in 83% of cases, RAS of 0-4 in 55% of cases, RAS of 5-15 in 97% of cases. At 2 years most significant predictor of RAS was sensation at 48hrs. incorporating all variables gave accurate two yr prediction of death in 96% of cases, RAS 0-4 in 71% cases, RAS 5-15 in 96% cases.	?

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SF-36	Rudick, Miller, Hass et al 2007	secondary objective to measure the relationship between common measures of MS severity - EDSS, MSFC	1. EXCELLENT 2. n/a 3. EXCELLENT b sample 4 poor unclear what was expected 5 good 6 good 7 poor no info provided 8 poor no info 9 excellent 10 excellent	poor	Higher EDSS scores were associated with lowerPCS scores at study entry. Each 1 point increasein EDSS score was associated with a decrease in PCS score of 3.7 points ($p<0.001$; $r^2 = 0.24$) and a decrease in MCS score of 0.99 point ($p<0.001$; r2 = 0.02). Greater MSFC scores wereassociated with higher HRQoL, with each 1-SD unit improvement in MSFC score associated withan increase in PCS score of 4.4 points ($p<0.001$ r2 = 0.16) and an increase in mCS score of 2.2 points ($p<0.001$; r2 = 0.03). summary - SD-36 correlates with EDSS and MSFC.	?
SIS	Ali, Fulton, Quinn, Brady on behalf of VISTA 2013 SIS comparator	" we hypothesised that existing primary outcome measures (mRS, BI, NIHSS)may also reflect patient quality of life" COSMIN UPDATE USED SAME HYPOTHESES AS STUDY but x 2 (1 for pt responses, 1for proxy)	1. Excellent 2. GOOD 3 excellent4 fair 5 good 6 good 7 poor 8 fair9 excellent 10 excellent COSMINUPDATE 1. very good 2 adequate (minimal stated but reduced minimally die to availability in other research 3 very good	poor - excellent	at 3 mo patient responses to QoL had a strongerassociation with mRS (SIS n=2970 P<0.0001, r = -0.71, r2 = 0.52, EQ-5D weighted score n=2987 r=-0.7 r2=0.53) where as proxy responses had a stronger association with BI (SIS n=867 P<0.0001, r = 0.68, r2 = 0.48, EQ- 5D n=837, r=0.78, r2= 0.63) than with NIHSS. BI had more mismatches between good primaryoutcome and poor QoL (EQ-5D 11.3%, SIS 19.2%) than the mRS (EQ-5D 8.5% SIS 10%) but less than NIHSS (29%, 23.9%)and for poorprimary outcome and good QoL less mismatches than mRS(EQ-5D 3.1%), (SIS 4.1%) but more than NIHSS (0%)	plus
SIS	Ellis, Sukal, DeMott & Dewald 2008	kinematic variables from the Arm Coordination Training Device 3D (ACT) are relatedto existing standardised clinical assessments of arm movement following stroke	1. good- not specified assumed 02 n/a 3 n=11 poor 4 good 5 good6 good 7 excellent 8 excellent 9 minimal generalisability high functioning FAIR 10 excellent	fair	SIS domains 2-6 and 8-9 no significant correlation. ACT (Spearman rank Correlations)with SIS domain 1; 0% .014, 25% 0.62, 50% 0.31, 75% 0.56, 100% 0.37, 125% 0.36, 150% 0.44, 175% 0.47, 200% 0.17. SIS domain 7; 0% 0.13, 25% 0.62, 50% 0.51, 75% 0.67, 100% 0.61, 125% 0.46, 150%0.66, 175% 0.71, 200% 0.54. varying strengths of correlation. mild-moderate.	plus

Hypothesis Testing						
Outcome measure	**Author**	**Hypothesis**	**COSMIN**	**COSMIN result**	**Study statistical results**	**Terwee score**
SIS	O'Dell, Kim, Rivera, Fieo, Christos, Polistena, Fitzgerald& Gorga 2013	the AMAT-9 will hold a stronger relationship with the Hand Function rather than the Communication sub-score of the SIS (divergent validity)	1. EXCELLENT 2. not stated butcan be deduced FAIR 3. FAIR (n=32) 4. GOOD 5. EXCELLENT 6. GOOD 7. EXCELLENT 8. EXCELLENT 9.FAIR not representative sample, young highly motivated group who volunteered 10. EXCELLENT	fair - excellent	AMAT-9 and hand sub score of SIS 0.40 (p=0.25) and with communication sub score -0.16 (p= 0.39) spearman rank correlation coefficients.	plus
SIS	Boger, Hankins, Demain & Latter 2015	health status would be positively associated with selfmanagement skills, attitudes and behaviours hence higherscores within each domain ofthe SIS would correlate with higher scores on the Southampton stroke self- management questionnaire (SSSMQ)	1. EXCELLENT 87 from 95 werecomplete 2. EXCELLENT - incomplete were excluded 3. GOOD sample, 4 GOOD re no of hypotheses 5. EXCELLENT direction of correlation 6. GOOD 7 EXCELLENT 8 EXCELLENT 9EXCELLENT 10 EXCELLENT non parametric correlations	good - excellent	Non parametric correlations of SSSMQ (patientreported measure of self management) with SISdomains; strength 0.05, hand function 0.10, ADL/IADL 0.24, mobility 0.25, communication 0.38, emotion 0.59, memory and thinking 0.49, participation /role function 0.36, 0-100 perceivedrecovery 0.61. all positive as per hypothesis andranged from low to moderate strength correlations.	+
UL MAS	Horgan, Cunningham, Coakley, Walsh, O'Regan & Finn 2006	aim to establish the concurrent validity of the SASby comparing to MMAS and to evaluate time taken to complete each measure.	1. not stated assumed 0 GOOD 2n/a 3. FAIR 4 FAIR as reported to establish the concurrent validity of the SAS with the MMAS 5 GOOD 6 GOOD 7 GOOD 8 GOOD 9 FAIR 10 EXCELLENT	fair - excellent	nil floor or ceiling effect for UL MAS but ceiling for SAS (9.8%). SAS quicker to complete than UL MAS (2.8 v 10.4 mins F ratio 192.4, 1df, p<0.0001). Pearson correlation coefficient between mMAS and SAS 0.91. 95% limits of agreement (Bland Altman method) showed the SAS was likely to give a value between 88% and168% of that obtained by mMAS. the Bland Altman plot shows SAS proportional scores tended to be higher than mMAS particularly at higher end of scale.	?

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
10MWT	Salbach, Mayo, Higgins, Ahmed, Finch & Richards 2001	no	nil stated, purpose to identify the most responsive simple method for measuring gait speed in the first month after stroke and to compare responsiveness of gait speed to other measures. COSMIN UPDATE hypotheses x 5 10MWT will be more responsive than 5MWT, TUG,Berg Balance, BI, STRERAM as per ES and SRM.	1. EXCELLENT 2 EXCELLENT 3. GOOD 4. EXCELLENT 5. EXCELLENT 6. GOOD 7. GOOD 8. FAIR 9. GOOD 10. GOOD 11 EXCELLENT 12 GOOD 13 FAIR - nil significant language or cognitivedeficits initial assessment 8 days post and variations in how scales administered 14 FAIR COSMIN UPDATE 4. very good 5 very good6 doubtful 7 doubtful	FAIR - excellent doubtful	RESPONSIVENESS - 5mWT (comfortable pace) was mos responsive followed by 5mWT (maximum pace), 10mWT (comfortable pace) 10mWT (maximum pace). When compared to other measures the most responsive tests of gait speed were 5mWT (comfortable pace) then Berg balance Scale, then 5mWT (maximum pace). dueto the inability of patients to complete the TUG its responsiveness varied dependingon the strategy for substitution data. when looking at "slow" ,"moderate" or "fast" walkers 10mWT was never one of top 3 most responsive. effect sizes - 10mWT (max speed) effect size 0.55, 10mWT (comfortable speed) 0.74 5mWT (max speed) 0.66 5mWT (comfortable speed) 0.83. SRM: 10mWT max speed 0.83 comfortable speed 0.92 5mWT max 1.00, comfortable 1.22. SRM's - berg balance 1.04, Barthel index 0.99, STREAM 0.89TUG 0.73. MORE INFO RE RESPONSIVENESS MEASURE IN ARTICLE	0 COSIMIN UPDATE 10MWT with 5MWT -, TUG +, berg balance -, BI -, STREAM +, (2/5) -

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
10MWT	Scrivner, Schurr & Sherringt on 2014	no	nil specific stated - research q: how responsive are the 10MWT, Step Test, LL items of the MAS to change in performance during inpatient care after stroke. COSMIN UPDATE 1. 10mwt to have at least moderate responsiveness as per ES	1. EXCELLENT 2 EXCELLENT 3 EXCELLENT 4 EXCELLENT 5 EXCELLENT 6 EXCELLENT 7 FAIR nil evidence of global measure for example 8 Fair 9 GOOD 10 GOOD 11EXCELLENT 12 EXCELLENT 13 EXCELEENT 14 FAIR used Effect size, SRM and mES COSMIN UPDATE 1 n/a 2 n/a 3 doubtful 4 very good	Fair - excellent doubtful	ES 1.44 SRM 0.93 mES 0.45 51 or 26.8% of participants did not change on the 10MWT, 127 (66.8%) had a floor effect on admission and 49 (25.8%) floor effect on discharge nil effects for ceiling on admission or discharge.26.8% did not change their scores. In comparison to other included measures (Step Test MAS ll items) The MAS mobility summed score had highest consistent responsiveness (ES 1.42 SRM 0.71, mES 0.92) and the Step test had the largest proportion of pts who did not change, item 4 on MAS had largest ceiling effect on discharge.	? Hypotheses 1 met = (1/1)
9HP	Beebe & Lang 2009	no	to determine how responsive 6 clinical tests are to change over the first 6 months after stroke	1. excellent 2. fair 3. poor 4. excellent 5 excellent 6 good 7. excellent 8 fair 9 good 10 good 11 excellent 12 excellent 13 fair 14 fairas used effect size for responsiveness COSMIN UPDATE4. very good 5 adequate as onlydiscussed retest reliability 6. no hypotheses with use of ES DOUBTFUL 7 very good	Poor - excellent doubtful	responsiveness determined usingsingle population effect size method 1-3 months 0.52 (moderate responsiveness) 1-6 months 0.66 (high responsiveness)	? - no clear hypothesis COSMIN UPDATE 2/2 + both timeframes moderate responsive ness

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
9HPT	Jacob-Lloyd, Dunn, Brain & Lamb 2005	no	nil stated, aim and purpose only to investigate measurement properties. COSMIN UPDATE NHPT would have a moderate ES (Cohens criteria)	1. EXCELLENT 2. GOOD 3. GOOD 4. EXCELLENT 5. EXCELLENT 6. FAIR 7. FAIR 8. FAIR no hypotheses able to deduce 9. GOOD 10 GOOD 11FAIR 12 EXCELLENT 13 EXCELLENT 14 FAIR as effect size for responsiveness COSMIN UPDATE 4. n/a 5 n/a 6 doubtful 7 very good	poor-excellent doubtful	effect sizes and efficiency based on Wilcoxin test statistics were calculated to evaluate responsiveness. Effect sizes showed a large improvement detected by UL MI scores r: 0.49 compared to 9HPT r=0.65 The non parametric calculation of the relative efficiency of the 9HPT versus the UL MI was 1.34 indicating that the 9HPT was the more efficient measure of changein the subsample	? No hypotheses Sample < 50 COSMIN UPDATE hyp 1 +1/1 +
9HPT	Sunderland, Tinson, Bradley & Hewer 1989	no	not clearly stated; the study investigated the relationship between grip strength, spasticity and functional recovery to discover whether in fact it (grip strength) may be a valuable marker of recovery in the typical stroke patient.	1. EXCELLENT yes 7 lost to follow-up 2. GOOD - not clear but can see that n=31 were included in the analyses 3. FAIR sample size, 4. EXCELLENT 5. EXCELLENT stating of time interval 6. GOOD stated continued with physio but not clear 7. EXCELLENT 8. FAIR hypotheses vague 9. GOOD not stated 10 GOOD not stated 11. EXCELLENT 12 EXCELLENT 13 FAIR - not stated who completed assessments? Blinded nil info to determine presence of bias 14 FAIR 15 FAIR	Poor - excellent	responsiveness was rated by counting the number of cases where scores increased between adjoining sessions at initial- 1month: 12 1-3mnths: 16 3-6 months: 14. most increase of adjoining initial - 1 month: MI n=22, 1-3mnths MI n=22, 3-6 mnths grip strength n=19, least numbers initial - 1 month FAT and9HPT n=12, 1-3 mnths FAT n=11, 3-6 mnths motor club n=5 (FAT n=6).	correlations not performed ?

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
ARAT	Beebe & Lang 2009	no	to determine how responsive 6 clinical tests are to change over the first 6 months after stroke COSMIN UPDATE hypotheses ARAT will have 1)mod responsiveness un 1-3 mo and 2) mod in 1-6 mo	1. excellent 2. fair 3. poor 4. excellent 5 excellent 6 good 7. excellent 8 fair 9 good 10 good 11 excellent 12 excellent 13 fair 14 fairas used effect size for responsiveness COSMIN UPDATE 4. very good 5 adequate as only discussed retest reliability 6. no hypotheses with use of ES DOUBTFUL 7 very good	excellent - poor sample excluded fair COSMIN UPDATE doubtful	responsiveness calculated using singe population effect size method values closer to 1.00 = more responsive to change ARATat 1-3 mo: 0.55 (moderate responsiveness) 1-6 months 0.63.(high responsiveness)	COSMIN UPDATE 2/2 + both timeframes moderate responsiveness

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
ARAT	de Weerdt& Harrison 1985	no	"…study quantifies and compares the arm-hand recovery, recorded at two points in time via the administration of the B-FM and the ARAT to a group of stroke patients at 2 and 8 weeks post onset of stroke"	percentage of missing items not described but pre and post n are equal (EXCELLENT), no missing items n/a, sample size (GOOD), longitudinal design (EXCELLENT), time interval described (EXCELLENT), not described but assumable (GOOD), part of patients changed (EXCELLENT) hypothesis not formulated but possible to deduce (FAIR), direction of correlations or mean differences of change scores NOT stated (GOOD), absolute or relative magnitude NOT stated (GOOD) adequate description of comparator instrument (EXCELLENT), adequate description of psychometric properties of comparator instruments (EXCELLENT), researcher not blinded (FAIR) statsUPDATED COSMIN 10 b 4. Very good 5. very good 6 inadequate asused significance tests wilcoxin matched pairs signed ranks test INADEQUATE	excellent - fair inadequate	6 week difference in scores usingWilcoxin matched-pairs signed ranks test (ARAT T=1, n=28 z=4.60 B-FM T=36, n=38 z=4.85 acute stroke	unsure how to rate -

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
ARAT	Lang, Wagner, Dromerick& Edwards 2006	no	not stated, aim to examine responsiveness of ARAT in a population of hemiparetic acute stroke patients. COSMIN UPDATE hypotheses that should have an ES of at least moderate in both 0-14 and 0-90 days (x2)	percentage of missing items was given (EXCELLENT), not clear howmissing items were handled (FAIR), sample size (GOOD), longitudinal design (EXCELLENT), time interval described (EXCELLENT), assume what happened as only says post treatment (GOOD), portion of patients changed FAIR, hypothesisnot formulated but possible to deduce (FAIR), direction of correlations not included in hypothesis GOOD, magnitude not included in hypothesis GOOD, adequate description of constructs measured by comparator instruments (EXCELLENT), adequate description of psychometric properties of comparator instruments (EXCELLENT), no other flaws (EXCELLENT) stats ES FAIR. CPOSMIN UPDATE 4 n/a 5 n/a 6 doubtful 7 very good	FAIR - excellent - doubtful	responsiveness was determined by the single population effect sizemethod (the effect size within the first weeks of stroke was calculated as the mean change from day 0 to day 14 divided by the SD at day 0). Same process for months post. Secondly responsiveness was determined by using the responsiveness ratio method. Findings: Responsiveness ratio ARAT total score Day 0 - 14 5.200, day 0 – 90 7.067. single population effect size ARAT total (day0-14) 1.018 day 0-90 1.390ARAT gross sub score day 0-14 0.729 day 0-90 0.984 ARAT grasp sub score day0-14 1.042 day 0-90 1.224 ARATgrip day 0-14 1.017 day 0-90 1.324 ARAT pinch day 0-14 0.854day 0-90 4.494. Large effect sizes usually around 0.8. acute stroke	plus (based on responsiveness ratio >0.70) (+ 2/2) moderate and much higher ES
ARAT	Rabadi & Rabadi 2006	no	assess responsiveness to change over time COSMIN UPDATE hypotheses - correlations with changes in instruments measuring similarconstruct (> 0.5)	1 good 2 fair 3 excellent 4 excellent 5 good 6 fair 7 good 8 fair 9 good 10 good 11excellent 12 excellent 13 fair 14 fair (SRM article defines that there is no consensus regarding how to best measure responsiveness COSMIN UPDATE4 very good 5 very good 6 adequate 7 adequate	fair - excellent adequate	SRM for ARAT 0.68 , for FMA motor score 0.74 , change in ARAT score correlated well with the change in FMA (R^2 = 0.56 P<0.001) acute stroke	? COSMIN UPDATE hypoth #1 =0.56 +

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
ArmA	Ashford, Turner-Stokes, Siegert & Slade 2013	no	"it was expected that ArmA would identify a significant difference between the responder and the non-responder groups for passive function as defined by the primary goal outcome (GAS)"	8 very good - discuss as responderor non responder re: GAS outcome9 adequate as discuss significanceonly not strength of correlations (quadratic weighted kappa)	adequate	significant difference between responder and non responder groups for passive function subscale at 8 weeks (U = 98.5; p=0.01) - more responsive than LASIS passive items, LASIS active items, DASH active items or Barthel Index - none of which demonstrated any difference between the two groups. A significant difference was not shown for the active function subscale (U=163.4; p=0.35) as expected due to 4 of 58 having goals related ot active function.	plus (2/2)
ArmA	Ashford, Slade, Nair & Turner-Stokes 2014	no	nil hypothesis - the aim wasto assess the ability of the ArmA to detect changes in clinical presentations. Hypotheses would be that #1 Detect change in passive function post spasticity intervention (compared to LASIS passive) #2 active subscale with LASIS active, changes in MAS indicate true change occurred	construct approach 4 very good 5 inadequate 6 very good 7 very good	inadequate	ArmA detected change at baselineto 8 weeks and was maintained at16 weeks, appeared more sensitive than LASIS. No change was identified on the active subscale nor the LASIS active	plus 2/2
BI	Filiatrault et al 1991	no	Nil given COSMIN update at least mod responsiveness	COSMIN update 4 very good 5 doubtful nil info on Functional test 6 inadequate 8 very good	inadequate	BI was equally as sensitive to change as the Fugl-Meyer but more sensitive than the FunctionalTest W = 0.39, (FM W = 0.41, Functional Test W = 0.21) W = Kendall's coefficient of concordance max value is 1	? Due to stats given

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
CMA	Gowland, Stratford, Ward, Moreland, Torresin, Van Hullenar, Sanford, Barreca, Vanspall &Plews 1993	nil	tested the change that took place between admission anddc, hypothesising that it wouldbe both statistically significantand greater than the change noted in the FIM COSMIN UPDATE CMSA disability more change than FIM	1. EXCELLENT 2 GOOD 3 poor 4 excellent 5 excellent 3-25 weeks admission 6 good assumable as were admitted to rehab 7 good no evidence provided but assumable changed as other measures demonstrated change also 8 excellent 9 excellent as expected direction greater than FIM 10 good 11 poor description FAIR 12 FAIR 13 excellent 14 excellent COSMIN UPDATE 1very good 2 adequate 3 adequate vague hypotheses very good	poor - excellent adequate	CMA disability inventory was considerably more responsive than FIM in subacute stroke, relative efficiency was 1.92 timesgreater. CMA disability inventoryvariance ratio 0.53, FIM 0.39, CMA disability inventory F 37.25 FIM 19.40.	plus (1/1)

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
EQ-5D	Peters 'Change inhealth status' (2014)	no	The specific aim of this article is to report the evidence of whether change in health status occurs after a one yearperiod in a primary care sample of people with Long term conditions. The assumption that change can be assessed was based on 2 considerations: first...'the possibility of changes over time in the quality of life in LTCs and second, the trajectory of many LTCs is a deterioration in health status. COSMIN UPDATE hypotheses assigned (as nil identified in study) 1) EQ-5D index and VAS would be responsive to change over one year period post stroke 2)EQ5D would be as responsive as disease specific measure SIS v3 in same period	COSMIN UPDATE 1. very good 2 adequate 3 inadequate not appropriate stats - statistical significance, no hypotheses 4 inadequate - missing data authors discuss in 'discussion' that data was not imputed as method not reported by measure developers and two was a purpose of study to look at return rates 0 but this was not mentioned in aims/methods	inadequate	0/2 - neitheEQ-5D and SIS was responsive (-)	

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
EQ-5D	Pickard, Johnson &Feeney, 2005	no	We hypothesized the following: 1) that the change scores of the preference based generic HRQOL measures would have moderate (0.35-0.50) to strong (>0.50) correlations with each other; 2) and based on results from a previous study, that HUI3 and EQ5D index would demonstrate larger change scores and consequently generate more QALYs, in patients categorizes as 'large improvement' than the other preference-based measures.' COSMIN UPDATE study hypotheses 1) that the change scores of the preference based generic HRQOL measures (EQ5D VAS with , EQ5D index score,SF-6D, HU12, HU13 AND EQ5D index with EQ5D VAS,SF-6D, HU12, HU13) would have moderate (0.35-0.50) tostrong (>0.50) correlations with each other; 2) EQ-5D index larger change scores than other measures 3) reviewer hypotheses - correlations with ADL (BI) #1,MRS (disability) #1	all 'very good' excluding evidence of psychometric properties of comparators (Lisa comment information about the measurement properties of only one (out of 7) of the comparator instruments was provided. (CES-D). ADEQUATE	adequate	EQ5D-Index moderate correlations with the change scores of SF-6D, HUI2 and MRS Strong correlation with HUI3 and Barthel index. Low correlations with EQ-VAS. VAS exhibited moderate correlations with changescores of SF-6D, HUI2, HUI3 and CESD (considered more EQ-VAS scores more strongly related to change in mental functioning than measures of disability whilst EQ5D Index more strongly correlated with measures of disability (MRS) and ADL's (BI). EQ5D index large change scores,EQ5D VAS less	EQ5D index with EQ5D VAS0 r = 0.31 (+), SF-6D r = 0.45 (+), HU12 r = 0.48 (+) HU13 r = 0.59 (+) with BI r = 0.57 (+) with MRS r = -0.36 (+). EQ VAS with EQ5D Index r = 0.31 (+), SF-6D r = 0.50 (+), HU12 r = 0.38 (+) HU13 r = 0.37 (+) BIr = 0.22 (-) MRS r = -0.18 (-) change scores EQ5D index (+) 11/13 overall +

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
FIM	Rabadi & Vincent 2013	no	nil "we set out to compare the EDSS to the FIM scale as an effective measure of MS-related disability and to determine which of the two scales is more responsive to clinical change following treatment. COSMIN UPDATE hypotheses FIM more responsive than EDSS hyp #2	1. good 2. N/a 3 good sample size 4 excellent 5. excellent 6. excellent fair - unclear if patients changed poor 9 good 10 good 11 excellent 12 excellent 13 not a representative sample, pts questioned re onset info several years prior 14 fair not optimal - SRM. COSMIN UPDATE 4 very good 5 very good 6 doubtful 7 doubtful	poor doubtful	FIM SRM 0.53, EDSS SRM 0.15	plus (1/1)
FIM	Houlden, Edwards, McNeil, Greenwo o d 2006	no	to compare BI and FIM - COSMIN UPDATE FIM to be at least Moderately responsive (Cohens ES) mod FIM total FIMm FIMc	responsiveness fair **Use of effect sizes**	doubtful	FIM total ES (0.52-0.0.72). ES FIM cognitive score (0.35-0.43)	2/3 (-)
FIM	Sharrack, Hughes, Soudain & Dunn 1999	no	COSMIN UPDATE FIM at least mod ES	doubtful - use of effect size		FIM total score (ES= 0.46, $p < 0.001$). N only 25	minus (0/1)
FIM	van der putten	no	COSMIN update hypotheses -generic 1. mod responsiveness for MS and stroke 2 hypotheses (at least ES 0.5)	Responsiveness was fair given the stats that were used. The old version of COSMIN did look favourably of ES but rather than mark poor I raised it to fair given that this is a stat that is commonly used within our field.	doubtful	ES FIM=0.30 in MS patients and 0.82 in stroke patients	plus stroke /minus MS 1/2 (-) 50%

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
GAS	Doig, Fleming, Kuipers & Cornwell 2010	no	1. "determining their relative sensitivity in measuring change" used paired t tests tolook at statistically significant change between pre and postIx scores	1. yes 2 good 3 sample re people or goals n=13 people as all goals converted to 1 GAS T-score per participant POOR 4. EXCELLENT 5 EXCELLENT 12 weeks 6. GOOD7 GOOD 8 POOR 9 GOOD 10 GOOD 11 EXCELLENT 12 EXCELLENT 13 EXCELLENT 14 POOR t tests and significance onlyreported COSMIN UPDATE inadequate as stats method used and no hypotheses	poor - excellent inadequate	using paired t tests: GAS T-scorePre Ix M(SD) 36.9 (6.3) post Ix 52.8 (6.2) t -9.65 p<0.01. COPM performance client pre 5.0 (1.9) post 8.0 (1.5) t -5.07 p<0.01 COPM satisfaction pre 4.6 (1.8) post 8.0 (1.5) t -6.40 p<0.01 performance significant other pre4.3 (1.5) 7.6 (1.2) t -7.92 p<0.01. Statistically significant improvements across GAS and COPM.	? Plus(1/1)
GAS	Khan, Pallant & Turner-Stokes 2008	nil	nil specific but objectives (1) explore the type and nature ofperson-centred goals that arecommonly set and achieved during the program (2) compare the responsiveness and relative efficiency of GASwith the FIM and BI as outcome measures for rehab in MS -- so would include as hypotheses GAS (A) to be more responsive than FIM (B)and BI (C) as per Cohens andat least moderate so 2 hypotheses	1. GOOD not described 2 N/A 3. POOR sample size n 24) 4 EXCELLENT 5 EXCELLENT 6. EXCELLENT 7 GOOD 8 POOR 9 GOOD 10 GOOD 11 EXCELLENT 12 EXCELLENT 13 EXCELLENT 14 FAIR SRM and effect size used COSMIN UPDATE 4 very good 5 very good 6 doubtful 7 very good	poor - excellent doubtful	GAS: Effect size 9.0, SRM 2.4, Relative efficiency (t value) 10.0 relative efficiency (z value) 1.4. FIM ES 0.4 SRM 1.0 relative efficiency (t value) 1.1 relative efficiency (z value) 1.1 BI: Effect size 0.4, SRM 0.8, Relative efficiency (t value) 1.0 relative efficiency (z value) 1.0. GAS mostresponsive then FIm then BI in MS.	?

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
GAS	Lannin 2003	no	"The aim of this study was to determine the clinical utility ofGAS in measuring the outcomes of a home-based occupational therapy rehabilitation service for adults"	1. excellent 2 n/a 3. fair n=36 goals 4. excellent 5. yes 5. poor 6. good 7 good 8 fair 9 good 10 good 11 n/a 12 n/a 13 excellent 14 ??fair COSMIN UPDATE 10d before and after intervention **doubtful** for both description of intervention and stats, in addition to limited methodological detail provided	poor - excellent	53% of pts achieved expected level of performance on discharge , 33% achieved above expected levels, 14% not attaining expected levels of performance on dc. Average change was greater than 1.5 standard deviations (mode of 1 SD). Program as a whole results - mean GAS at admission 36.4 (4.87) mode -1, dc 52.5 (8.74) mode outcome 0 - results closelylinked to theoretical expected results mean of 50 SD of 10 .	? Plus (1/1) as per results in accordanc e with theory
BI(C&W)	Houlden, Edwards, McNeil, Greenwoo d 2006	no	to compare BI and FIM - COSMIN UPDATE mBI to be at least Moderately responsive (Cohens ES)	responsiveness fair **Use of effect sizes**	doubtful	ES BI (C&W) =0.65 for all conditions, 0.79 for infarct, 0.52 intracerebral haemorrhage, 0.64 Subarachnoid haemorrhage and 0.55 for Traumatic brain injury. Correlation between BI (C&W) and FIM change scores r2=0.733	plus
BI(C&W)	van der putten	no	COSMIN update hypotheses -generic 1. mod responsiveness for MS and stroke 2 hypotheses (at least ES 0.5)	Responsiveness was fair given the stats that were used. The old version of COSMIN did look favourably of ES but rather than mark poor I raised it to fair given that this is a stat that is commonly used within our field.	doubtful	ES BI(C&W) =0.37 and FIM=0.30 in MS patients and ES BI (C&W) =0.95 and FIM=0.82 in stroke patients	plus stroke /minus MS 1/2 (-) 50%

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
MI	Jacob-Lloyd, Dunn, Brain & Lamb 2005	no	nil stated, aim and purpose only to investigate measurement properties. COSMIN UPDATE MI would have a moderate ES (Cohens criteria)	1. EXCELLENT 2. GOOD 3. GOOD 4. EXCELLENT 5. EXCELLENT 6. FAIR 7. FAIR 8. FAIR no hypotheses able to deduce 9. GOOD 10 GOOD 11FAIR 12 EXCELLENT 13 EXCELLENT 14 FAIR as effect size for responsiveness COSMIN UPDATE 4. n/a 5 n/a 6 doubtful 7 very good	poor-excellent doubtful	effect sizes and efficiency based on Wilcoxin test statistics were calculated to evaluate responsiveness. Effect sizes showed a large improvement detected by UL MI scores r: 0.49 compared to 9HPT r=0.65 The non parametric calculation of the relative efficiency of the 9HPT versus the UL MI was 1.34 indicating that the 9HPT was the more efficient measure of changein the subsample	? Less than 50 sample COSMIN UPDATE hyp 1 - as <0.50
SF-36	Freeman, Hobart, Langdon & Thompson2000	no	nil given COSMIN UPDATE hypothesis moderate responsiveness	fair COSMIN UPDATE 1. very good 2 doubtful as ES with no hypothesis	doubtful	Effect sizes for the SF-36 dimensions ranged from negligibleto small (effect sizes 0.01–0.30). The dimensions demonstrating the largest effect size were the emotional role limitations (effect size 0.27) and pain (effect size 0.30).	minus (0/1)
SF-36	Hagen, Bugge & Alexander 2003	no	Hypothesised that since the health of stroke survivors will tend to improve over time, their SF-36 scores would improve also, #1 with most improvement taking place over the first few months following onset.#2 still responsive 3-6 but less than first 1-3 In addition, any changes in sf-36 scores would conceivably be positively related to changes in other health measures.	1. very good 2 very good 3. doubtful	poor	statistically significant improvements were seen between1 and 3 months except bodily pain, general health and Mental health. Significant improvements were also seen in mBI, CNS and MMSE.NO significant changes between 3 and 6 mo however CNS and MMSE did. (mBI assumed not as no info given). 1- 3 SRM 0.1175 - 0.3879, 3-6 - 0.0136 - -0.1457 summary not responsive 3-6 mo post. Low responsiveness first 3 mo	hypotheses #1 - (0/1) #2 - (0/1) overall -(0/2)

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
SF-36	Madden etal 2006	no	nil given was a secondary aimof the paper COSMIN UPDATE hypotheses SF-36 would be moderately responsive	poor COSMIN UPDATE 1. very good 2 very good 3 inadequate stats, statistical significance only 4	inadequate	Re: correlations between changesin FIM and SF-36 Subscales 'No correlation co-efficient attained thea priori level of 0.25.' Similar constructs did not correlate with similar constructs e.g. FIM motor summary with SF-36 Physical function correlation0.183 SHANS NOTE this appears to be the results related ot hypotheses testing for RESPONSIVENESS: all mean change scores were significant for the fIM, for the SF- 36 4 domain and the PCS were statistically significant. nether of the summary scores were associated with changes in any of the FIM outcomes. no relationship. all <0.30	minus as only half of SF-26 showed statistical significant change - (0/1)
Excluded at COSMIN update							
ARAT	Barden, Baguley, Nott, Chapparo 2014	nil	nil specific - aimed to evaluate change in UL performance following BTX-A as measured by clinical measures and dynamometry	1. GOOD 2 N?A 3. POOR 4 EXCELLENT 5. EXCELLENT 6 FAIR unclear what occurred in interim period 7 FAIR nil info found re this 8 FAIR 9 GODD 10 GOOD 11 EXCELLENT 12 FAIR only dynamometry info found 13 EXCELLENT 14 FAIR ES only	poor-excellent sample excluded fair	"Low" (≤3) "high" (4-57) neither group demonstrated change in ULperformance. Pre and post median scores for low grp remained at 0 the high grp achieved non significant 4 - point median improvement	?

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
ARAT	Barreca, Stratford, Lambert, Masters & Streiner 2005	no	1. the CAHAI change scores should correlate more highly with the CMSA and ARAT change scores than with the CMSA shoulder pain score 2.the CAHAI is more adept than the CMSA and ARAT at distinguishing change in patients with acute, mild/moderate impairment presentations from patients with chronic, sever impairment presentations	1. no but can be deduced as 0 GOOD 2 GOOD 3 FAIR n=39 4 EXCELLENT 5 EXCELLENT 6 Excellent 7 GOOD 8 EXCELLENT 9 EXCELLENT 10 GOOD 11 EXCELLENT 12EXCELLENT 13 EXCELLENT 14 EXCELLENT	fair-excellent	CAHAI more adept to distinguishchange than ARAT, ROC curve ARAT 0.88, CAHAI ROC curve 0.95 so latter is more adept at distinguishing change	plus
ARAT	Barreca, Stratford, Masters, Lambert, Griffiths 2006	no	was the CAHAI-13 more adept than the ARAT in detecting true change in upper limb function over time. Objective: to determine whether the longitudinal validity of scores on 2 versions of the CAHAI was significantly greater than that of scores on the ARAT.	1. % of missing items not given can see n=105 at initial and n=100 at follow up GOOD 2. not clear how handled FAIR 3. excellent 4. excellent 5. states in abstract only 2-6 weeks GOOD 6. GOOD 7. FAIR unclear as nil global rating scale to indicate if change occurred 8. FAIR hypothesis vague 9. EXCELLENT " was CAHAI scores greater than ARAT" 10 GOOD 11 EXCELLENT 12 EXCELLENT 13. EXCELLENT 14 EXCELLENT	Fair - excellent	longitudinal validity ROC curves: CAHAI-13 = 0.86 (95% CI = 0.78- 0.93) CAHAI-9 = 0.82 (95%CI = 0.73 - 0.90) and ARAT = 0.72 (95% CI = 0.62 - 0.83). Curve areas for 2 versions of CAHAI were significantly greater than thatof ARAT. longitudinal validity of both CAHAI versions superior to that of ARAT.	plus

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
ARAT	Blennerhassett, Avery & Carey 2010	no	not stated specifically but aim: investigate HFS ability todetect change and comparedto ARAT	1. Yes EXCELLENT 2 no FAIR 3 POOR 4 EXCELLENT 5 EXCELLENT 6 EXCELLENT 7 GOOD no evidence provided but assumable 8 FAIR 9 GOOD 10 GOOD 11 EXCELLENT 12 EXCELLENT 13 EXCELLENT 14 EXCELLENT	Poor - excellent sample excluded fair	amount of change observed for HFS moderately agreed to that ofARAT (Rho_c= 0.62, 95% CI: 0.35-0.90; Kw = 0.65). The median IQR change relative to thescale was 7.7% (0-28%) for the HFS and 2.6% (0-22.4%) for ARAT. Floor and ceiling effect noted in ARAT and HFS. acute stroke	? COSMIN UPDATE plus as met hypotheses #1 same results
ARAT	Edwards, Lang, Wagner, Birkenmeier & Dromerick 2012	no	not stated	1. yes EXCELLENT 2. GOOD 3 GOOD 4 EXCELLENT 5 EXCELLENT 6 EXCELLENT 7 GOOD 8 FAIR 9 GOOD 10 GOOD 11 EXCELLENT 12 EXCELLENT 13 GOOD representative due to impairments (mild to mod) 14 EXCELLENT - single population effect size with comparator	fair - excellent	single population effect size method used to calculate effect size coefficients as indicators of scale responsiveness. Based on Cohens criteria coefficients of 0.8or greater are considered large, 05 to 0.8 moderate and less than 0.5 small. Responsiveness of ARAT at Day 0: to day 14: ARATtotal: 1.018ARAT gross:0.729 ARAT grasp: 1.042 ARAT grip: 1.017 ARAT pinch 0.854 Day 0 today 90 ARAT total: 1.390, ARAT gross : 0.984 ARA grasp: 1.224 ARAT grip: 1.324 ARAT pinch: 1.494. acute stroke	

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
ARAT	O'Dell, Kim, Rivera, Fieo, Christos, Polistena, Fitzgerald & Gorga 2013	no	not stated	1 yes EXCELLENT 2 GOOD not escribed but can be deduced 3 POOR N=29 4 EXCELLENT 5 EXCELLENT 6 EXCELLENT 7 GOOD 8 POOR 9 GOOD 10 GOOD 11 EXCELLENT 12 EXCELLENT 13 not representative sample, assessor not blinded, measures not administered in random order FAIR 14 used SRM FAIR	Poor - excellent	ARAT SRM 0.89 in chronic stroke	?
BI	Ashford, Turner-Stokes, Siegert & Slade 2013	no	nil specific reported	1. excellent 2 n/a 3 good sample size 58 4 excellent 5 excellent 8 weeks 6 good - assumable undergoing intervention 7 good 8 fair 9 good 10 good 11 excellent 12 excellent 13 fair - no information provided on raters, blinding etc minimal detail 14 excellent	fair - excellent	ArmA more responsive in detecting change between responders and non responders on passive items U = 127.0; p = 0.07 and active items U = 176.5; p = 0.92 or BI U = 200.5; p = 0.17 - none demonstrated any difference between the two groups	?
BI	Khan, Pallant & Turner-Stokes 2008	nil	nil specific but objectives (1) explore the type and nature of person-centred goals that are commonly set and achieved during the program (2) compare the responsiveness and relative efficiency of GAS with the FIM and BI as outcome measures for rehab in MS	1. GOOD not described 2 N/A 3. POOR sample size n 24) 4 EXCELLENT 5 EXCELLENT 6. EXCELLENT 7 GOOD 8 POOR 9 GOOD 10 GOOD 11 EXCELLENT 12 EXCELLENT 13 EXCELLENT 14 FAIR SRM and effect size used	poor - excellent	GAS: Effect size 9.0, SRM 2.4, Relative efficiency (t value) 10.0 relative efficiency (z value) 1.4. FIM ES 0.4 SRM 1.0 relative efficiency (t value) 1.1 relative efficiency (z value) 1.1 BI: Effect size 0.4, SRM 0.8, Relative efficiency (t value) 1.0 relative efficiency (z value) 1.0. GAS most responsive then Film then BI in MS.	

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
FIM	Khan, Pallant & Turner-Stokes 2008	nil	nil specific but objectives (1) explore the type and nature ofperson-centred goals that arecommonly set and achieved during the program (2) compare the responsiveness and relative efficiency of GASwith the FIM and BI as outcome measures for rehab in MS	1. GOOD not described 2 N/A 3. POOR sample size n 24) 4 EXCELLENT 5 EXCELLENT 6. EXCELLENT 7 GOOD 8 POOR 9 GOOD 10 GOOD 11 EXCELLENT 12 EXCELLENT 13 EXCELLENT 14 GOOD SRM and effect size used but no hypotheses	poor - excellent	GAS: Effect size 9.0, SRM 2.4, Relative efficiency (t value) 10.0 relative efficiency (z value) 1.4. FIM ES 0.4 SRM 1.0 relative efficiency (t value) 1.1 relative efficiency (z value) 1.1 BI: Effect size 0.4, SRM 0.8, Relative efficiency (t value) 1.0 relative efficiency (z value) 1.0. GAS mostresponsive then FIm then BI in MS.	? + (2/2)
Frenchay Arm Test	Sunderland, Tinson, Bradley & Hewer 1989	no	not clearly stated; the study investigated the relationship between grip strength, spasticity and functional recovery to discover whether in fact it (grip strength) may be a valuable marker of recovery in the typical stroke patient.	1. EXCELLENT yes 7 lost to follow up 2. GOOD not clear but can see that n=31 were included in the analysis 3. FAIR sample size 4. EXCELLENT 5. EXCELLENT 6. GOOD stated continued with physio but not clear 7. EXCELLENT 8. FAIR hypotheses vague 9. GOOD not stated 10 GOOD not stated 11. EXCELLENT12 EXCELLENT 13 FAIR - not stated who completed assessments? Blinded nil info to determine presence of bias 14 FAIR	Poor - excellent	responsiveness was rated by counting the number of cases where scores increased between adjoining sessions at initial- 1month: 12 1-3mnths: 11 3-6 months: 6. most increase of adjoining initial - 1 month: MI n=22, 1-3mnths MI n=22, 3-6 months grip strength n=19, least numbers initial - 1 month FAT and9HPT n=12, 1-3 months FAT n=11, 3-6 months motor club n=5 (FAT n=6).	? correlations not performed
GAS	Barden, Baguley, Nott, Chapparo 2014	nil	nil specific - aimed to evaluate change in UL performance following BTX-A as measured by clinical measures and dynamometry	1. GOOD 2 N/A 3. POOR 4 EXCELLENT 5. EXCELLENT 6 FAIR unclear what occurred in interim period 7 FAIR nil info found re this 8 FAIR 9 GODD 10 GOOD 11 EXCELLENT 12 FAIR only dynamometry info found 13 EXCELLENT 14 FAIR ES only	Poor - excellent	ES 0.78	? No hypothesis

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
GAS	Barden, Baguley, Nott, Chapparo 2014 (b)	no	study aimed to evaluate change in hand performance of adults with the UMN syndrome after BTX-A as measured by DCD and current clinical measures	1. EXCELLENT 2. N/A 3. poor sample size 4. EXCELENT 5 EXCELLENT 4 weeks 6. EXCELLENT 7. GOOD 8 fair 9 GOOD 10 GOOD 11 GOOD 12 GOOD 13 EXCELLENT 14 FAIR stats not optimal as only effect size without clear hypothesis	fair	GAS effect size 0.79	?
LASIS	Ashford, Slade, Nair & Turner-Stokes 2014	no	to assess the ability of the ArmA to detect changes (LASIS also compared)	1. EXCELLENT 2. GOOD 3. Fair sample size 4. Excellent 5. excellent 6 excellent 7 Good - didn't include global measure of change but various other measures also included and detected changes 8 Poor unclear what was expected no hypothesis 9 Good 10 good 11 excellent 12 excellent 13 excellent	poor - excellent	LASIS less sensitive than ArmA in detecting passive function change	?
LASIS	Ashford, Turner-Stokes, Siegert & Slade 2013	no	nil specific reported	1. excellent 2 n/a 3 good sample size 58 4 excellent 5 excellent 8 weeks 6 good - assumable undergoing intervention 7 good 8 fair 9 good 10 good 11 excellent 12 excellent 13 fair - no information provided on raters, blinding etc minimal detail 14 excellent	fair - excellent	ArmA more responsive in detecting change between responders and non responders on passive items U = 127.0; p = 0.07 and active items U = 176.5; p = 0.92 or BI U = 200.5; p = 0.17 - none demonstrated any difference between the two groups	?

Responsiveness							
Outcome Measure	**Author**	**Gold standard**	**Hypothesis**	**Findings**	**COSMIN result**	**Study statistical results**	**Terwee criteria**
LASIS (pt disability scale, carer burden scale)	Barden, Baguley, Nott, Chapparo 2014	nil	nil specific - aimed to evaluate change in UL performance following BTX-A as measured by clinical measures and dynamometry	1. GOOD 2 N?A 3. POOR 4 EXCELLENT 5. EXCELLENT 6 FAIR unclear what occurred in interim period 7 FAIR nil info found re this 8 FAIR 9 GODD 10 GOOD 11 EXCELLENT 12 FAIR only dynamometry info found 13 EXCELLENT 14 FAIR ES only	Poor - excellent	Pt disability scale ES 0.45 carer burden scale not computed	
Motricity Index	Sunderland, Tinson, Bradley & Hewer 1989	no	not clearly stated; the study investigated the relationship between grip strength, spasticity and functional recovery to discover whether in fact it (grip strength) may be a valuable marker of recovery in the typical stroke patient.	1. EXCELLENT yes 7 lost to follow up 2. GOOD not clear but can see that n=31 were included in the analysis 3. FAIR sample size 4. EXCELLENT 5. EXCELLENT 6. GOOD stated continued with physio but not clear 7. EXCELLENT 8. FAIR hypotheses vague 9. GOOD not stated 10 GOOD not stated 11. EXCELLENT12 EXCELLENT 13 FAIR - not stated who completed assessments? Blinded nil info to determine presence of bias 14 FAIR	Poor - excellent	responsiveness was rated by counting the number of cases where scores increased betweenadjoining sessions at initial- 1month: 22 1-3mnths:22 3-6 months:9. MI was most sensitiveto detect early change. most increase of adjoining initial - 1 month: MI n=22, 1-3mnths MI n=22, 3-6 months grip strength n=19, least numbers initial - 1 month FAT and 9HPT n=12, 1-3 months FAT n=11, 3-6 months motor club n=5 (FAT n=6).	correlations not performed

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
10MWT	Donovan, Lord, McNaughton & Weatherall 2008	nil missing	N/A	not given	not given	not given	mean and SD all participants 40.6 (7.9) IQR 35.1-46.6 Range 20-50	not stated
10MWT	Hirsch, Williams, Norton & Hammond2014	yes nil	not applicable	no	Nil	nil	mean and SD given for each of 6 trials of both fast and self selected paces. For self selected pace average was given for first three and second three trials as was for fast pace	not stated
10MWT	Kuys, Bew, Lynch, Morrison & Brauer 2009	yes 0	no	no	not stated	not stated	admission 10MWT mean and SD	not stated
10MWT	Miller, Combs, Van Puymbroeck, Altenburger, Kean, Dierks & Schmid 2013	nil	n/a	not given	not given	not given	high fatigue and high pain mean and SD 1.28 +/- 0.66 low fatigue and pain 1.47 +/- 0.60 plus whole sample 1.32 +/- 0.64	not stated
10MWT	Mudge & Stott 2009	1 participant did not havefull data	not included in analysis	range only 0.12-1.42	not given	not given	mean SD 0.67+/-0.32 of one group only	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
10MWT	Salbach, Mayo, Higgins, Ahmed, Finch & Richards 2001	YES n=3 (for true missing data) some missing for TUG but substituted data applied	yes	not given	0 - however assessments not started until participants could walk and exclusion criteria non ambulant by 3 weeks..	evaluation 1: comfortable pace 2 (4%) max pace 3(6%) evaluation 2 comfortable pace 10(20%) max pace 8(16%)	mean and SD given for evaluation 1 and 2	not given
10MWT	Schmid, Van Puymbroeck, Altenburger, Dierks, Miller, Damush & Williams 2012	no	N/A	no	not given	not given	gait speed mean and SD given	not stated
10MWT	Scrivner, Schurr & Sherrington 2014	yes 3/190 had incomplete data	data was excluded from study	yes	127 (66.8) on admission 49 (25.8%) on dc	nil at admissionor dc	analysis for 10MWT on admission and sdc only admission: mean 0.17 (0.3) median 0.00 IQR 0.3, discharge mean 0.60 (0.5) median 0.56m.s IQR 1.0.	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
9HPT	Beebe & Lang 2009	yes over thethree testing periods n= 33, 28, 19	completed a paired t test to determine whether dfferences existed between the subjects who werelost to attrition andthose between 3 and 6 months, analyses showed no difference, nothing else statedbut assumed analyses were done without inclusion.	no	not stated	not stated	mean and SD range given	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
9HPT	Benedict, Holtzer, Motl, Foley, Kaur, Hojnacki & Weinstock-Guttman2011	not stated but can deduce 0	not required	frequency distribution given - the distribution approximate da Gaussian distribution (Kurtosis 41.80 Skewness 5.50) (log transformatio n due to considerable positive skew)2 - 47 seconds taken	2%	1%	??	not discussed
9HPT	Costelloe, O'Rourke, McGuigan, Walsh, Tubridy & Hutchinson 2008	not stated but can deduce 0	not required	not given	not given	not given	mean baseline scores and SD (in z score) - 0.16 (1.15), mean follow up -0.26 (1.12)	not stated for 9HPT

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
9HPT	Goodkin, Hertsgaard & Seminary 1988	not stated but can be deduced 0	not stated	not given	not given	not given	control group n = 126 mean 26.46 SD between individuals 10.62 SD within individuals 7.74. %change on successive trials n = 84 mean 0.62 SD between individuals 0.088 SD within individuals 0.078. Prospective group n = 310 mean 32.62 SD between 17.84 SD within 12.66 % changein successive trials n =179 mean 0.034 SD between 0.2927 SD within 0.2484	change of more than 20% would be less than 5% due to chance.

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
9HPT	Jacob-Lloyd, Dunn, Brain & Lamb 2005	yes, 3 lost between discharge and follow up	not stated but can deduce that were not included in the analysis	not stated	not stated for whole sample- but was discussed that participants included in the subsample - able to complete both9HPT and UL MI - 42% only were able to at discharge and 54% at follow up. At discharge 1st quartileat lowest score and an outlier at min score on follow-up	nil	raw score at discharge median: 56 IQR 40-70 , at follow up median 44.5 IQR 34-52.	not stated
9HPT	Poole, Nakamoto, McNulty, Montoya, Weill, Dieruf & Skipper 2010	nil	not required	not stated	not stated	not stated	geometric means and 95% confidence intervals stated for each MS subtype	not stated
9HPT	Schwid, Goodman, McDermott, Bever &Cook 2002	not stated but can deduce 0	not required	15.1 to 65.5 seconds	not given	not given	not given	not stated
9HPT	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987(a)	yes nil	n/a	not given	17 of the pts that scored 0/5 on Frenchay and 1 of those that scored 1-4/5.	17 of those that scored 5/5 on FAT scored above the cut off of 18 seconds	not given	not discussed

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
ARAT	Beebe & Lang 2009	yes	completed a paired t test to determine whether differences existedbetween the subjects who werelost to attrition andthose between 3 and 6 months, analyses showed no difference, nothing else statedbut assumed analyses were done without inclusion.	no	not stated	not stated	mean and SD range given	not stated
ARAT	Burridge, Turk, Notley, Pickering & Simpson 2009	yes	Not given, not clear how handled but can see that were treated as missing as were not included in the analysis	yes 3-37	no	no	no	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
ARAT	de Weerdt & Harrison 1985	nil	N/A	yes in columntable of 10 increments	data presented in column table of 10 increments so unableto determine % of lowest scores	data presented incolumn table of 10 increments sounable to determine % of highest scores	yes	not stated - other: 11 mins toadminister BFM and 8 for ARAT.
ARAT	Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, Edwards2006	no	no	no	not stated	yes n=16 (41%)	no	not stated
ARAT	Fleming, Newham, Roberts-Lewis & Sorinola 2014	yes	yes	yes	0%	0%	yes; change scores mean and SD and min and max change	? Not as such but does state what score on ARAT id required to predict MAL AOU.

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
ARAT	Lang, Edwards, Birkenmeier, Dromerick 2008	Not given but can be deduced from the paper (77% i.e. (52-12/12)	Not given, not clear how handled but can see that were treated as missing as were not included in the analysis	no	not stated	not stated	Mean +/- SD Day 0: 22.5 +/- 15.3 Day 14: 38.1 +/- 16.6 change +/- SD: 15.1 +/- 11.4	early after strokeestimated MCID (dominant side affected) Raw value: 12, Percentage of Scale: 21, effect size: 0.78. Estimated MCID (non dominant side affected) Raw value: 17, Percentage of Scale: 30, effect size: 1.10 SUMMARY MCID ARAT DOMINANT AND NON DOMINANT SIDES 12, 17 POINTS
ARAT	Lang, Wagner, Dromerick, Edwards2006	10 participants from 50 missed to follow up at 90 day Ax (5%)	no	no	not stated	not stated	scores and change scores all given	MCID 10% 6 points or more isa real and important change

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
ARAT	Lyle 1981	nil	n/a	not given	not given	not given	not given	not given
ARAT	Morris, van Wijck, Joice & Donaghy 2013	yes	yes	no	not stated	not stated	mean and SD	not stated
ARAT	Rabadi & Rabadi 206	no	no	mean and SD only	not stated	not stated	yes	not stated
ARAT	Rand & Eng 2015	45%	yes	median and IQR given only dc = 42.0 (19/57) 12 mo = 57.0 (39.5-57)	not given	not given but noted IQR at max score	not given	not given
ARAT	Yozbatiran, Der-Yeghiain & Cramer 2008	not stated but can deduce nil missing items	n/A	range of each subscale given for datafor 2 examiners	nil scored 0 at baseline nor follow up	not stated	mean and SD for each subscale given (for inter intra rater reliability)	not stated
ArmA	Ashford, Turner-Stokes, Siegert & Slade 2013	yes	yes	table 1	nil	37% for active function	reported	criterion based method MIC 2.5 passive, 1.1 active and distribution based 3 and 2.5 respectively.
BI	Ali, Fulton, Quinn, Brady on behalf of VISTA 2013	yes	no but can be deduced	no	no	no	not given	no

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
BI	Dennis, Wellwood & Warlow 1996	2	no but assumed omitted	yes	not given	not given	no	no
CMA	Dang, Ramsaran, Street, Syed, Barclay-Goddard, Stratford & Miller 2011	yes n=30	no assumed excluded	not given	no	no	mean SD and quartilesfor each section od AI II	not given
CMA	Gowland, Stratford, Ward, Moreland, Torresin, Van Hullenar, Sanford, Barreca, Vanspall & Plews 1993	yes	deduced	no	no	no	not given	not stated
DAS	Brashear, Zafonte, Corcoran, Galvez-Jimenez, Gracies, Gordon, Mcafee, Ruffing, Thompson, Williams, Lee & Turkel 2002	yes 1	yes excluded from analysis	not given	not given	not given	mean and SD given for evaluation 1 and 2 and for mean of evaluations 1 and 2	not stated
DAS	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	nil missing	n/a	given	0	0		not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
EQ-5D	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	nil	n/a	not given	yes to check	yes to check	yes mean and SD at baseline given in respect to each DAS domain for	not given
EQ-5D	Kuspinar, Finch, Pickard & Mayo 2014	yes	n/a	not given	not given	not given	given	not stated
EQ-5D	Moore, Wolfson, Alexandrov & Lapierre 2004	yes	n/a	not given	not given	not given	given	not stated
FIM	Brown, Therneau, Schultz, Niewczyk &Granger 2015	no	n/a	no	FIM mobility 55.1%	FIM mobility 0.2%	yes many subgroups	not given
FIM	Cullen, Krakowski & Taggart 2014	yes	Yes	not given	not given	not given	drivers and non drivers	not given
FIM	Cuthbert, Harrison-Felix, Corrigan, Bell, Haarbauer-Krupa &Miller 2015	yes	Yes	not given	not given	not given	yes	not given
FIM	Egan, Davis, Dubouloz, Kessler &Kubina 2014	yes	Yes	not given	not given	not given	not given	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
FIM	Grant, Goldsmith & Anton 2014	yes	Yes	mean, median and 25th, 75th percentiles given	not given	not given	yes	not given
IM	Kuys, Bew, Lynch, Morrison & Brauer 2009	yes 0	No	no	not stated	not stated	admission FIM mean and SD	not stated
FIM	Ouellette, Timple, Kaplan, Rosenberg & Rosario 2015	no	n/a	yes	not given	not given	yes	not given
FIM	Rabadi & Vincent	yes	n/a	yes	not given	not given	yes - different types	not given
mFrencha y Arm Test	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987sub study 1(a)	yes nil	n/a	not given	this sub study only discussed those that had achieved 5/5	all participants asonly those scoring 5/5 analysed	not given	not discussed
mFrencha y Arm Test	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987sub study 2(b)	not identified how many were actually missing	yes described what was done butnot for how many missing items	not given	17/56 30%	19-56(34%)	not given	not discussed

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
GAS	Bovend'Eerdt, Dawes, Izadi, Wade2011	n=1	not stated	not given	not stated	not stated	mean SD given for ARAT, NEADL, RMI, BI, GAS therapist 51.99 (11.01) assessor 53.51 (10.29)	not stated-check
GAS	Brock, Black, Cotton, Kennedy, Wilson & Sutton 2009	not stated but can deduce	not included in analysis	not given	7% made no progress or declined in function	not given		not statedcheck
GAS	Joyce, Rockwood & Mate-Kole 1994	not stated	n/a	no	not given	not given	mean and SD admission and Dc scores for standardised outcome measure	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
GAS	Khan, Pallant & Turner-Stokes 2008	not stated but can deduce o	n/a	given	all non responders has GAS change score of 16 or less, all responders 17 or more	all non responders has GAS change score of 16 or less, all responders 17 ormore	total sample GAS admission; 32.3 (29.0 - 33.0) DC 58.4 (43.7 - 61.2) Change 25.8 (16.2-28.4) CGI responders admission 32.8 (31.2 - 33.3) DC 60.6 (56.9-61.6) change 27.2 (25.5-29.3) CGI non responders admission 27.9 (27.0-31.0) DC 40.5 (35.9-43.6) change 11.5 (7.6-16.2).	GAS T-score change from baseline to evaluation of more than 10.
GAS	Lannin 2003	0	n/a	not given	not given	not given	no	no
GAS	Malec 1999	yes 54/88 only completed one of the outcome measures	no can assume	not given	not given	not given	not given	not stated
GAS	Malec 2001	yes	no assumed excluded	no	no	no	not given	no

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
GAS	Malec, Smigielski &DePompolo 1991	yes 7 dropout initially then 2 excluded from analysis as 2 mo and final scores identical	not included in analysis	not given	N/A	N/A	not given	not given
MAL	Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, Edwards2006	no	no	no	not stated	no participant reported doing all30 tasks, more than half reported not taking off shoes or putting key in door.	given for those with fullARAT and WMFT.	not stated
MAL	Harris & Eng 2007	not stated	not stated	not given	not given	not given	Mean 3.1, SD 1.6 range 0-5	not stated
MAL	Chen, Wolf, Zhang, Thompson & Winstein 2012	not stated but can deduce 0	not stated but deduced were excluded	baseline mean and SD	not given	not given	mean and SD given for individual measures	not given
MAL	Uswatte, Taub, Morris, Light & Tompson 2006	not stated but can deduce 0	n/a	not given	not given	not given	yes for immediate and delayed treatment groups and all participants	no

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
MAL - 28	Uswatte, Taub, Morris, Light & Tompson 2006	not stated but can deduce 0	n/a	not given	not given	not given	yes for immediate and delayed treatment groups and all participants	referred to another study <0.5
Motricity index	Bohannon 1991	n=0	N/A	range given for individual sections	not given	not given	mean, median and range for individual sections	not given
Motricity index	Collin & Wade 1990	yes	n/a	no	for MI LL only	for MI LL only	mean for each observer (n=2) 3 assessments but no change scores	not stated
Motricity index	Stone, Patel & Greenwood 1993	yes	not clearly	not given	not given	not given	not given	not stated
Motricity index	Wade & Hewer 1987	yes varied across 3 time points	no but can be deduced	not given	not given	not given	not given	not stated
otricity index	Jacob-Lloyd, Dunn, Brain & Lamb 2005	yes, 3 lost between discharge and follow up	not stated but can deduce that were not included in the analysis	not stated	0	4/22 (18%) scored max at discharge.	not whole sample but sub sample completingboth 9HPT and UL MI; standardised score at discharge median: 77 IQR 77-84 , at follow up median 100 IQR 77- 100.	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
NHPT	Morris, van Wijck, Joice & Donaghy 2013	yes	yes	no	63 (75%)	not stated	mean and SD	not stated
OHS	Rigby, Gubitz, Eskes, Reidy, Christian, Grover & Phillips 2009	0	n/a	not given	not given	not given	not given	not stated
OHS	Simon, Kumar & Kendrick 2008	no	no	not given	not given	not given	not given	not stated
RMA	Collin & Wade 1990	yes	n/a	no	2 scored 0 on RMA LL (other sections not given)	not clearly stated4 pts were in rank scoring 3-10	not given	not stated
RMA	Jones 1998	not stated but can deduce 0	n/a	not given	% not given but lowest score was obtained at both baseline (6 wks post) and follow up (12 wks post)	% not given but max scored at follow up not at baseline	mean median and SD given for 3 sections	not stated
RMA	Sackley 1990	not stated can deduce 0	n/a	not given	not given	not given	mean and SD given for all RMA sections and ADL scores	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
RMA	Adams, Pickering & Taylor 1997 (acute)	yes	not stated but deduced were excluded	no	discusses low and high, one quarter were at the extremes and a "substantial" group were only one item from the end of the scale - indicates a proportion of patients recovery may not be captured	see <---	not given	not stated
RMA	Adams, Pickering, Ashburn & Lincoln 1997 (2 non acute)	yes	no	no	partly given	partly given	not given	not stated
RMA - UL	Morris, van Wijck, Joice & Donaghy 2013	yes	yes	no	not stated	not stated	mean and SD	not stated
SA-SIP	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	nil missing	n/a	given	0	0		not given
SA-SIP	Edwards, Hahn, Baum & Dromerick 2006	yes n=0	n/a	yes via mean and SD	not given	not given	yes	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
SF-36	Robinson et al2009	yes	yes	yes	not given	not given	yes	MID: Physical functioning 4-9, role physical 6-8, social functioning 6-7, PCS 6 points.
SF-36	Dorman et al 1999	approx 3%'	Interpolation procedure	Table 1	Table 1 (itemised for each domain)	Table 1 (itemised for each domain)	Not reported	not reported
SIS	Duncan, Bode, Lai, Perera 2003	14/640	not stated		not given	not given	not given	not given
SIS	Duncan, Lai, Tyler, Perera, Reker & Studenski 2002	0	n/a	no	no	no	yes each SIS domain	no
SIS	Duncan, Reker, Kwon, Lai, Studenski, Perera, Alfrey & Marquez 2005	for whole study 13 incomplete for telephone nil	not described	yes - mean and SD given	not given	not given	yes given	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
SIS	Eriksson, Baum, Wolf & Connor2013	yes	no	yes - 19-100	0%	32.20%	yes	not given
SIS	Jenkinson, Fitzpatrick, Crocker & Peters 2013	yes - 73 cases only available (48.34%)	excluded	no	no	no	not given	no
SIS	Kwon, Duncan, Studenski, Perera, Lai & Reker 2006	n=198 reduced at 12 weeks and further at 16 weeksn=90	compared responders to non responders no significant differences, excluded	given	tables indicate 0	tables indicate 0	given	not stated
SIS	Lai, Studenski, Duncan & Perera 2002	stated missing but actual number not given	not stated	for each SIS domain given	not given	not given	given	not given
SIS	Wolf & Koster 2013	yes	not stated but deduced were excluded	no	no	no	not given	no
UL MAS	Miller, Slade, Pallant, Galea 2010	not stated but can deduce	not stated but can deduce that were not included in the analysis	yes	14%	9%	not given	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
UL MAS	Khan, Chien & Brauer 2013	not stated but can deduce 0	not stated	not given	not given	not given	not given ??	not given
UL MAS	Kuys, Bew, Lynch, Morrison & Brauer 2009	yes n=15	no	no	not stated	not stated	admission MAS items mean and SD	not stated
UL MAS	Loewen & Anderson 1988	not stated assumed 0	n/a	no	no	no	no	no
UL MAS	Loewen & Anderson 1990	yes 7	data excluded	not given	not given	not given	mean and SD of mMAS	not given
UL MAS	Pickering, Hubbard, Baker & Parsons 2010	yes 1 excluded 2 wrist #	not described just"excluded"	not given	subset 6: nil stated 7:nil stated 8: 13/25 (52%) unable to achieve a score greater than 0	6: 17/25 (67%) assessments indicated subject achieved max score, 7 12/25 (48%) achieved max 8: nil stated	not given	not given
UL MAS	Carr, Shephard, Nordholm & Lynne 1985	0	n.a	no only complete MAS	not given	not given	not given	not given
UL MAS	Lannin 2004	yes	n/a	no	no	no	not given	no

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
UL MAS	Sabari, Lim, Velozo,Lehman, Kieran & Lai 2005	not stated but can be deduced	not stated but can deduce 0	not given	upper arm function (6) 31% participants ,hand movements (7) 31%, advanced hand activities (8) 38%	upper arm (6) 28% hand movement (7) 28% advanced (8) 9% ranging sample - acute tochronic stroke attending for UL therapy	not given	not given
Excluded at update								
10MWT	Baer & Smith 2001	yes n=5	no but can be deduced	not given	?	?		not given
10MWT	Bower, McGinley, Miller, Clark 2014	yes	not detailed	not given	not given	not given	day 1 and day 2 change mean (95% CI) -0.01 (-0.18, 0.16)	not given
10MWT	Smith & Baer 1999	yes 9	data was excluded	not given	22.3% were never able to achieve 10mWT within study timeframe	not given essentially just those that were able to achieve 10mWT 77.7%	not given	not given
10MWT	Vernon, Paterson, Bower, McGinley, Miller, Pua & Clark 2015	yes	n/a	not given	not given	not given	day 1 and 2 change mean and SD given	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
10MWT	Wolf, Catlin, Gage, Gurucharri, Robertson & Stephen 1999	not stated but can seen=0	n/a	not given	not given	not given	mean, SD and rage given for without impairment and with stroke	not given
10MWT - comfortabl eand fast pace	Combs, Dugan, Passmore, Reisner, Whipker, Yingling & Curtis 2010	3	not described can assume omitted	given	not given	not given	given	not given
9HPT	Alusi, Worthington, Glickman, Findley & Bain 2000	7, however unclear in article	not described but assumed omitted from correlation analyses	not stated	not stated	not stated	not stated	not stated
9HPT	Fisk, Brown, Sketris, Metz & Stadnyk 2005	yes	can be deduced	not given	not given	not given	mean 9HPT 87 seconds, SD 105 rang 12-300	not stated
9HPT	Marrie & Goldman 2011	nil	not stated, can deduce from data nil missing	not stated	not stated	not stated	mean 0.18 SD 0.97	not stated
9HPT	Rossier & Wade 2002	yes 0.39%	yes	not given	not given	not given	not given	not given
9HPT	Sunderland, Tinson,Bradley & Hewer 1989	yes 7	not stated but can deduce that were not included in the analysis	not stated	29/38 at initial	nil scored highest at initial	mean 0.03 SD 0.7	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
ARAT	Barden, Baguley, Nott, Chapparo 2014	not stated but can deduce 0	n/a	not given	11/11 100% in low grp pre ≤3 /57	0	low and high ARAT median and IQR given for pre and post injections	not stated
ARAT	Barden, Nott, Heard, Chapparo & Baguley 2012	yes	not applicable	median and range of scores given	26% of TBI group scored 0	5% of TbI group	median and range given (no change scores as one off assessment completed)	not stated
ARAT	Barreca, Stratford, Lambert, Masters & Streiner 2005	not stated but can deduce 0	n/a	not stated	not stated	not stated	mean and SD for initial, follow up and change score	refers to van DerLee reference of 10% of total score 5.7
ARAT	Barreca, Stratford, Masters, Lambert, Griffiths 2006	yes	no	mean and SDonly	not stated	not stated	initial Ax, follow up Ax and change mean andSD's given	reported the longitudinal validity of data for both CAHAI versions exceeded the clinically importance difference in area under the ROC curve of 0.07

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
ARAT	Barreca, Stratford, Masters, Lambert, Griffiths & McBay 2006	not stated but can deduce 0	n/a	not stated	not stated	not stated	no	not stated
ARAT	Blennerhassett, Avery & Carey 2010	yes	no	yes	not given	% not given	% not given	not stated
ARAT	Celik, O'Malley, Boake, Levin, Yozbatiran & Reistetter 2010	not stated but can deduce 0	n/a	each participant pre and post ARAT total score stated	nil	2 of 9 participants	not stated but can be deduced from table	not stated
ARAT	Edwards, Lang, Wagner, Birkenmeier & Dromerick 2012	yes	not given but can deduce that were not included in analysis	no	floor % for 3 time points given; day 0: 5.9, day 14: 2 day 90: 2.1	scores at ceiling given day 0: 3.9% day 14: 22% day 90 33%	yes	not stated
ARAT	McDonnell, Hillier, Ridding & Miles 2006	not stated but can deduce is 0	n/a	each participant ARAT total score stated	% not given	% not given	mean and SD, medianand range given for both affected and non affected side	not stated
ARAT	Notley, Turk, Pickering, Simpson & Burridge 2007	yes	not given but can be deduced	no	0%	0%	scores given only no change scores or SD	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
ARAT	O'Dell, Kim, Rivera, Fieo, Christos, Polistena, Fitzgerald & Gorga 2013	stated 3 lostto follow up with reasons given	not stated but can deduce that were not included in the analysis	range 0-39 stated	normalised scores groups towards lowerend of scale in sample with severe functional limitations	normalise scoresindicated none atmax score	mean 11.3 SD 10.3	not stated
ARAT	Page, Hade & Persch 2015	not stated	not described	mean and SDof pre test 1 and 2 given	not given	not given	given	not given
ARAT	Page, Levine, Hade2012	Not given but can be deduced from the paper (77% i.e. (52-12/12)	not given but can be deduced nil action needed	mean and SDonly	not stated	not stated	not given	minimal detectable change given ARAT 22.54 - not interpretability but measurement error
ARAT	Stinear, Barber, Petoe, Anwar & Byblow 2012	8 lost from consent to 12 week follow up	not stated assumed were not included in analysis	stratified scores given	3 at 72 hrs 4 at 2 weeks but difficult to clearly ascertain fromgraph	difficult to determine from graph but appears at higher end of scale particularly at 2 weeks mark	median and range for baseline ARAT 34 (0-57)	12

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
ARAT	Urbin, Waddell & Lang 2015	not stated	not stated	8-46, 10-57	not given	not given	<30 days 23./4 (13.2), >6 mo 33.2 (14.2)	not given
BI	Khan, Pallant & Turner-Stokes 2008	not stated but can deduce o	n/a	given	nil	nil	as above	not given
BI	Lai, Studenski, Duncan & Perera 2002	stated missing but actual number not given	not stated	not given but all equal to orgreater than 95	95	not given	not given	not given
BI	Maujean, Davis, Kendall, Casey & Loxton 2014	yes	n/a	not given	not given	not given	not given	not given
BI	Sprigg, Selby, Fox, Berge, Whynes, Philip & Bath 2013	yes	yes	no	not given	not given	not given	not given
BI	Wolf & Koster 2013	1	not stated but deduced were excluded	no	no	no	not given	no
BI	Ashford, Turner-Stokes, Siegert & Slade 2013	0	n/a	not given	not given	not given	at 3 timepoints	not given
CMA	Barreca, Stratford, Masters, Lambert, Griffiths & McBay 2006	not stated but can deduce 0	n/a	not stated	not stated	not stated	no	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
CMA	Barreca, Stratford, Lambert, Masters & Streiner 2005	not stated but can deduce 0	n/a	not stated	not stated	not stated	mean and SD for initial, follow up and change score	not stated
CMA	Ellis, Sukal, DeMott & Dewald 2008	not stated assumed 0	no	yes given	0%	o%	not given	not given
CMA	Huijbregts, Gowland& Gruber 2000	yes - 4 werenot fully interviewed	assumed relied oncaregivers and/or data excluded from analysis	given in varying formsas applicableto the objectives of the study	nil	nil	given for each item of activity inventory	MCID on activityinventory is 8 points (20 is a large and important change)
CMA	Levin, Desrosiers, Beauchemin, Bergeron & Rochette 2004	not state but can deduce 0	n/a	yes given	CMA hand 0 arm 0	CMA hand 6 arm9/28	scores in categories given i.e. 2-3 4-5 etc	not stated
CMA	Manns, Tomczak, Jelani, Cress & Haennel 2009	not given but can deduce 0	n/a	not given	not given	not given	CMA mean and SD given	not given
CMA	Oczkowski & Barreca 1993	3 died	no	no	no	no	not given	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
CMA arm and hand impairment	Coderre, Zeid, Dukelow, Demmer, Moore, Demers, Bretzke, Herter, Glasgow, Norman, Bagg & Scott 2010	o	n/a	given	0	left affected handscore 24/26 (92.31%) arm score 23/26 (88.46%) right affected hand score 10/26 (38.46%) arm score	not given	not given
CMSA	Semrau, Herter, Scott, Dukelow 2015	yes 9%	yes	not given	not given	not given	yes all individual clinical measures	not stated
DAS	Brashear, Gordon, Elovic, Kassicieh, Marciniak, Do, Lee, Jenkins, Turkel, 2002	yes 4 in control group	not given unsure	not given	not given	not given	changes in mean scores given	not given
EQ-5D	Ali, Fulton, Quinn, Brady on behalf of VISTA 2013	yes	no but can be deduced	no	no	no	not given	no
EQ-5D	Golicki et al	yes	yes	not given	not given	not given	not given	not given
EQ-5D	Jenkinson, Fitzpatrick, Crocker & Peters 2013	yes - 73 cases only available (48.34%)	excluded	no	no	no	not given	no
EQ-5D	Kim, Jo & Lee 2015	yes	no	not given	not given	not given	yes	0.08 - 0.12

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
EQ-5D	Lunde 2013	15.40%	not stated but can be deduced	yes	nil	indicates ceiling effect but numbers not given	given	not stated
EQ-5D	Mitosek-Szewczyk et al 2014	no	n/a	yes	via graph	30%	yes	not given
EQ-5D	Sprigg, Selby, Fox, Berge, Whynes, Philip & Bath 2013	yes	yes	yes	not given	<15%	yes	not given
FIM	Cusick, Lannin, Hanssen & Allaous 2014	yes	yes	yes	not given	not given	yes FIM subscales	not given
FIM	Joseph, Pandit, Aziz et al 2013	no	no	yes	not given	not given	yes	not given
FIM	Khan, Pallant & Turner-Stokes 2008	not stated but can deduce o	n/a	given	nil	nil	as above	not given
FIM	Kwon, Duncan, Studenski, Perera, Lai & Reker 2006	n=198 reduced at 12 weeks and further at 16 weeksn=90	compared responders to non responders no significant differences, excluded	given	tables indicate 0	ceiling effect	given	not stated

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
FIM	McNett, Amato, Gianakis, Grimm, Philippbar, Belle & Moran 2014	yes	assumed	mean and range	not given	not given	yes	not given
FIM	Perrin, Niemeier, Mougeot et al 2015	yes	yes	not given	not given	not given	yes	not given
FIM	Semrau, Herter, Scott, Dukelow 2015	yes 9%	yes	not given	not given	not given	yes all individual clinical measures	not stated
FIM	Tyryshkin, Coderre, Glasgow, Herter, Bagg, Dukelow & Scott 2014	not given assumed 0	n/a	not given	not given	not given	yes	not given
FIM	Wu, Burgard & Radel 2014	0	-	mean and SD	not given	not given	yes	not given
mFrenchay Arm Test	Sunderland, Tinson,Bradley & Hewer 1989	yes 7	not stated but can deduce that were not included in the analysis	not stated	25/38 at initial assessment	6/38 at initial	mean 1.1 SD 1.9	not stated
GAS	Barden, Baguley, Nott, Chapparo 2014	not stated but can deduce 0	not included in analysis	not given	not stated	not stated		not stated - check
GAS	Barden, Baguley, Nott, Chapparo 2014 (b)	not stated but can be deuced 0	n/a	mand and SDonly	not given	not given	mean and SD for all sample	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
LASIS	Barden, Baguley, Nott, Chapparo 2014	not stated but can deduce 0	not included in analysis	not given	not stated	not stated		not stated - check
LASIS	Ashford, Slade, Nair & Turner-Stokes 2014	0	n/a	not given	no	no	no	no
LASIS	Ashford, Turner-Stokes, Siegert & Slade 2013	0	n/a	not given	not given	not given	at 3 timepoints	not given
MAL	Atler, Malcolm & Griefe 2015	0	n/a	no	not given	not given	not given	not given
MAL	Borstad & Nichols-Larsen 2016	0	n/a	yes	nil lowest possible	1/12 had highest score	not given	not given
MAL	Celik, O'Malley, Boake, Levin, Yozbatiran & Reistetter 2010	not stated but can deduce 0	n/a	each participant pre and post MAL averageof all ratings (both QOM and AOU) stated	not stated but can deduce from table - nil floor effect noticed	not stated but can deduce from table - nil ceiling effect noticed	not stated but can be deduced from table	not stated
MAL	Harris & Eng 2006	not given but can deduce 0	n/a	not given	not give	not given	given for all measures	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
MAL	Mark, Woods, Mennemeier, Abbas& Taub 2006	yes several across different cognitive assessmen ts	not clear	mean and SDgiven	not given	not given	mean and SD given	not given
Motricity Index	Kopp, Kunkel, Flor, Platz, Rose, Mauritz, Gresser, McCulloch & Taub 1997	not stated	not stated	66 - 96	not given	not given	median MI 89.8 when broken into 2 groups week 1 group median 92 and two week group 87.5 (high functioning)	not stated
Motricity Index	Smith-Arena, Edelstein, Rabadi, 2006	yes n= 6	no	no	not given	not given	mean and SD given for UL MI for total pop, those who failed driving evaluation and those who passed	not stated
Motricity index	Sunderland, Tinson,Bradley & Hewer 1989	yes 7	not stated but can deduce that were not included in the analysis	not stated	16/38 at initial assessment	1	mean 34 SD 36	not stated
OHS	Dennis, Wellwood & Warlow 1996	2	no but assumed omitted	yes	not given	not given	no	no
OHS	Gubitz, Reidy, Christian & Phillips 2012	72/522 deaths	unsure	not given	not given	not given	not given	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
OHS	Herrman, Black, Lawrence, Szekely & Szalai 1998	no	no	not given	not given	not given	not given	not stated
OHS	Pittock, Meldrum, Dhuill, Hardiman & Moroney 2003	yes	not stated but deduced were excluded	not stated	not given	not given	yes for different time points and OPS groups	not stated
OHS	SCOPE Collaborations 7 IST 2007	yes	not stated but deduced were excluded	not given	not given	not given	not given	not stated
RMA	Pittock, Meldrum, Dhuill, Hardiman & Moroney 2003	yes	not stated but deduced were excluded	not stated	not given	not given	yes for different time points and OPS groups	not stated
RMA	Taylor, Ashburn & Ward 1994	yes 1 unable to sitat 6 weeks	appears to have been excluded from analysis	median and IQR only	% not given	% not given	median and IQR given at week 1, 3, 6	not given
RMA - UL	Meldrum, Pittock, Hardiman, Dhuill, O'Regan & Moroney2004	yes	assumed excluded	no mean and SD only	no	25 (22%) had max RAS of 15at 48hrs	yes mean and SD for 3different grps	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
SF-36	Rudick, Miller, Hass et al 2007	yes	n/a	not given	not given	not given	yes	discussed as 5 points from prev studies this study had 16.8 to 28.5% achieving MCID -improvement on scores and 16.5 -22.2% when worsening in scores.
SIS	Ali, Fulton, Quinn, Brady on behalf of VISTA 2013	yes	no but can be deduced	no	no	no	not given	no
SIS	Ellis, Sukal, DeMott & Dewald 2008	not stated assumed 0	no	no	no	no	not given	no
SIS	O'Dell, Kim, Rivera, Fieo, Christos, Polistena, Fitzgerald & Gorga 2013	stated 3 lostto follow up with reasons given	not stated but can deduce that were not included in the analysis	range 0-39 stated	not given	not given	SIS hand 15.3 (21.6) [5.0] (0-80)	not stated
SIS	Boger, Hankins & Latter 2015	% paper returned 23 (24%), online 72 (76%)	excluded	?	not given	not given	not given	not given

Interpretability								
Measure	**Author**	**% of missin gitems given**	**Description ofhow missing items were handled**	**Distributio nof the (total) scores**	**% of the respondents whohad the lowest possible (total) score**	**% of the respondents who had the highest possible (total)score**	**Scores and change scores (i.e. mean andSD) for relevant (sub)groups, e.g. for normative groups, subgroups of patients or the general population**	**Minimal Important Change (MIC)or Minimal Important Difference (MID)**
ULMAS	Horgan, Cunningham, Coakley, Walsh, O'Regan & Finn 2006	not given assumed 0	n/a	yes	0%	0%	given for total scores and time taken	not stated
ULMAS	Johnson & Selfe 2004	not stated	not stated	not given	not given	not given	not given	not given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
10MWT	Donovan, Lord, McNaughton & Weatherall 2008	61.3 (11.1)	70% male n=21	side of stroke, months post, MMSE, Berg balance	community stroke survivors	New Zealand	English	convenience	30 selected from 71 43%
10MWT	Kuys, Bew, Lynch, Morrison & Brauer2009	70 (13)	n = 64, 53% male	side of stroke, time post	Hospital / rehabilitation	Australia	English	consecutive	not stated but assumed0
10MWT	Miller, Combs, Van Puymbroeck, Altenburger, Kean,Dierks & Schmid 2013	whole sample 64.1 years 48-89, high fatigue level 63.9 48 89 high pain 65.3 53-84	58 males 12females	modified Rankin scale, education, race time post stroke and lesion R or L	community	USA	English	convenience	not given
10MWT	Mudge & Stott 2009	67.4 SD 12.5years	29 men 20 women	time post stroke 6 to 219 months post	community dwelling	New Zealand	English	convenience	not given
10MWT	Salbach, Mayo, Higgins, Ahmed,Finch & Richards 2001	68+/-13	31 (62%) male 19 (38%) female	side of lesion, stroke type and severity	hospital and community	? Canada	English	consecutive	357 admitted, 170 met eligibility, 65 approached,53 consented, data available for 50

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
10MWT	Schmid, Van Puymbroeck, Altenburger, Dierks, Miller, Damush & Williams 2012	mean SD 64.06 (8.78)	n= 19 women (25%)	stroke characteristics, time post, type and side of, modified Rankin score	community	USA	English	volunteer - convenience ?	not given
10MWT	Scrivner, Schurr &Sherrington 2014	76.0 (12.7)	93 female 49%	stroke type, modified Rankin score, Charlston co morbidity index	inpatient	Australia	English	consecutive	1014 admitted to unit, 200 met criteria, 7 died throughout, 3 had incomplete data, total of 190 analysed
10MWT	Hirsch, Williams, Norton & Hammond 2014	35.8 (14.2) years	22 male 1 female	initial FIM walk testing FIM walk no days in rehab and time post injury	inpatient rehabilit ation	USA	English	consecutive	not given
9HPT	Beebe & Lang2009	mean age 53.9 (SD10.2) years range 31-77	19 men (58%) 14 women (42%)	time since, typeof stroke info given	inpatient rehaband community	not stated	not stated	recruited from Cognitive Rehabilitatio nResearch Group Stroke Registry	the 33 included accounted for approx 10% of totalsubjects screened.

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
9HPT	Benedict, Holtzer,Motl, Foley, Kaur, Hojnacki & Weinstock-Guttman 2011	44.9 SD 10	79.1% female	MS type education ethnicity	outpatients	USA	English	retrospectiv edata analysis	not stated
9HPT	Costelloe, O'Rourke, McGuigan, Walsh, Tubridy & Hutchinson 2008	not given	not given	MS only	outpatients	Ireland	English	not stated	not stated
9HPT	Goodkin, Hertsgaard & Seminary 1988	control grp mean 45.24 SD 16.50 prospective grp mean 47.16 SD11.73	control M/F 7/14 prospective 25/43	disease duration given and type	outpatient clinic	not stated assumed USA	assumed English	consecutive	not stated
9HPT	Heller, Wade, Wood, Sunderland, Hewer & Ward1987 (a)	68.1 SD 11.4	24 men 32 women	medical diagnoses only	Hospital	Frenchay Hospital	English	consecutive	61 excluded
9HPT	Jacob-Lloyd, Dunn, Brain &Lamb 2005	not given; 47/55 or 85% were over 60 years	31 men 24 women	stroke deficits noted only including affected side.	inpatient rehabilitation and community	UK	English	consecutive	99 people admitted, 55 assessed. 56%

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
9HPT	Poole, Nakamoto, McNulty, Montoya,Weill, Dieruf & Skipper 2010	mean 46.8 years, range 23 to 66	45 women 11men	MS subtypes	outpatient clinics	USA	English	convenience	not stated
9HPT	Schwid, Goodman, McDermott, Bever& Cook 2002	51.9 years SD +/- 9.0 years	74% women	MS and ambulating	not stated	USA	English	not stated	not stated
ARAT	Beebe & Lang2009	mean age 53.9 (SD10.2) years range 31-77	19 men (58%) 14 women (42%)	time since, typeof stroke info given	inpatient rehaband community	not stated	not stated	recruited from Cognitive Rehabilitatio nResearch Group Stroke Registry	the 33 included accounted for approx 10% of totalsubjects screened.
ARAT	Burridge, Turk, Notley, Pickering& Simpson 2009	57 (13)	male 11 (65%) female 6 (35%)	time post stroke, hemiplegic side, dominance	outpatie nt rehabilit ation clinics	UK	English	convenience	yes data for2 participants for tracking and FMI lostdue to technical issue.
ARAT	de Weerdt & Harrison 1985	mean 68.6 SD 9.3	25 male 28 female	acute stroke	acute medical ward	not stated	English	consecutive admission	nil

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
ARAT	Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, Edwards 2006	mean age 64.54 SD14.13	male 17 (44%) female 22 (56%)	race, stroke type affected side	Barnes-Jewish Hospital	USA	English	consecutive admission and screened eligible	not stated
ARAT	Fleming, Newham, Roberts-Lewis & Sorinola 2014	61.5yrs, no SDgiven	Female 13 males 20 (noted 3 lost atfollow up Ax)	mean time post stroke	univer sity laborat ory - outpati ents	UK	English	Recruited from National Heath Service, stroke support groups, word of mouth	not stated
ARAT	Lang, Edwards, Birkenmeier, Dromerick 2008	age 64 +/- 14	male 21 (40) female 31 (60)	pre morbid Barthel 99.6 +/- 2.2, pre morbid modified rankin 0.3 +/- 0.6, 79% post ischaemic stroke, 56% had non dominant side affected	inpatient rehabilitation unit	St Louis, USA	English	recruited via the Cognitive Rehabilitati on Research Group Stroke Registry fromacute neurology service	1850 patients screened to include 52.

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
ARAT	Lang, Wagner, Dromerick, Edwards 2006	63.7+/-13.6	males: 21 (42%)female	78% ischaemic 22%haemorrhage stroke, acute stroke	Barnes-Jewish Hospital, St Louis	USA	English	29 (58%)	not stated
ARAT	Lyle 1981	53.2 mean range 26 72	65% male (13 male 7 female)	stroke, ABI, aneurysm	outpatient	?	English	not stated	not given
ARAT	Morris, van Wijck Joice & Donaghy2013	median 69 range 36-88	male 49 female 36	type and duration post stroke	recruited within acute presentation but followed up and assessed in community	unknown	not stated	consecutive admission	given
ARAT	Rabadi & Rabadi2006	72 +/- 13	male/female43/61	stroke type, race, comorbidities	inpatient rehabilitation	not stated	not stated	consecutive admissions	not stated
ARAT	Rand & Eng 2015	58.1 (12.4)	M/F 25/7 (78% male)	equal L and R hemisphere infarct	community	Canada	English	consecutivethen volunteered	68 / 125 eligible 3 decided not to participatetotal of 10 dropped out for various reasons

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
ARAT	Yozbatiran, Der-Yeghiain & Cramer 2008	mean age 61yrs SD:+/- 15, range 39-86	6 males 6 females	chronic stroke (>3 months post)	not stated	not stated - deduced to be USA	English	not stated	not stated
BI	Ali, Fulton, Quinn, Brady 2013	71 (60-78)	male 2715 (54.9) of whole sample	other medical conditions	assumed hospital	international repository	English	?	not given
BI (C&W)	Dennis, Wellwood& Warlow 1996	73.1 mean	68 (44%) men	nil info given	community	UK	English	consecutive	266 admitted 152 recruited 2 missing data
CMA	Dang, Ramsaran, Street, Syed, Barclay-Goddard, Stratford & Miller 2011	65.3 (12.4)	26 (35%) female	less than 45 days post, co morbidities	hospital	Canada	English	convenience (database data)	not given
CMA	Gowland, Stratford, Ward, Moreland, Torresin, Van Hullenar, Sanford, Barreca, Vanspall& Plews 1993	mean 64 range 18-86	18 women 14men	14 right hemi 14 left hemi 4 bilateral, stroke type and sensory motor deficits discussed, comorbidities discussed, time post mean 9 weeks	rehabilitation - inpatient and day hospital	Canada	English	consecutive	not given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
DAS	Brashear, Zafonte, Corcoran, Galvez-Jimenez, Gracies, Gordon, Mcafee, Ruffing, Thompson, Williams, Lee & Turkel 2002	age of pts 37.5 +/-4.53, median 39	not given	professionals completing assessments neurologist n=4 physiatrist n=2 OT n=2 physical therapist n=2 with mean experience 6.6 years +/- 4.99	outpatients	USA	English	not stated	not given
DAS	Doan, Brashear, Gillard, Varon, Vandenburgh Turkel & Elovic 2012	mean 58.2 range 21 88	150 (53.8%) men	thrombotic stroke 45.5%, mean time since strokeonset 5 years range 0.2 - 31.5	community	USA	English	not stated	not given
EQ-5D	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	mean 58.2 range 21 88	150 (53.8%) men	thrombotic stroke 45.5%, mean time since strokeonset 5 years range 0.2 - 31.5	community	USA	English	not stated	not given
EQ-5D	Kuspinar, Finch, Pickard & Mayo 2014	43 (10.2)	49 (26)	mild disability	community	Canada	English	random	not stated
EQ-5D	Moore, Wolfson, Alexandrov &Lapierre 2004	45 (11)	18 (45)	mild	community	Canada	English/French Canadian	random	not stated
FIM	Brown, Therneau, Schultz,	70.6 (13.1)	71, 726 (48)	acute	hospital	USA	English	consecutive	not reported

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
	Niewczyk& Granger 2015								
FIM	Cullen, Krakowski& Taggart 2014	drivers 49.77 (15.25) non drivers 51.42 (15.73)	driver 28 (80) non driver 19 (79)		community	Canada	English	convenience	not reported
FIM	Cuthbert, Harrison-Felix, Corrigan, Bell, Haarbauer-Krupa & Miller 2015	76% <80	41204 (64.3)	mod to severe TBI	community	USA	English	consecutive	given
FIM	Egan, Davis, Dubouloz, Kessler& Kubina 2014	64.8 (13.3)	39(58.2)	high functioning, no cognitive impairment	community	Canada	English	consecutive	not given
FIM	Grant, Goldsmith& Anton 2014	median 72 (25th, 75th percentile 61, 81)	6581 (55)	43% R) body paresis	hospital	Canada	English	consecutive	not given
FIM	Kuys, Bew, Lynch, Morrison & Brauer2009	70 (13)	n = 64, 53% male	side of stroke, time post	Hospital / rehabilitation	Australia	English	consecutive	not stated but assumed0
FIM	Ouellette, Timple,Kaplan, Rosenberg &Rosario 2015	68.2 (13.9)	not given		hospital	USA	English	consecutive	not given
FIM	Rabadi & Vincent	53.6 (10.9)	63 (83)	MS type	community	USA	English	consecutive	not given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
GAS	Bovend'Eerdt, Dawes, Izadi, Wade 2011	50.28 (13.88)	11F/18M	time since onset (weeks) (n=28) 18.86 (16.19)	hospital rehabilitation	UK		not stated	not stated
GAS	Brock, Black, Cotton, Kennedy, Wilson & Sutton 2009	median 66 range 35-87	56% male	median length of stay in acute 14 (quartiles 9-23.5) median length ofIP rehab stay 31 days (quartiles 19.5 - 64.5)	community	Australia	English	consecutive convenience	not given
GAS	Joyce, Rockwood& Mate-Kole 1994	mean 27 (range 17-49)	9 men 7 women	average time post injury 3 months (range 2-5)	inpatient rehabilitation	Canada	English	consecutive	not given
GAS	Khan, Pallant & Turner-Stokes 2008	52.0 (8.3) range 37-62	10M/14F	mean time post diagnosis 11.1 (5) years range 2-23 years	inpatient rehabilitation	Australia	English	consecutive admissions	not stated
GAS	Lannin 2003	average 56.5 range 26 to 79	not given	ABI	community	Australia	English	consecutive	not given assumed 0

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
GAS	Malec 1999	mean 33.8 no SD given, median 32 range 18-69	64 men 24 women	75% TBI, 17% stroke, 8% anoxia plus other, schooling levels given, timepost injury ranged from 38 days to over 30 years, 39% injured within 1 year of admission, 61% more than one year prior to admission	community	USA	English	consecutive	not stated
GAS	Malec, Smigielski& DePompolo 1991	34.3 (12.2) median 32.5 range 18-60	not given	time since injury (days) 782.3 (1098.6) median 269.5 range 38 days to 8 plus years	outpatient braininjury program	? USA	English	convenience (consecutive)	0%
GAS	Turner-Stokes, Baguley, De Graaff, Katrak, Davies, McCrory & Hughes 2010	54.5 (13.2)	male : female 54 :36 60%	mean time post stroke 5.9 yrs (10.5)	community	Australia	English	consecutive	122 screened, 102 eligible 96 consented
MAL	Chen, Wolf, Zhang, Thompson& Winstein 2012	CIMT sample 60.98 (13.47) 28.6-84, control sample 63.26 (12.56) 18.5-89.8	CIMT sample %M/F 69/37 (65.1/34.9) control sample 73/43 (62.9/37.1)	affected hand, high/low function, Fugl-Meyer range	assum ed comm unity	USA	English	convenience	not given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
MAL	Dromerick, Lang, Birkenmeier, Hahn, Sahrmann, Edwards 2006	mean age 64.54 SD14.13	male 17 (44%) female 22 (56%)	race, stroke type affected side	Barnes-Jewish Hospital	USA	English	consecutive admission and screened eligible	not stated
MAL	Harris & Eng 2006	68.7 (9.4) 50-93	m/f 61/32	side of paresis, dominance, various impairment andfx measures	community	Canada	English	volunteer	not given
MAL	Uswatte, Taub, Morris, Light & Tompson 2006	all participants 62.2 (13)	female 80 (36)	stroke side, type	Community	USA	English	check EXCITEtrial	not given
MAL - 28	Uswatte, Taub, Morris, Light & Tompson 2006	all participants 62.2 (13)	female 80 (36)	stroke side, type	check EXCITE trial data ? Community	USA	English	check EXCITEtrial	not given
mFrenchay	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987 sub study 1	68.3 years range 5587	5 men 9 women	stroke only	community	not stated	?	convenience	nil
mFrenchay	Heller, Wade, Wood, Sunderland, Hewer & Ward 1987 sub study 2	68.1 SD 11.4	24 men 32 women	medical diagnoses only	Hospital	Frenchay Hospital	English	consecutive	61 excluded

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
Motricity Index	Stone, Patel & Greenwood 1993	72.37 (12.11) years	not given	stroke type neglect presence	hospital	UK	English	consecutive	99 from 171 included at 3 months
Motricity Index	Bohannon 1999	66.7 range 46-81	not given	within 15 days of stroke, intact proprioception and could follow instruction	assumed hospital not stated	USA	English	measures from database no other information	not given
Motricity Index	Collin & Wade 1990	male range 15-77 years mean 56.1 years female 45-69 mean 59.9	12 female 24 male (total)	right or left hemiplegia	inpatient rehab	UK	English	not stated	not stated
Motricity Index	Jacob-Lloyd, Dunn, Brain & Lamb 2005	not given; 47/55 or 85% were over 60 years	31 men 24 women	stroke deficits noted only including affected side.	inpatient rehabilitation and community	UK	English	consecutive	99 people admitted, 55 assessed. 56%
Motricity Index	Wade & Hewer1987	not given	not given	other than stroke only added if were admitted tohospital during first 6 months post or not	acute hospital to community	UK	English	consecutive	initial 976 alive 545 seen
NHPT	Morris, van Wijck Joice & Donaghy 2013	median 69 range 36-88	male 49 female 36	type and duration post stroke	recruited within acute presentation but followed up and assessed in community	unknown	not stated	consecutive admission	given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
OHS	Rigby, Gubitz, Eskes, Reidy, Christian, Grover & Phillips 2009	median 73 (IQR) 63.0 - 79.0	64 female (41.3%) 91 male (58.7%)		acute	Canada	English	consecutive	671 admitted 155 were included and had data
OHS	Simon, Kumar &Kendrick 2008	carer age 6.8(12.4) prior to discharge, 65.9 (13.4) 6 week post, 65.6(12.1) 15 mo post	prior to discharge77 female (73.2%), 6 week post 54 female (3%), 15 mo post 3 female (72.2%)	other care commitments	acute to community	UK	English	invited on carerecipient admission	105 carers at baseline to 53 at 15mo follow up
RMA	Adams, Pickering, Ashburn & Lincoln1997 (2 non acute)	grp 1 > 65 mean 75.39 (6.41) range 65-101, Grp 2 <65 yrs 56.54 (5.73) 44-64 grp 3 56.33 (5.95) 44-64	65> 50.5% men <65 62.2% men grp 3 54.2% men	site of lesion, previous stroke, timing of assessment (time post)	community	UK	English	consecutive	327 eligible
RMA	Collin & Wade1990	male range 15-77 years mean 56.1 years female 45-69 mean 59.9	12 female 24 male (total)	right or left hemiplegia	neuro rehab inpatient unit	UK	English	not stated	not stated
RMA	Jones 1998	66 (9.4)	16 females 13 males	16 left hemi 13 right hemi	inpatient rehab	UK	English	consecutive	not stated
RMA	Sackley 1990	right hemi - 63.4 (11.4) range 21-87 left hemi 63.2 (11.9) range 33-86	right hemi 23 male 29 female left hemi 23 male15 female	right or left hemi	hospital	UK	English	consecutive	not stated

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
SA-SIP	Doan, Brashear, Gillard, Varon, Vandenburgh, Turkel & Elovic 2012	mean 58.2 range 21 88	150 (53.8%) men	thrombotic stroke 45.5%, mean time since stroke onset 5 years range 0.2 - 31.5	community	USA	English	not stated	not given
SA-SIP	Edwards, Hahn, Baum & Dromerick2006	64.74 (15.87)	94 (43)	education, race, dc location	community	USA	English	consecutive	884 admitted, complete data for 771, 377 met criteria for mild stroke following exclusion n= 254. 18 lost to follow-up 3 died, 5 second stroke 5 institutionalised 4 declined.
SF-36	Moore, Wolfson, Alexandrov &Lapierre 2004	45 (11)	18 (45)	mild	community	Canada	English / French Canadian	random	not stated
SIS	Duncan, Bode, Lai, Perera 2003	68.6 (12.5) median70	female 310 (45) male 386 (55)	acute stroke between 1-3 mo post	acute and community	USA Canada	English	consecutive	not reported
SIS	Duncan, Lai, Tyler, Perera, Reker & Studenski 2002	pt 72.6 (10.0) proxy 59.8 (15.5)	pt 135 males (47%) 152 females, proxy males 78(27.2% females 209 (72.8%)	education, race, stroke type	community	USA	English	consecutive	20%

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
SIS	Duncan, Reker, Kwon, Lai, Studenski, Perera,Alfrey & Marquez 2005	whole sample mail68.48 (11.4) telephone 68.84 (12.2)	whole sample mail female 5/224(2.2%) telephone 4/234 (1.7%)	MRS, cognitive deficit	community	USA	English	random for this section but initial recruitment consecutive	mail response rate 45% telephone mode 68.7%
SIS	Eriksson, Baum, Wolf & Connor 2013	62.4 (12.7)	56 (48)	mild impairment	community	USA	English	convenience	not given
SIS	Jenkinson, Fitzpatrick, Crocker & Peters 2013	4 (2.7%) 18-44, 45 (30.4%) 45-64 41 (27.1%) 65-74 58 (39.2%) >75	56 (37.1%) women 88 (58.3%) men	mean time post stroke 7.3 years (6.1)	community	UK	English	convenience -recruited from database	525 eligible, 418 sent survey
SIS	Kwon, Duncan, Studenski, Perera,Lai & Reker 2006	68.05 (12.0)	unable to clearly state as missing values but original n=136 male 133 (97.8%)	stroke type prior living/function	community	USA	English	consecutive	48%
SIS	Lai, Studenski, Duncan & Perera2002	76.0 (6.56)	male 48 (59.4%)		community	USA	English	convenience	not given
SIS	Wolf & Koster 2013	grp 1 64.2 (13.4) grp 2 60.5(12.8)	grp 1 males 28 grp 2 31		community	USA	English	convenience	not given
UL MAS	Johnson & Selfe 2004	77 (9) range 45-88	13 men 13 women	29 days (18) range 7-83	hospital	Australia	English	consecutive	23 ineligible 47%
UL MAS	Khan, Chien & Brauer 2013	18-101	53% male 47% female	nil additional	inpatient rehabilitation	Australia	English	not stated	not stated

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
UL MAS	Kuys, Bew, Lynch,Morrison & Brauer 2009	70 (13)	n = 64, 53% male	side of stroke, time post	Hospital / rehabilitation	Australia	English	consecutive	not stated but assumed0
UL MAS	Lannin 2004	67(10.1)	male: female 15:15	lesion location	inpatient rehabilitation	Australia	English	not stated	not stated
UL MAS	Loewen & Anderson 1988	73.6 +/- 8.3	2 men 5 women	side of stroke	hospital	Canada	English	volunteer - convenience	not stated
UL MAS	Loewen & Anderson 1990	68 +/- 10 range 44-84	28 men 22 women	stroke cause,	hospital	Canada	English	consecutive	0
UL MAS	Pickering, Hubbard, Baker &Parsons 2010	69.96 (11.97)	m 14(56) f 11 (44)	time since stroke (days) mean 4.58 (2.93) IQR 2.75-6.0 affected side l 16(64) r 9 (36)	hospital	Australia	English	not stated	not stated
UL MAS	Miller, Slade, Pallant, Galea2010	67.4 (15.6) range28-90	male 46 (57.5%) female 34 (42.5%)	stroke type and location	rehabilit ation facilities	Australia	English	not stated	not stated
UL MAS	Carr, Shephard, Nordholm & Lynne1985	65 (range 55-78)	1 male 4 female 20%	average 14weeks post (range 6-40 weeks)	community	Australia	English	convenience	not stated
UL MAS	Sabari, Lim, Velozo, Lehman, Kieran & Lai 2005	54 (mean and median) range 18-94	67 males 33women	length of time post-ranged 3 days to 6.5 years,83% within 3 months	hospital-inpatientand outpatient OT	USA	English	consecutive referrals	not stated
Excluded at update									

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
10MWT	Baer & Smith 2001	71.8 SD 11.2 years	92 (49.7%) male 93 (50.3%) female	stroke classification	hospital	UK	English	consecutive	238 admitted 23 excluded secondary to intracerebral haemorrhage, 30 died prior to meeting milestone, 185 available for analysis
10MWT	Bower, McGinley,Miller, Clark 2014	68.3 (15.1)	male 21 (70)	co morbidities, median months post stroke 13.5 (5-45)	community	Australia	English	consecutive	30/65 recruited
10MWT	Smith & Baer 1999	mean 69.7 years SD 11.9	110 (48%) men 119 (52%) female	stroke type	hospital	UK	English	consecutive	238 records available for analysis 9 were omittedas missing data 17 diedthroughout but data was available so analysed
10MWT	Vernon, Paterson, Bower, McGinley, Miller, Pua & Clark 2015	68 (15)	21 (70)	time since stroke months (21 (19)	community	Australia	English	not stated	not given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
10MWT	Wolf, Catlin, Gage, Gurucharri, Robertson & Stephen 1999	without impairmentmean 56.43 SD13.82 range 34- 89 with stroke mean 56.04 SD 12.80 range 38-88	without 8 women 20 men with stroke not given	months post and side of stroke, assistive devices used	community	USA	English	convenience	not stated
10MWT - fast and comfortab l e	Combs, Dugan, Passmore, Reisner, Whipker, Yingling & Curtis 2010	61.1 (11.7)	31% men	3.8 (3.2) years post 44% right hemiparesis	community	USA	English	convenience	not given
9HPT	Alusi, Worthington, Glickman, Findley& Bain 2000	unclear which patients underwent9HPT testing	unclear which patients underwent 9HPTtesting	MS only	outpatient clinics	UK	English	not stated	not given
9HPT	Fisk, Brown, Sketris, Metz & Stadnyk 2005	51 years, SD 10 range 26-84 years	female n=14075%	MS type	outpatient clinics	Canada	English	random	not given
9HPT	Marrie & Goldman2011	42.2 SD: 8.1	35 (79.5%) women 9 (20.5%) men	MS type only	not stated	Canada	English	not stated	not stated
9HPT	Rossier & Wade 2002	combined grp 53.8 (10.3)	m/f 14/29	19.2 (10.8) years duration	community	UK	English	convenience (volunteer)	not given
9HPT	Sunderland, Tinson, Bradley &Hewer 1989	mean age 67, range31-82	17 men 21 women	not given, stroke diagnosis only - MCA 2 brain stem	hospital	UK	English	consecutive	not given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
ARAT	McDonnell, Hillier,Ridding & Miles 2006	each individual participant age stated so can calculate as required, range 45- 94	F: 8 M:9	lesion site on CT, time post,	not stated	Australia	English	not stated	not stated
AQoL	Turner-Stokes, Baguley, De Graaff, Katrak, Davies, McCrory & Hughes 2010	54.5 (13.2)	male : female 54 :36 60%	mean time post stroke 5.9 yrs (10.5)	community	Australia	English	consecutive	122 screened, 102 eligible 96 consented
ARAT	Barden, Baguley,Nott, Chapparo 2014	51 (17)	15/13 M/F	years post median and range 2.5 (0.5-39)	outpatient	Australia	English	not stated	not given
ARAT	Barden, Nott, Heard, Chapparo& Baguley 2012	median and range:50 (18-81)	M/F 22/16	hand dominance, diagnosis (strokeor TBI), years after injury/event	outpatient spasticity clinic	Australia	English	consecutive	not stated
ARAT	Barreca, Stratford, Lambert, Masters & Streiner 2005	acute group = 71.4 (50.9 - 90.0) chronic group = 64.0 (44.7-76.6)	F = 19 M=20	days post, affected side, ICD-9-CM code, depression present or absent and neglect, hemianopsia and proprioception present or absent.	Inpatient / outpatient rehabilitation facilities	Canada	English	not stated	not stated

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
ARAT	Barreca, Stratford, Masters, Lambert, Griffiths 2006	age quartile of 2 groups given; mild-mod impairment: 66, 76, 81 severe impairment 59, 69, 77	F/M: 51/54	days post, hand dominance, affected side, infarct or haemorrhage or missing and ICD-9 CM classification	inpatient and outpatient rehabilitation	Canada	English	not stated	not stated
ARAT	Barreca, Stratford, Masters, Lambert, Griffiths & McBay 2006	median age 71yrs (1st, 3rd quartiles 51, 90) for acute group, 64 yrs (1st, 3rd quartiles 45, 77) for chronic group	total women 19 (12 in acute grp 7in chronic grp)	time post stroke	inpatient rehabilitation	Canada	English	not stated howpatients were selected	not stated and cannotbe deduced
ARAT	Barreca, Stratford, Masters, Lambert, Griffiths & McBay 2006	median age 71yrs (1st, 3rd quartiles 51, 90) for acute group, 64 yrs (1st, 3rd quartiles 45, 77)for chronic group	total women 19 (12 in acute grp 7in chronic grp)	time post stroke	inpatient rehabilit ation	Canada	English	not stated howpatients were selected	not stated and cannotbe deduced
ARAT	Blennerhass ett,Avery & Carey 2010	median (IQR) = 63 (50-69) range 23-80	male: female frequency 17:5	days post, type of stroke, side affected, dominance	inpatient rehabilit ation	Australia	English	convenience	not stated
ARAT	Celik, O'Malley, Boake, Levin, Yozbatiran & Reistetter 2010	each of n=9 ages given, range 48-67	M =7 F = 2	side affected, months post, stroke location	assumed community based	USA	English	those exhibiting under use were selected	not stated

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
ARAT	Edwards, Lang, Wagner, Birkenmeier & Dromerick 2012	63.7 +/-13.6	men 21 (42) women 29 (58)	admission NIHSS score, time post, stroke type, affected side	inpatient rehabilit ation hospital	USA	English	not stated butprior study with such details referenced	not stated
ARAT	Notley, Turk, Pickering, Simpson & Burridge 2007	63 (13.8)	male: 6 female: 4	side assessed, time post stroke	hospit al outpati ents	UK	English	? Review of database list and eligible were invited toparticipate	yes 21 invited, 15 accepted, 3 excluded following screening, further 2 later excluded.
ARAT	O'Dell, Kim, Rivera, Fieo, Christos, Polistena, Fitzgerald & Gorga 2013	mean age 56 yrs,SD 12.4, median 57, range 35-85	M: 72% of 32 participants	ethnicity, time post stroke 4.1 yrs (4.5) range 0.8-25.2), type of and lesion location	outpatie nt rehabilit ation	USA	English	convenience -participants volunteered recruited fromflyers, outpatients clinics, support groups - part of larger trial.	not stated
ARAT	Page, Hade & Persch 2015	56.6 (10.1) range38-75	15 male 17female	mean time post stroke 4.6 years (5.8)	community	USA	English	convenience	not given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
ARAT	Page, Levine &Hade 2012	mean +/- SD = 60.8 +/- 12.3 years agerange 21-76years	23 men 26 females	time post stroke, ischaemic	outpatient rehabilitation clinics	USA	English	not stated butreference provided for article with such details (this is a secondary analysis of preinterventi on scores fromanother trial)	not stated
ARAT	Stinear, Barber, Petoe, Anwar & Byblow 2012	median 70 range 31-91	16 males 24females	vascular risk factors, stroke characteristics,	hospital and community	New Zealand	English	prospective	not stated
ARAT	Urbin, Waddell & Lang 2015	56 (10.4), 62 (9.4)	6 (75), 20 (74)		inpatient and community	USA	English	not stated	not given
BI	Duncan, Lai, Tyler, Perera, Reker & Studenski2002	pt 72.6 (10.0) proxy 59.8 (15.5)	pt 135 males (47%) 152 females, proxy males 78(27.2% females 209 (72.8%)	education, race, stroke type	community	USA	English	consecutive	20%
BI	Khan, Pallant & Turner-Stokes 2008	52.0 (8.3) range 37-62	10M/14F	mean time post diagnosis 11.1 (5) years range 2-23 years	inpatient rehabilit ation	Australia	English	consecuti ve admission s	not stated

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
BI	Lai, Studenski, Duncan & Perera2002	76.0 (6.56)	male 48 (59.4%)		community	USA	English	convenience	not given
BI	Maujean, Davis, Kendall, Casey &Loxton 2014	62.77 (11.24)	40 (50)	> 7 years post, no cognitive impairment	community	Australia	English	convenience	27% of 1300 sent questionnaires returned
BI	Sprigg, Selby, Fox, Berge, Whynes, Philip &Bath 2013	70 (12.11)	1474 (58)	acute	community	multi site international	multi	consecutive	11.8% died
BI	Wolf & Koster 2013	grp 1 64.2 (13.4) grp 2 60.5(12.8)	grp 1 males 28 grp 2 31		community	USA	English	convenience	not given
BI	Ashford, Turner-Stokes, Siegert &Slade 2013	47 (17.5)	32 (55)		outpatient spasticity clinic	UK	English	convenience purposeful	34 from 40
CMA	Barreca, Stratford, Lambert, Masters & Streiner 2005	acute group = 71.4 (50.9 - 90.0) chronic group = 64.0 (44.7-76.6)	F = 19 M=20	days post, affected side, ICD-9-CM code, depression present or absent and neglect, hemianopsia and proprioception present or absent.	inpatient/outpatient rehabilitation facilities	Canada	English	not stated	not stated

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
CMA	Coderre, Zeid, Dukelow, Demmer, Moore, Demers, Bretzke,Herter, Glasgow, Norman, Bagg & Scott 2010	left affected median63 (min max 22, 90) right affected 66 (29,92)	left affected 13/13right affected male 14 f 12	handedness, type of stroke and location dayspost left 31 right 31	hospital	Canada	English	consecutive	not given
CMA	Ellis, Sukal, DeMott & Dewald2008	51-78	m-8 f =3	affected side, dominance, lesion location	community	USA	English	convenience	not given
CMA	Huijbregts, Gowland & Gruber2000	stroke survivors atleast 18 years no other info provided	stroke survivorsmale - 19 (56%) female 15 (44%) nil sex details for caregivers	44% right hemi, co morbidities listed	acute through to community	Canada	English	taken from another study assumed consecutiv e	32 from original 66
CMA	Levin, Desrosiers, Beauchemin, Bergeron & Rochette 2004	54.9 (18.6) range20-80	14 men 14 women	time since stroke mean 16 months SD 18.6 range 1-70, 12 right side stroke,	outpatients	Canada	English	convenience	not given
CMA	Manns, Tomczak, Jelani, Cress & Haennel 2009	54 (11)	4 men 6 women	time since - 7.5 years (8.3)	community	Canada	English	convenience	not given
CMA	Oczkowski & Barreca 1993	65.7 women average 65.8 men	54 women 59 men	side of hemiparesis	inpatient rehabilitation	Canada	English	consecutive	not given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
CMSA	Semrau, Herter, Scott, Dukelow 2015	not given	not given		hospital and community	Canada	English	not stated	not given
DAS	Brashear, Gordon,Elovic, Kassicieh, Marciniak, Do, Lee, Jenkins, Turkel, 2002	placebo 62 (range 23-87) Botox group 61 (23-88)	placebo 27 (44%) Botox 36 (56%)	duration ofspasticity	community assumed	USA	English	not stated	not stated
EQ-5D	Ali, Fulton, Quinn,Brady on behalf ofVISTA 2013	71 (60-78)	male 2715 (54.9) of whole sample	other medical conditions	assumed hospital	international repository	English	?	not given
EQ-5D	Jenkinson, Fitzpatrick, Crocker & Peters2013	4 (2.7%) 18-44, 45 (30.4%) 45-64 41 (27.1%) 65-74 58 (39.2%) >75	56 (37.1%) women 88 (58.3%) men	mean time post stroke 7.3 years (6.1)	community	UK	English	convenience -recruited from database	525 eligible, 418 sent survey
EQ-5D	Kim, Jo & Lee 2015	68.3 (8.1)	287 (59)	nil	community	Korea	mixed	voluntary	not given
EQ-5D	Lunde 2013	68.74 (12.93)	not given		community	Norway	assumed Norwegian	consecutive	not stated
EQ-5D	Mitosek-Szewczyket al 2014	40.7 (11.7)	1020 (29)	majority had relapsing remitting form,	community	Poland	polish	convenience	not stated
EQ-5D	Sprigg, Selby, Fox, Berge, Whynes, Philip &Bath 2013	70 (12.11)	1474 (58)	acute	community	multi site international	multi	consecutive	11.8% died

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
FIM	Cusick, Lannin, Hanssen & Allaous 2014	28 (12)	27 (82)	acute - slow to recover, severe TBI	hospital	Australia	English	consecutive	yes 33/37
FIM	Joseph, Pandit,Aziz et al 2013	54.9 (21.8)	110 (69)	TBI	hospital	USA	English	consecutive	not given
FIM	McNett, Amato, Gianakis, Grimm, Philippbar, Belle &Moran 2014	53.1 (21.40)	10 (75)		hospital	USA	English	consecutive	not given
FIM	Perrin, Niemeier, Mougeot et al2015	39.90 (18.48)	67 (67)	evenly mixed severity - mild, mod and severe	hospital	USA	English	consecutive	100/133
FIM	Semrau, Herter, Scott, Dukelow 2015	not given	not given		hospital and commu nity	Canada	English	not stated	not given
FIM	Tyryshkin, Coderre, Glasgow,Herter, Bagg, Dukelow & Scott 2014	L) affected median61 (20, 89), R) 62 (21, 86)	L) M/F 62/29 R)38/25		hospital	Canada	English	not stated	not given
FIM	Chen, Chen, Hreha, Goedert & Barrett 2015	median (IQR) 70 (61-81)	53 (44)	hospital	rehabilitation	USA	English	consecutive	not given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
FIM	Wu, Burgard &Radel 2014	no apraxia 52 (8.80) with apraxia 66 (8.11)	total 10/15 (67)	left hemispheric stroke only	inpatient rehabilit ation	USA	English	consecutive	32 excluded
mFrencha y Arm Test	Sunderland, Tinson, Bradley &Hewer 1989	mean age 67, range31-82	17 men 21 women	not given, stroke diagnosis only - MCA 2 brain stem	hospital	UK	English	consecutive	not given
GAS	Barden, Baguley,Nott, Chapparo 2014	51 (17)	15/13 M/F	years post median and range 2.5 (0.5-39)	outpatient	Australia	English	not stated	not given
GAS	Barden, Baguley, Nott, Chapparo 2014 (b)	51	13 female	greater than 7 years post event	community	Australia	English	not stated	not given
GAS	Malec 2001	34.2 (12.2)	m 73%	time since mean 4.6, SD6.6 median 1.5, typeof injury stroke 19% TBI 72% other 9		USA	English	consecutive	not given
LASIS	Turner-Stokes, Baguley, De Graaff, Katrak, Davies, McCrory & Hughes 2010	54.5 (13.2)	male : female 54 :36 60%	mean time post stroke 5.9 yrs (10.5)	community	Australia	English	consecutive	122 screened, 102 eligible 96 consented

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
LASIS	Ashford, Slade, Nair & Turner-Stokes 2014	47 (17.5)	32 (55)	high level of disability and dependence	outpatient spasticity clinic	UK	English	consecutive	103 screened 58 included
LASIS	Ashford, Turner-Stokes, Siegert &Slade 2013	44.5 (16.7)	54 (59)	mixed level of function	outpatient spasticity clinic	UK	English	consecutive and part convenience	92 from 103
MAL	Atler, Malcolm & Griefe 2015	65.08 (10.95) range 41-85	7 male (58.3)	chronicity mean years post 3.41	community	USA	English	convenience	not given
MAL	Borstad & Nichols-Larsen 2016	mean 64	7 female 5 male	lesion location, chronicity	community	USA	English	convenience	not given
MAL	Celik, O'Malley, Boake, Levin, Yozbatiran & Reistetter 2010	each of n=9 ages given, range 48-67	M =7 F = 2	side affected, months post, stroke location	assumed community based	USA	English	those exhibiting under use were selected	not stated
MAL	Mark, Woods, Mennemeier, Abbas & Taub2006	59.6 (20.6)	9 men 6 women	7 right, 8 left hemi. 14.6, education: 14.6 (2.0) years, NIHS stroke score 2.6 (1.7), MMSE 27.5 (3.5) charlson Comorbidity index 3.7 (1.3)	rehabilitation clinic	USA	English	convenience - volunteers	43 potential, 15 UE 14 LE

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
MI	Kopp, Kunkel, Flor, Platz, Rose, Mauritz, Gresser, McCulloch & Taub 1997	66 years	12/33 were women	nil other details other than stroke given	inpatient rehabilitation clinic	no stated	not stated	consecutive	not stated
mMAS	Horgan, Cunningham, Coakley, Walsh, O'Regan & Finn 2006	mean and SD 77.8(7.5 years)	19 (46% male)	side of stroke, living situation	rehabilitation	Ireland	English	convenience	0%
Motricit yIndex	Smith-Arena, Edelstein, Rabadi,2006	total pop 71 +/-9.8	male: female 29:10	stroke type LOS MMSE stroke severity, FIM,	inpatient and community	USA	English	consecutive	381 admitted 45 participated in driving evaluation at their institute, n=39 (6 excluded with missingdata)
Motricit yIndex	Sunderland, Tinson, Bradley &Hewer 1989	mean age 67, range31-82	17 men 21 women	not given, stroke diagnosis only - MCA 2 brain stem	hospital	UK	English	consecutive	not given
OHS	Dennis, Wellwood& Warlow 1996	73.1 mean	68 (44%) men	nil info given	community	UK	English	consecutive	266 admitted 152 recruited 2 missing data

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
OHS	Gubitz, Reidy, Christian & Phillips2012	not available forwhole sample	unavailable forwhole sample		hospital and community	Canada	English	consecutive	not stated
OHS	Herrman, Black, Lawrence, Szekely& Szalai 1998	74.9 +/-11.6	51% female	history and type of stroke	acute	Canada	English	consecutive	436 with 450 meeting hemispheric stroke criteria, of those n= 150 at 3 months and n=133 at 1 year
OHS	Pittock, Meldrum, Dhuill, Hardiman & Moroney 2003	2 weeks 76 (20) 6 mo 77 (16) 2 yrs 75.4 (15.2)	not given	stroke type	acute	Ireland	English	consecutive	117/126 then following loss over time periods
OHS	SCOPE Collaborations 7 IST 2007	30 day grp given not 6 mo when OHScompleted; 74 (12.2)	291 (54%) males	time since stroke (hr)	acute and community followup	international		consecutive	not stated
RMA	Pittock, Meldrum, Dhuill, Hardiman & Moroney 2003	2 weeks 76 (20) 6 mo 77 (16) 2 yrs 75.4 (15.2)	not given			Ireland	English	consecutive	117/126 then following loss over time periods
RMA	Taylor, Ashburn &Ward 1994	mean 72 range 49-86	20 women 18men	17 right hemi 21 left hemi	hospital	UK	English	consecutive	73 referred, 42 eligible, 31 excludedfor many reasons and4 did not complete - 3 died 1 withdrew

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
RMA	Adams, Pickering& Taylor 1997 (acute)	74.37 (9.38) 49-88, 42% > 65yrs	27 women 24men	62.7% left hemiplegia	acute hospital	UK	English	consecutive	16% (26) of the 173 were eligiblefrom recruitment site 1 second site rate not stated
SF-36	Rudick, Miller, Hass et al 2007	range 18 - 50	591 (28)	MS - relapsing	community	USA	English	not stated	not stated
SIS	Ali, Fulton, Quinn,Brady on behalf ofVISTA 2013	71 (60-78)	male 2715 (54.9) of whole sample	other medical conditions	assumed hospital	international repository	English	?	not given
SIS	O'Dell, Kim, Rivera, Fieo, Christos, Polistena, Fitzgerald & Gorga 2013	mean age 56 yrs,SD 12.4, median 57, range 35-85	M: 72% of 32 participants	ethnicity, time post stroke 4.1 yrs (4.5) range 0.8-25.2), type of and lesion location	outpatient rehabilitation	USA	English	convenience -participants volunteered recruited fromflyers, outpatients clinics, support groups - part of larger trial.	not stated
SIS	Boger, Hankins & Latter 2015	57.99 (14.66) range 27-89	male 37 (50) female 37 (50)	51 mo post stroke (mean)	community	UK	English	convenience	not given
UL-RMA	Meldrum, Pittock, Hardiman, Dhuill,O'Regan & Moroney 2004	69 (12.6)	68 male (60%) female 16 (40%)	stroke classification, therapies received, mortality	acute to community	Ireland	English	consecutive	not given

Generalisability									
Measure	**Author**	**Median or mean age (with standard deviation or range)**	**Distribution of sex**	**Important disease characteristics**	**Setting(s) in which the study was conducted**	**Countries in which the study was**	**Language in which the HR-PRO**	**Method usedto select patients**	**Percentage of missing responses (response**
UL-RMA	Morris, van Wijck Joice & Donaghy2013	median 69 range 36-88	male 49 female 36	type and duration post stroke	recruited within acute presentation but followed up and assessed in community	unknown	not stated	consecutive admission	given

Appendix G: Supplementary material from Study 4

- Participant Demographic Information
- GAS-Light tool
- Goal Attainment Scaling (GAS) and Goal Attainment Scaling – Light (GAS-Light) training package outline
- SPSS outputs
- Observed power analysis

Participant Demographic Information

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Date: / /

Participant Number: ☐☐☐☐

DEMOGRAPHICS

Date of Birth:		Age:	Gender:	
Marital Status	☐ Never Married	☐ Widowed	☐ Divorced	☐ Separated
	☐ Married	☐ Other – De facto	☐ Other – single/not married	
Indigenous Status	☐ Aboriginal Origin		☐ Torres Strait Island Origin	
	☐ Aboriginal and Torres Strait Islander Origin		☐ Not Indigenous Australian	
Language Spoken				
Education *How much school have you completed?*	☐ Less than high school	☐ Finished high school	☐ TAFE or technical college	
	☐ University	☐ Postgraduate school or degree		
Employment Status	☐ Full-time	☐ Part-time	☐ Casual	☐ Unemployed
	☐ Retired	☐ Other (e.g. student, volunteer/unpaid work)		
Current Living Situation	☐ Hospital		☐ Home with support	
	☐ Rehabilitation (Inpatient)		☐ Home without support	
	☐ Low level Residential Care		☐ Transitional care services	
	☐ High level Residential Care		Other:	
Do you live on your own?	☐ Yes, I live entirely on my own		☐ No, I live with others	
Involved Community Services	☐ Domiciliary Care		☐ DVA	
	☐ Meals on wheels		☐ COP	
	☐ RDNS		☐ Council	
	☐ CACP		☐ Other:	

Clinical Information

Diagnosis:

Date of diagnosis/condition onset: ☐☐/☐☐/☐☐☐☐ (dd/mm/yyyy) Estimated date: ☐

Hemiplegic/Affected Side: ☐ Right ☐ Left

Motor Function:

Upper Limb	
	No weakness
	Mild, able to move against gravity but incomplete resistance to force
	Significant, cannot completely overcome gravity in range of motion
	Total paralysis, absence of motion or contraction of muscles without joint movement

Lower Limb	
	No weakness
	Mild, able to move against gravity but incomplete resistance to force
	Significant, cannot completely overcome gravity in range of motion
	Total paralysis, absence of motion or contraction of muscles without joint movement

Transfer Ability: ☐ Independent ☐ Supervision ☐ Standby Assist ☐ 1 Person Assist ☐ 2 Person Assist

Ambulation: ☐ Walking – independent (nil aides) ☐ Walking – independent (with aides)
☐ Walking- supervision (with or without aides) ☐ Wheelchair

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Date: / /

Participant Number: ☐☐☐☐

NIH STROKE SCALE

Interval: [ ] Baseline [ ] 2 hours post treatment [ ] 24 hours post onset of symptoms ±20 minutes [ ] 7-10 days
[ ] 3 months [ ] Other _____(___ ___)

Time: ___ ___:___ ___ [ ]am [ ]pm

Person Administering Scale _____

Administer stroke scale items in the order listed. Record performance in each category after each subscale exam. Do not go back and change scores. Follow directions provided for each exam technique. Scores should reflect what the patient does, not what the clinician thinks the patient can do. The clinician should record answers while administering the exam and work quickly. Except where indicated, the patient should not be coached (i.e., repeated requests to patient to make a special effort).

Instructions	Scale Definition	Score
1a. Level of Consciousness: The investigator must choose a response if a full evaluation is prevented by such obstacles as an endotracheal tube, language barrier, orotracheal trauma/bandages. A 3 is scored only if the patient makes no movement (other than reflexive posturing) in response to noxious stimulation.	0 = **Alert;** keenly responsive. 1 = **Not alert;** but arousable by minor stimulation to obey, answer, or respond. 2 = **Not alert;** requires repeated stimulation to attend, or is obtunded and requires strong or painful stimulation to make movements (not stereotyped). 3 = Responds only with reflex motor or autonomic effects or totally unresponsive, flaccid, and areflexic.	____
1b. LOC Questions: The patient is asked the month and his/her age. The answer must be correct - there is no partial credit for being close. Aphasic and stuporous patients who do not comprehend the questions will score 2. Patients unable to speak because of endotracheal intubation, orotracheal trauma, severe dysarthria from any cause, language barrier, or any other problem not secondary to aphasia are given a 1. It is important that only the initial answer be graded and that the examiner not "help" the patient with verbal or non-verbal cues.	0 = **Answers** both questions correctly. 1 = **Answers** one question correctly. 2 = **Answers** neither question correctly.	____
1c. LOC Commands: The patient is asked to open and close the eyes and then to grip and release the non-paretic hand. Substitute another one step command if the hands cannot be used. Credit is given if an unequivocal attempt is made but not completed due to weakness. If the patient does not respond to command, the task should be demonstrated to him or her (pantomime), and the result scored (i.e., follows none, one or two commands). Patients with trauma, amputation, or other physical impediments should be given suitable one-step commands. Only the first attempt is scored.	0 = **Performs** both tasks correctly. 1 = **Performs** one task correctly. 2 = **Performs** neither task correctly.	____
2. Best Gaze: Only horizontal eye movements will be tested. Voluntary or reflexive (oculocephalic) eye movements will be scored, but caloric testing is not done. If the patient has a conjugate deviation of the eyes that can be overcome by voluntary or reflexive activity, the score will be 1. If a patient has an isolated peripheral nerve paresis (CN III, IV or VI), score a 1. Gaze is testable in all aphasic patients. Patients with ocular trauma, bandages, pre-existing blindness, or other disorder of visual acuity or fields should be tested with reflexive movements, and a choice made by the investigator. Establishing eye contact and then moving about the patient from side to side will occasionally clarify the presence of a partial gaze palsy.	0 = **Normal.** 1 = **Partial gaze palsy;** gaze is abnormal in one or both eyes, but forced deviation or total gaze paresis is not present. 2 = **Forced deviation,** or total gaze paresis not overcome by the oculocephalic maneuver.	____

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Date: / /

Participant Number: | | | | |

NIH STROKE SCALE

Interval: [ ] Baseline [ ] 2 hours post treatment [ ] 24 hours post onset of symptoms ±20 minutes [ ] 7-10 days
[ ] 3 months [ ] Other _____(___ ___)

3. Visual: Visual fields (upper and lower quadrants) are tested by confrontation, using finger counting or visual threat, as appropriate. Patients may be encouraged, but if they look at the side of the moving fingers appropriately, this can be scored as normal. If there is unilateral blindness or enucleation, visual fields in the remaining eye are scored. Score 1 only if a clear-cut asymmetry, including quadrantanopia, is found. If patient is blind from any cause, score 3. Double simultaneous stimulation is performed at this point. If there is extinction, patient receives a 1, and the results are used to respond to item 11.	0 = **No visual loss.** 1 = **Partial hemianopia.** 2 = **Complete hemianopia.** 3 = **Bilateral hemianopia** (blind including cortical blindness).	____
4. Facial Palsy: Ask – or use pantomime to encourage – the patient to show teeth or raise eyebrows and close eyes. Score symmetry of grimace in response to noxious stimuli in the poorly responsive or non-comprehending patient. If facial trauma/bandages, orotracheal tube, tape or other physical barriers obscure the face, these should be removed to the extent possible.	0 = **Normal** symmetrical movements. 1 = **Minor paralysis** (flattened nasolabial fold, asymmetry on smiling). 2 = **Partial paralysis** (total or near-total paralysis of lower face). 3 = **Complete paralysis** of one or both sides (absence of facial movement in the upper and lower face).	____
5. Motor Arm: The limb is placed in the appropriate position: extend the arms (palms down) 90 degrees (if sitting) or 45 degrees (if supine). Drift is scored if the arm falls before 10 seconds. The aphasic patient is encouraged using urgency in the voice and pantomime, but not noxious stimulation. Each limb is tested in turn, beginning with the non-paretic arm. Only in the case of amputation or joint fusion at the shoulder, the examiner should record the score as untestable (UN), and clearly write the explanation for this choice.	0 = **No drift;** limb holds 90 (or 45) degrees for full 10 seconds. 1 = **Drift;** limb holds 90 (or 45) degrees, but drifts down before full 10 seconds; does not hit bed or other support. 2 = **Some effort against gravity;** limb cannot get to or maintain (if cued) 90 (or 45) degrees, drifts down to bed, but has some effort against gravity. 3 = **No effort against gravity;** limb falls. 4 = **No movement.** UN = **Amputation** or joint fusion, explain: _____ **5a. Left Arm** **5b. Right Arm**	____ ____
6. Motor Leg: The limb is placed in the appropriate position: hold the leg at 30 degrees (always tested supine). Drift is scored if the leg falls before 5 seconds. The aphasic patient is encouraged using urgency in the voice and pantomime, but not noxious stimulation. Each limb is tested in turn, beginning with the non-paretic leg. Only in the case of amputation or joint fusion at the hip, the examiner should record the score as untestable (UN), and clearly write the explanation for this choice.	0 = **No drift;** leg holds 30-degree position for full 5 seconds. 1 = **Drift;** leg falls by the end of the 5-second period but does not hit bed. 2 = **Some effort against gravity;** leg falls to bed by 5 seconds, but has some effort against gravity. 3 = **No effort against gravity;** leg falls to bed immediately. 4 = **No movement.** UN = **Amputation** or joint fusion, explain: _____ **6a. Left Leg** **6b. Right Leg**	____

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Date: / /

Participant Number: | | | | |

NIH STROKE SCALE

Interval: [ ] Baseline [ ] 2 hours post treatment [ ] 24 hours post onset of symptoms ±20 minutes [ ] 7-10 days
[ ] 3 months [ ] Other _____(___ ___)

7. **Limb Ataxia:** This item is aimed at finding evidence of a unilateral cerebellar lesion. Test with eyes open. In case of visual defect, ensure testing is done in intact visual field. The finger-nose-finger and heel-shin tests are performed on both sides, and ataxia is scored only if present out of proportion to weakness. Ataxia is absent in the patient who cannot understand or is paralyzed. Only in the case of amputation or joint fusion, the examiner should record the score as untestable (UN), and clearly write the explanation for this choice. In case of blindness, test by having the patient touch nose from extended arm position.	0 = **Absent.** 1 = **Present in one limb.** 2 = **Present in two limbs.** UN = **Amputation** or joint fusion, explain: _____	____
8. **Sensory:** Sensation or grimace to pinprick when tested, or withdrawal from noxious stimulus in the obtunded or aphasic patient. Only sensory loss attributed to stroke is scored as abnormal and the examiner should test as many body areas (arms [not hands], legs, trunk, face) as needed to accurately check for hemisensory loss. A score of 2, "severe or total sensory loss," should only be given when a severe or total loss of sensation can be clearly demonstrated. Stuporous and aphasic patients will, therefore, probably score 1 or 0. The patient with brainstem stroke who has bilateral loss of sensation is scored 2. If the patient does not respond and is quadriplegic, score 2. Patients in a coma (item 1a=3) are automatically given a 2 on this item.	0 = **Normal;** no sensory loss. 1 = **Mild-to-moderate sensory loss;** patient feels pinprick is less sharp or is dull on the affected side; or there is a loss of superficial pain with pinprick, but patient is aware of being touched. 2 = **Severe to total sensory loss;** patient is not aware of being touched in the face, arm, and leg.	____
9. **Best Language:** A great deal of information about comprehension will be obtained during the preceding sections of the examination. For this scale item, the patient is asked to describe what is happening in the attached picture, to name the items on the attached naming sheet and to read from the attached list of sentences. Comprehension is judged from responses here, as well as to all of the commands in the preceding general neurological exam. If visual loss interferes with the tests, ask the patient to identify objects placed in the hand, repeat, and produce speech. The intubated patient should be asked to write. The patient in a coma (item 1a=3) will automatically score 3 on this item. The examiner must choose a score for the patient with stupor or limited cooperation, but a score of 3 should be used only if the patient is mute and follows no one-step commands.	0 = **No aphasia;** normal. 1 = **Mild-to-moderate aphasia;** some obvious loss of fluency or facility of comprehension, without significant limitation on ideas expressed or form of expression. Reduction of speech and/or comprehension, however, makes conversation about provided materials difficult or impossible. For example, in conversation about provided materials, examiner can identify picture or naming card content from patient's response. 2 = **Severe aphasia;** all communication is through fragmentary expression; great need for inference, questioning, and guessing by the listener. Range of information that can be exchanged is limited; listener carries burden of communication. Examiner cannot identify materials provided from patient response. 3 = **Mute, global aphasia;** no usable speech or auditory comprehension.	____
10. **Dysarthria:** If patient is thought to be normal, an adequate sample of speech must be obtained by asking patient to read or repeat words from the attached list. If the patient has severe aphasia, the clarity of articulation of spontaneous speech can be rated. Only if the patient is intubated or has other physical barriers to producing speech, the examiner should record the score as untestable (UN), and clearly write an explanation for this choice. Do not tell the patient why he or she is being tested.	0 = **Normal.** 1 = **Mild-to-moderate dysarthria;** patient slurs at least some words and, at worst, can be understood with some difficulty. 2 = **Severe dysarthria;** patient's speech is so slurred as to be unintelligible in the absence of or out of proportion to any dysphasia, or is mute/anarthric. UN = **Intubated** or other physical barrier, explain:_____	____

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Date: / /

Participant Number: | | | | |

NIH STROKE SCALE

Interval: [ ] Baseline [ ] 2 hours post treatment [ ] 24 hours post onset of symptoms ±20 minutes [ ] 7-10 days [ ] 3 months [ ] Other _____(___ ___)

11. Extinction and Inattention (formerly Neglect): Sufficient information to identify neglect may be obtained during the prior testing. If the patient has a severe visual loss preventing visual double simultaneous stimulation, and the cutaneous stimuli are normal, the score is normal. If the patient has aphasia but does appear to attend to both sides, the score is normal. The presence of visual spatial neglect or anosagnosia may also be taken as evidence of abnormality. Since the abnormality is scored only if present, the item is never untestable.	**0 = No abnormality.** **1 = Visual, tactile, auditory, spatial, or personal inattention or extinction to bilateral simultaneous stimulation in one of the sensory modalities.** **2 = Profound hemi-inattention or extinction to more than one modality; does not recognize own hand or orients to only one side of space.**	____

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

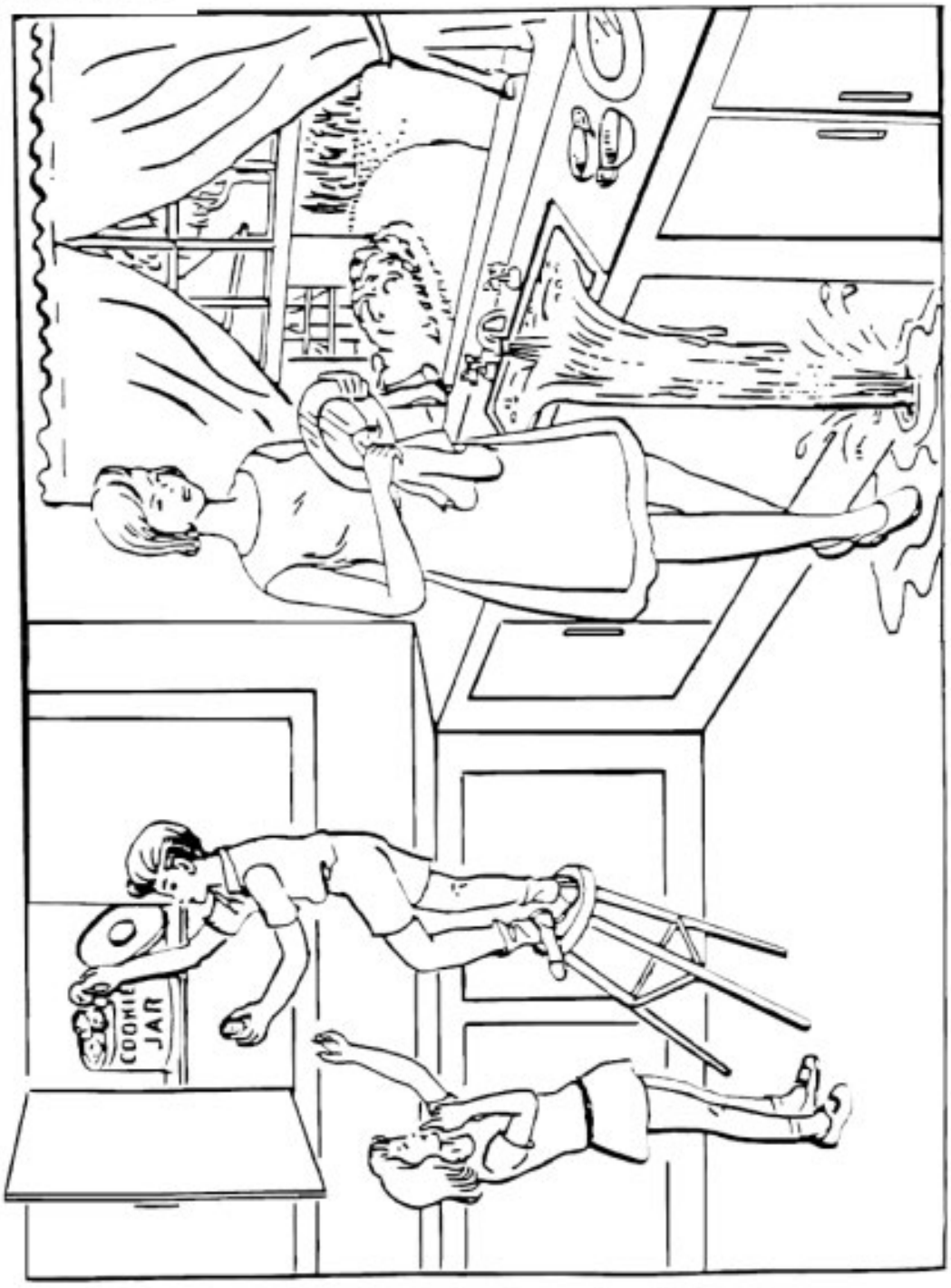
Date: / /

Participant Number: | | | | |

NIH STROKE SCALE

Question 9 Stimulus Page:

The patient is asked to describe what is happening in the picture



University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Date: / /

Participant Number: | | | | |

NIH STROKE SCALE

Question 9 Stimulus Page

The patient is asked read from this sentence list

You know how.

Down to earth.

I got home from work.

Near the table in the dining room.

They heard him speak on the radio last night.

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Date: / /

Participant Number: | | | | |

NIH STROKE SCALE

Question 9 Stimulus Page:
The patient is asked to name the items in the picture

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Date: / /

Participant Number: | | | | |

NIH STROKE SCALE

Question 9 Stimulus Page

The patient is asked read from this word list.

MAMA

TIP – TOP

FIFTY – FIFTY

THANKS

HUCKLEBERRY

BASEBALL PLAYER

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Date: / /

Participant Number: | | | | |

OXFORDSHIRE COMMUNITY STROKE PROJECT CLASSIFICATION

TAC — Total Anterior Circulation Stroke

LAC — Lacunar Stroke

PAC — Partial Anterior Circulation Stroke

POC — Posterior Circulation Stroke

Code last letter as follows:

(S) — Syndrome: Indeterminate pathogenesis, prior to imaging (e.g., TACS)

(I) — Infarct (e.g., TACI)

(H) — Hemorrhage (e.g., TACH)

FINAL CLASSIFICATION: ___ ___ ___ ___

References

Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. "Classification and natural history of clinically identifiable subtypes of cerebral infarction." **Lancet 22;337(8756):1521-6, 1991**

"How well does the Oxfordshire Community Stroke Project classification predict the site and size of the infarct on brain imaging?" ***J Neurol Neurosurg Psychiatry* 2000;68:558-562**

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

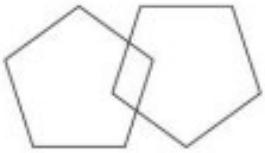
Date: / /

Participant Number: | | | | |

The Mini-Mental State Exam

Patient_____ Examiner _____ Date_____

Maximum	Score	
		Orientation
5	( )	What is the (year) (season) (date) (day) (month)?
5	( )	Where are we (state) (country) (town) (hospital) (floor)?
		Registration
3	( )	Name 3 objects: 1 second to say each. Then ask the patient all 3 after you have said them. Give 1 point for each correct answer. Then repeat them until he/she learns all 3. Count trials and record. Trials _____
		Attention and Calculation
5	( )	Serial 7's. 1 point for each correct answer. Stop after 5 answers. Alternatively spell "world" backward.
		Recall
3	( )	Ask for the 3 objects repeated above. Give 1 point for each correct answer.
		Language
2	( )	Name a pencil and watch.
1	( )	Repeat the following "No ifs, ands, or buts"
3	( )	Follow a 3-stage command: "Take a paper in your hand, fold it in half, and put it on the floor."
1	( )	Read and obey the following: CLOSE YOUR EYES
1	( )	Write a sentence.
1	( )	Copy the design shown.



____ Total Score

ASSESS level of consciousness along a continuum _____

Alert Drowsy Stupor Coma

GAS-Light tool

University of Wollongong | LA TROBE UNIVERSITY | NSW GOVERNMENT Health Murrumbidgee Local Health District

GAS-Light Record Sheet

Participant Number: | | | | |

Baseline date:						**Review date:**		
	Patient stated goal	**SMART goal**	**Imp**	**Diff**	**Baseline**	**Achieved**		**Variance** (Describe achievement if differs from expected and give reasons)
1.			**0** **1** **2** **3**	**0** **1** **2** **3**	❒ **Some** function ❒ **None** (as bad as can be)	❒ Yes	❒ Much better ❒ A little better ❒ **As expected**	
						❒ No	❒ Partially achieved ❒ Same as baseline ❒ Worse	
2.			**0** **1** **2** **3**	**0** **1** **2** **3**	❒ **Some** function ❒ **None** (as bad as can be)	❒ Yes	❒ Much better ❒ A little better ❒ **As expected**	
						❒ No	❒ Partially achieved ❒ Same as baseline ❒ Worse	
3.			**0** **1** **2** **3**	**0** **1** **2** **3**	❒ **Some** function ❒ **None** (as bad as can be)	❒ Yes	❒ Much better ❒ A little better ❒ **As expected**	
						❒ No	❒ Partially achieved ❒ Same as baseline ❒ Worse	

GAS-Light Record Sheet, Version 2, Date:22.05.2012
Turner-Stokes L., Goal Attainment Scaling (GAS) in Rehabilitation: The GAS-Light model, 2012, Kings College London.

Page 1 of 2

University of Wollongong | LA TROBE UNIVERSITY | NSW GOVERNMENT Health Murrumbidgee Local Health District

GAS-Light Record Sheet continued

Participant Number: | | | | |

Baseline date:						**Review date:**		
	Patient stated goal	**SMART goal**	**Imp**	**Diff**	**Baseline**	**Achieved**		**Variance** (Describe achievement if differs from expected and give reasons)
4.			**0** **1** **2** **3**	**0** **1** **2** **3**	❒ **Some** function ❒ **None** (as bad as can be)	❒ Yes	❒ Much better ❒ A little better ❒ **As expected**	
						❒ No	❒ Partially achieved ❒ Same as baseline ❒ Worse	
5.			**0** **1** **2** **3**	**0** **1** **2** **3**	❒ **Some** function ❒ **None** (as bad as can be)	❒ Yes	❒ Much better ❒ A little better ❒ **As expected**	
						❒ No	❒ Partially achieved ❒ Same as baseline ❒ Worse	

Summary

Baseline GAS T-score:	**Achieved GAS T-score:**	**Change in GAS T Score:**

GAS tool

University of Wollongong | La Trobe University | NSW Government Health Murrumbidgee Local Health District | AlfredHealth | Caulfield Hospital

Goal Attainment Scale

Participant Number: | | | | |

Goal Attainment Levels	Goal 1 Date set: Date for r/v:	Goal 2 Date set: Date for r/v:	Goal 3 Date set: Date for r/v:
Best anticipated success with treatment (+2)			
More than expected success with treatment (+1)			
Expected level of treatment success (0)			
Less than expected success with treatment (-1)			
Most unfavourable treatment outcome thought likely (-2)			
Importance			
Difficulty			
Level Achieved			

Baseline GAS T-Score:	Achieved GAS T-Score:	Change in GAS T-Score:

Importance	Difficulty	
0 = not at all (important) 1 = a little (important) 2 = moderately (important) 3= very (important)	0 = not at all (difficult) 1 = a little (difficult) 2 = moderately (difficult) 3 = very (difficult)	**NOTE: Please place an * next to baseline level**

Standard GAS form, Version 1, Date: 11.01.2013

Page1 of 1

GAS and GAS-Light training package outline

Topic	Content
Goal Attainment Scaling methods: original GAS and GAS-Light	1. Introduction to GAS 2. GAS development history and administration. a. Psychometric evidence b. Patient examples presented c. Exemplar completed GAS 3. Introduction to GAS-Light tool 4. GAS-Light development history and administration a. Rationale for development b. Proposed clinical utility c. Patient examples presented d. Exemplars completed GAS-Light 5. Goal setting a. Impairment, activity and participation based goals b. Exemplars presented to generate discussion: i. possible goal areas ii. how to improve goal quality: increasing objectivity and measurement iii. goal interpretation and scoring c. Trouble shooting i. Provision of a handy hints sheet 6. Excel spreadsheet for T-score calculations provided and discussed
Study Implementation as per protocol	1. Introduction to study aims and research questions 2. Present study protocol a. Clinicians and research team members roles defined. b. Explanation of study forms c. Video review demonstrating completion of NIHS stroke scale (Alfred Health site coordinator only) d. Provision of links to further available online training (Alfred Health site coordinator only)

	3. Patient participant packages a. Provision and review of content 4. Provision of additional resources a. Journal articles, guides for writing SMART goals and using GAS methods. 5. Contact number and email provided for support throughout the study.

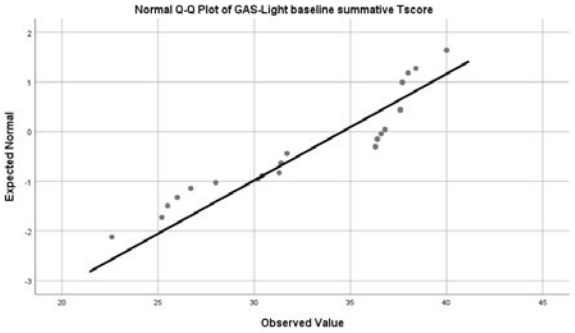
SPSS Outputs

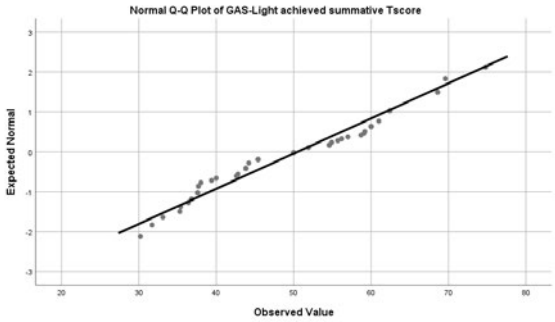
n=58

<table>
<tr><th colspan="2">Statistical Test</th><th>SPSS output</th><th>Interpretation</th></tr>
<tr><td>Kolmogorov -Smirnov

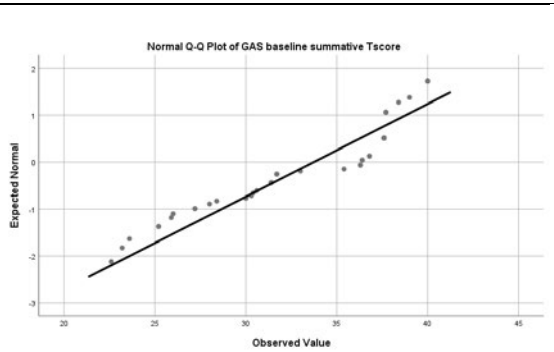
Shapiro-Wilk

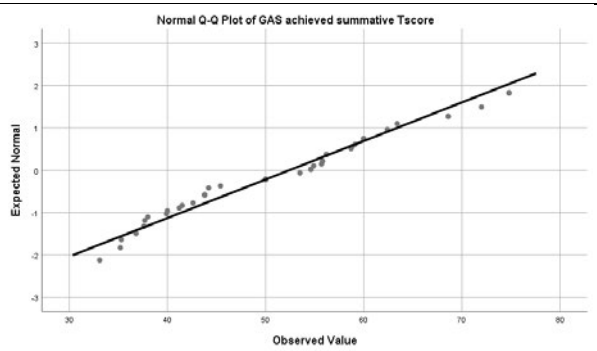
Q-Plot</td><td>Test of the normality of the data

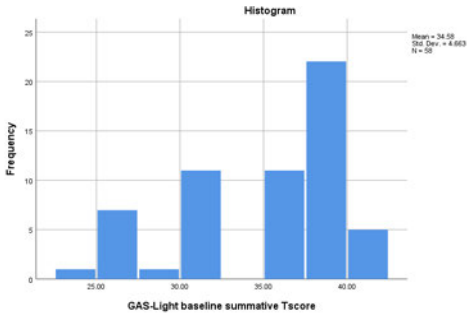
p value < 0.05 provides evidence the distribution is significantly different from normal (Barton & Peat, 2014)</td><td>
Tests of Normality
<table>
<tr><th></th><th colspan="3">Kolmogorov-Smirnov[a]</th><th colspan="3">Shapiro-Wilk</th></tr>
<tr><th></th><th>Statistic</th><th>df</th><th>Sig.</th><th>Statistic</th><th>df</th><th>Sig.</th></tr>
<tr><td>GAS-Light baseline summative T-score</td><td>.299</td><td>58</td><td>.000</td><td>.826</td><td>58</td><td>.000</td></tr>
<tr><td>GAS-Light achieved summative T-score</td><td>.121</td><td>58</td><td>.033</td><td>.955</td><td>58</td><td>.032</td></tr>
<tr><td>GAS baseline summative T-score</td><td>.245</td><td>58</td><td>.000</td><td>.867</td><td>58</td><td>.000</td></tr>
<tr><td>GAS achieved summative T-score</td><td>.117</td><td>58</td><td>.048</td><td>.962</td><td>58</td><td>.070</td></tr>
</table>
a. Lilliefors Significance Correction


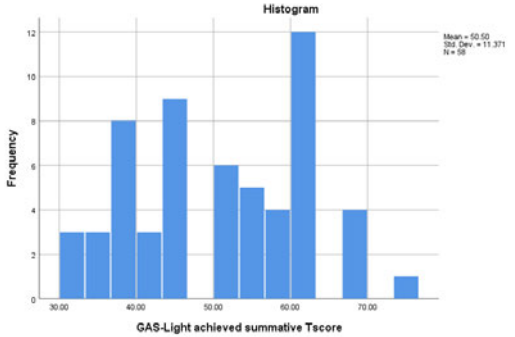


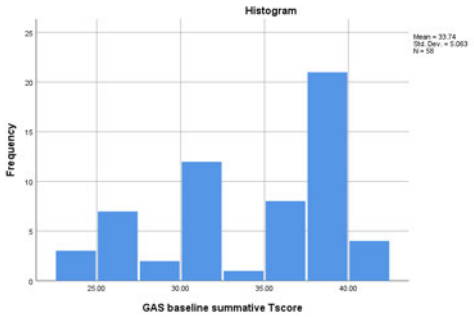
</td><td>Data conforms to normality therefore parametric tests can be used</td></tr>
</table>

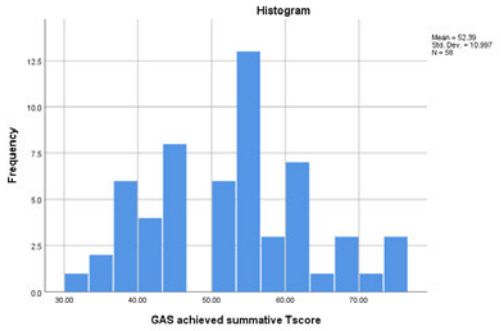

Normal Q-Q Plot of GAS baseline summative Tscore
Expected Normal
Observed Value

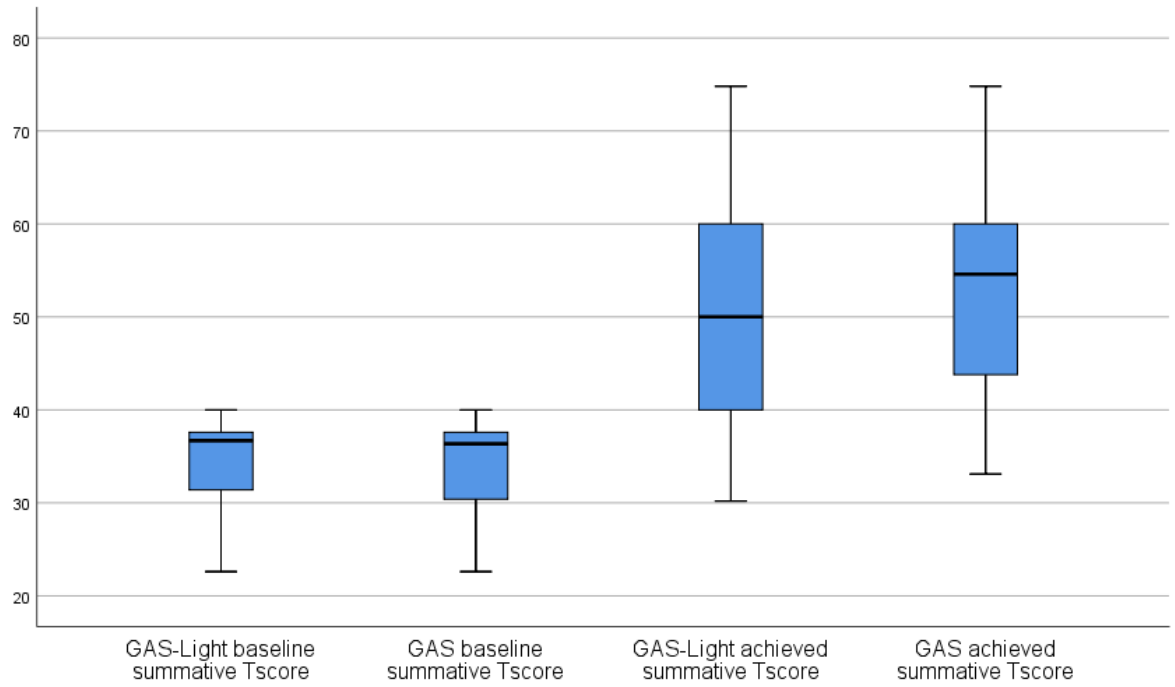

Normal Q-Q Plot of GAS achieved summative Tscore
Expected Normal
Observed Value


Histogram
Frequency
GAS-Light baseline summative Tscore


Histogram
Frequency
GAS-Light achieved summative Tscore


Histogram
Frequency
GAS baseline summative Tscore


Histogram
Frequency
GAS achieved summative Tscore

<table>
<tr>
<td>Boxplot</td>
<td>Provides visual summary and comparison data captured by the two tools at baseline and follow up.</td>
<td> 
</td>
<td>Baseline data: median, interquartile range and minimum and maximum values equal across tools.

Follow up data indicates GAS to have a slightly higher median and less spread of data however, the two tools remains quite similar.

Nil outliers identified at baseline or follow up.</td>
</tr>
<tr>
<td>Intraclass correlation using a two-way random effects model, absolute agreement, single measures</td>
<td>to assess the reliability between the two tools when evaluating goal attainment

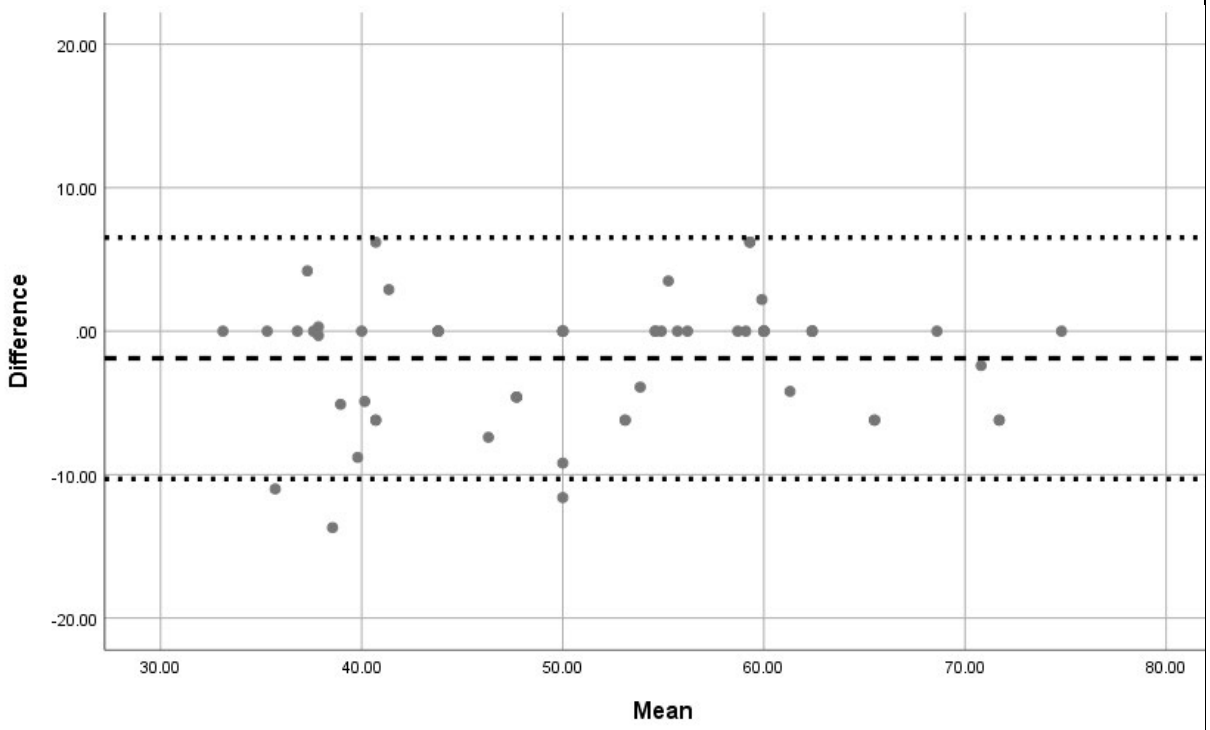
‘Single measures’ statistic is an</td>
<td>Reliability

Scale: ALL VARIABLES
Case Processing Summary
<table>
<tr><th></th><th></th><th>N</th><th>%</th></tr>
<tr><td>Cases</td><td>Valid</td><td>58</td><td>100.0</td></tr>
<tr><td></td><td>Excluded[a]</td><td>0</td><td>.0</td></tr>
<tr><td></td><td>Total</td><td>58</td><td>100.0</td></tr>
</table>
a. Listwise deletion based on all variables in the procedure.</td>
<td>ICC (2,k) = 0.91, 95% CI 0.84 to 0.95, P<.001 indicating a high degree of reliability and suggesting that goal attainment was measured</td>
</tr>
</table>

	index of reliability for typical single raters, which is the most common situation in clinical research. 'Average measures' statistic is an index of reliability for different raters averaged together – for example when the measurement are a mean of the measurements taken by different raters (Barton & Peat, 2014)	**Reliability Statistics** Cronbach's Alpha: .962; N of Items: 2 **Intraclass Correlation Coefficient** 95% Confidence Interval; F Test with True Value 0 Single Measures: Intraclass Correlation[b] .914[a], Lower Bound .836, Upper Bound .953, Value 26.165, df1 57, df2 57, Sig .000 Average Measures: .955, .911, .976, 26.165, 57, 57, .000 Two-way random effects model where both people effects and measures effects are random. a. The estimator is the same, whether the interaction effect is present or not. b. Type A intraclass correlation coefficients using an absolute agreement definition.	similarly by the two tools. ICC of 1 would indicate perfect concordance between measurements.

Bland-Altman method	Assesses the agreement between the two tools.
	
	Tests for any systematic
		`DESCRIPTIVES VARIABLES=Difference`
		`  /STATISTICS=MEAN STDDEV MIN MAX.`
		
		Descriptive Statistics
		Difference: N 58, Minimum -13.70, Maximum 6.20, Mean -1.8897, Std. Deviation 4.29217
		Valid N (listwise): 58
			The scatter on the plot demonstrates a small amount of systematic bias due.

	bias or differences	**Paired Samples Statistics** Mean / N / Std. Deviation / Std. Error Mean Pair 1 — GAS-Light achieved summative T-score: 50.5034 / 58 / 11.37081 / 1.49306 Pair 1 — GAS achieved summative T-score: 52.3931 / 58 / 10.99682 / 1.44395 **Paired Samples Correlations** N / Correlation / Sig. Pair 1 — GAS-Light achieved summative T-score & GAS achieved summative T-score: 58 / .927 / .000 **Paired Samples Test** Paired Differences: Mean / Std. Deviation / Std. Error Mean / 95% Confidence Interval of the Difference (Lower / Upper) / t / df / Sig. (2-tailed) Pair 1 — GAS-Light achieved summative T-score - GAS achieved summative T-score: -1.88966 / 4.29217 / .56359 / -3.01822 / -.76109 / -3.353 / 57 / .001	Demonstrates that 95% of the differences between the GAS and GAS-Light are included.

		To determine the LoA: SD x 1.96 = 8.14 Mean + 8.14= 6.52 Mean – 8.14 = -10.30  	

<table>
<tr>
<td>Measurement error</td>
<td>Used to describe agreement (measurement error), then converted into a range which indicates that the average of all possible scores is within the error range above and below the actual measurement taken.</td>
<td>

```
DESCRIPTIVES VARIABLES=Difference
  /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Difference	58	-13.70	6.20	-1.8897	4.29217
Valid N (listwise)	58				

Measurement Error = SD of Difference / √2
4.29217 / 1.414 = 3.03548

Converted to a range by multiplying by a critical value of 1.96 = 5.95

	Mean difference	Limits of agreement	Error range	ICC
Goal attainment	-1.8897	-10.30, 6.52	5.95	0.96

</td>
<td>Error range indicates that the average of all possible GAS-Light scores are within the range of 5.95 above and 5.95 below GAS ratings – the range within which the true score lies.

Clinically meaningful change is deemed to be 10% change (F. Khan, J. F. Pallant, & L. Turner-Stokes, 2008</td>
</tr>
<tr>
<td>Root Mean Square Error</td>
<td>An absolute measure of fit. Frequently used measure of the difference between the values predicted by a model or an</td>
<td>

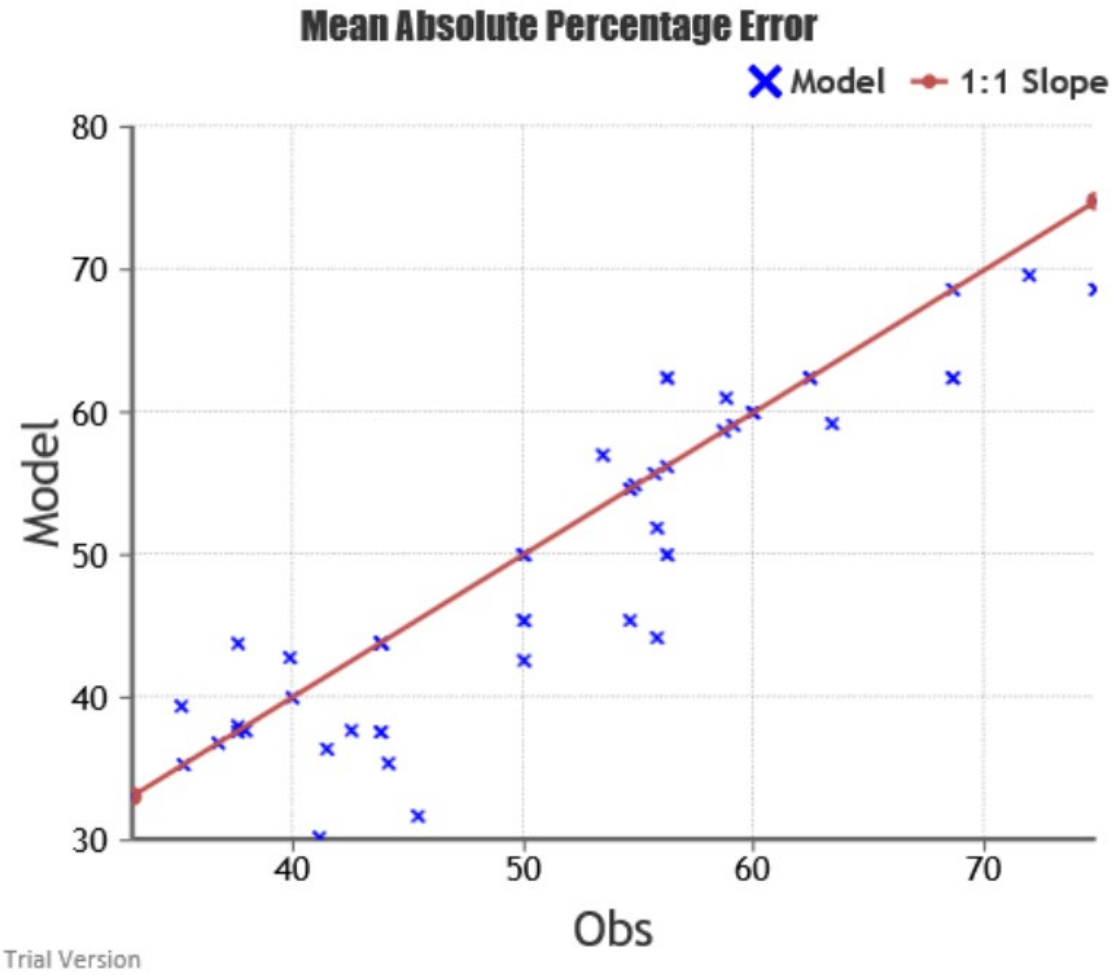
Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	GAS achieved summative T-score[b]	.	Enter

</td>
<td>RMSE: 4.30567 which indicates the GAS-Light would be ±4.31 either way of the true GAS mean, such as if the GAS-Light achieved</td>
</tr>
</table>

	estimator and the values observed = how much error between the two datasets with GAS – the observed and GAS-Light the predicted	a. Dependent Variable: GAS-Light achieved summative T-score b. All requested variables entered. **Model Summary** | Model	R	R Square	Adjusted R Square	Std. Error of the Estimate | |---|---|---|---|---| | 1	.927[a]	.859	.857	4.30567 | a. Predictors: (Constant), GAS achieved summative T-score	T-score was 50 the true GAS score would actually be somewhere within 54.31 and 45.69. 85.7% of the variance in GAS-Light can be predicted from GAS.

Mean absolute % error	equal to the error range calculated above.	 Mean Absolute Percentage Error = 5.9%	Agreed at meeting not to report this

<table>
<tr>
<td>Cohen’s d Effect size</td>
<td>Effect size: mean change / SD (of baseline score) (F. Khan, Pallant, et al., 2008; Rockwood et al., 2003)

Standardised Response Mean (mean change/SD of change score) (Khan, Pallant, et al., 2008)</td>
<td>
Descriptives
Descriptive Statistics
<table>
<tr><th></th><th>N</th><th>Mean</th><th>Std. Deviation</th></tr>
<tr><td>GAS-Light baseline summative T-score</td><td>58</td><td>34.5793</td><td>4.66257</td></tr>
<tr><td>GAS-Light achieved summative T-score</td><td>58</td><td>50.5034</td><td>11.37081</td></tr>
<tr><td>GAS baseline summative T-score</td><td>58</td><td>33.7362</td><td>5.06345</td></tr>
<tr><td>GAS achieved summative T-score</td><td>58</td><td>52.3931</td><td>10.99682</td></tr>
<tr><td>Valid N (listwise)</td><td>58</td><td></td><td></td></tr>
</table>
Descriptive Statistics
<table>
<tr><th></th><th>N</th><th>Mean</th><th>Std. Deviation</th></tr>
<tr><td>GAS-Light change summative T-score</td><td>58</td><td>15.9241</td><td>9.93393</td></tr>
<tr><td>GAS change summative T-score</td><td>58</td><td>18.7172</td><td>10.35410</td></tr>
<tr><td>Valid N (listwise)</td><td>58</td><td></td><td></td></tr>
</table>
Cohen d estimate
GAS-Light = 50.5034 – 34.5793 / 4.66257 = 3.42
GAS = 52.3931 – 33.7362 / 5.06345 = 3.68
Standardised Response Mean (SRM)
GAS-Light = 50.5034 – 34.5793 / 9.93393 = 1.60
GAS = 52.3931 – 33.7362 / 10.35410 = 1.81
</td>
<td>In comparison to the GAS, the GAS-Light captured a similar magnitude of change scores as demonstrated by similar and high effect size and SRM.</td>
</tr>
</table>

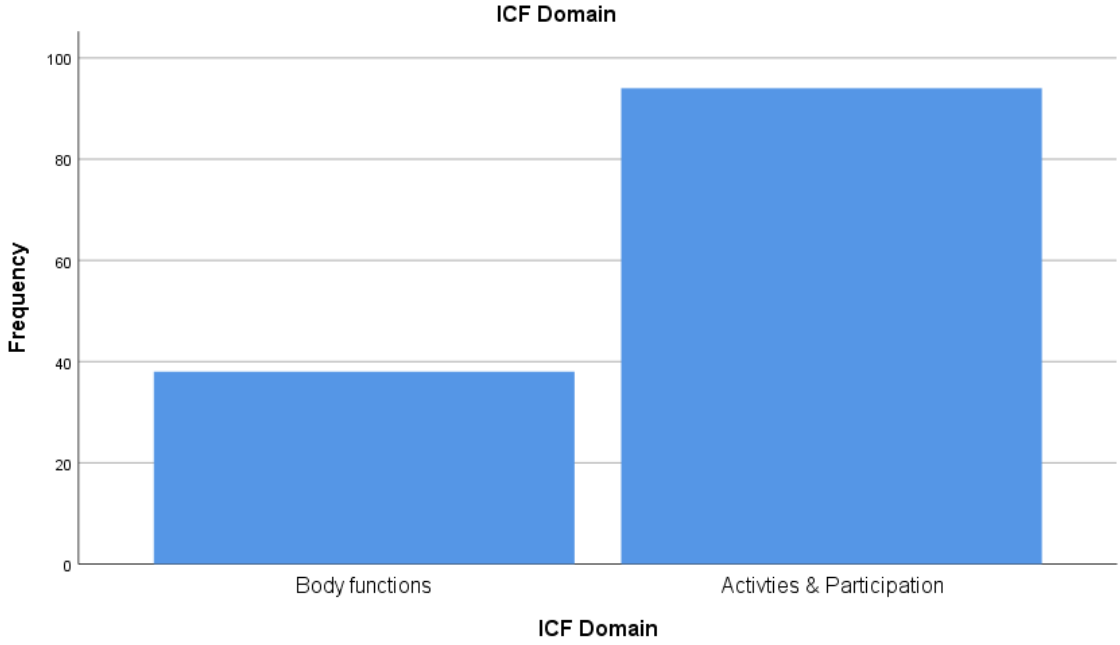
Descriptive information: Goal ICF domain, category and code

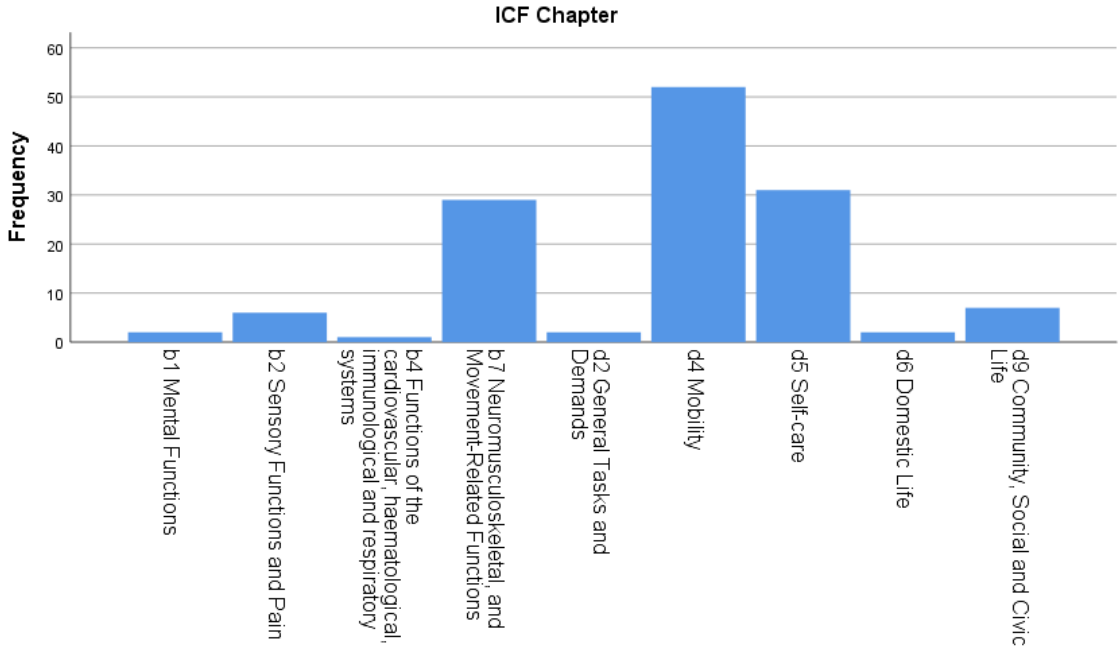
ICF Domain

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Body functions	38	28.8	28.8	28.8
	Activities & Participation	94	71.2	71.2	100.0
	Total	132	100.0	100.0	

ICF Chapter

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	b1 Mental Functions	2	1.5	1.5	1.5
	b2 Sensory Functions and Pain	6	4.5	4.5	6.1
	b4 Functions of the cardiovascular, haematological, immunological and respiratory systems	1	.8	.8	6.8
	b7 Neuromusculoskeletal, and Movement-Related Functions	29	22.0	22.0	28.8
	d2 General Tasks and Demands	2	1.5	1.5	30.3
	d4 Mobility	52	39.4	39.4	69.7
	d5 Self-care	31	23.5	23.5	93.2
	d6 Domestic Life	2	1.5	1.5	94.7
	d9 Community, Social and Civic Life	7	5.3	5.3	100.0
	Total	132	100.0	100.0	


ICF Domain
Frequency
100
80
60
40
20
0
Body functions
Activties & Participation
ICF Domain


ICF Chapter
Frequency
60
50
40
30
20
10
0
b1 Mental Functions
b2 Sensory Functions and Pain
b4 Functions of the cardiovascular, haematological, immunological and respiratory systems
b7 Neuromusculoskeletal, and Movement-Related Functions
d2 General Tasks and Demands
d4 Mobility
d5 Self-care
d6 Domestic Life
d9 Community, Social and Civic Life

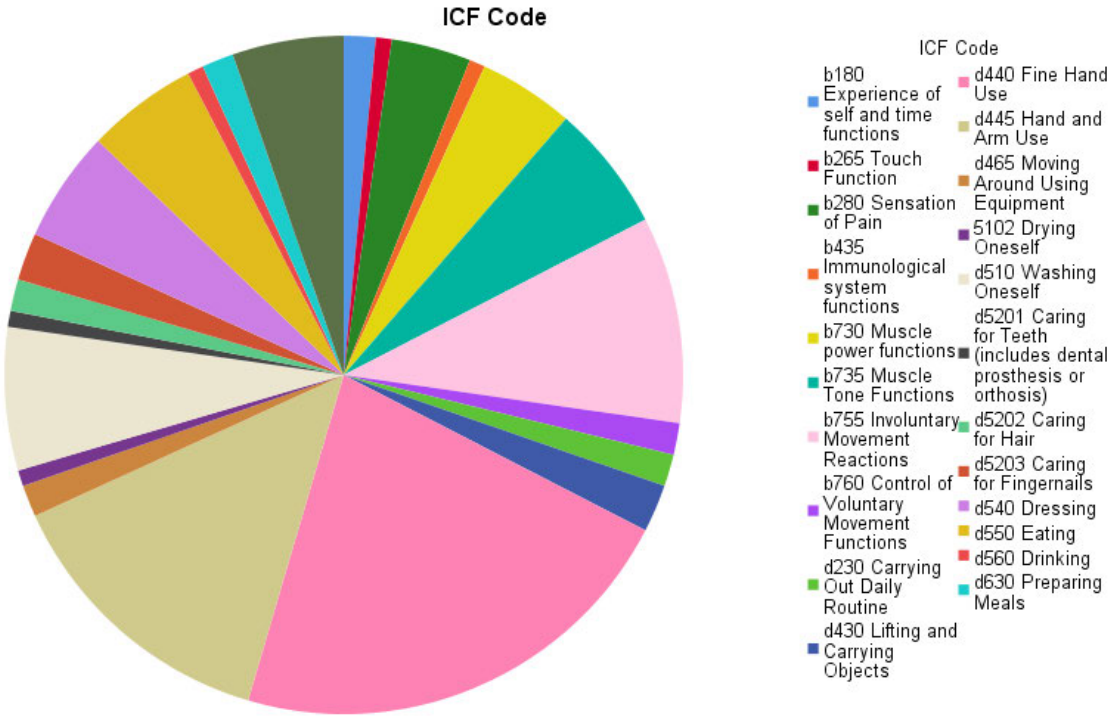
Statistics

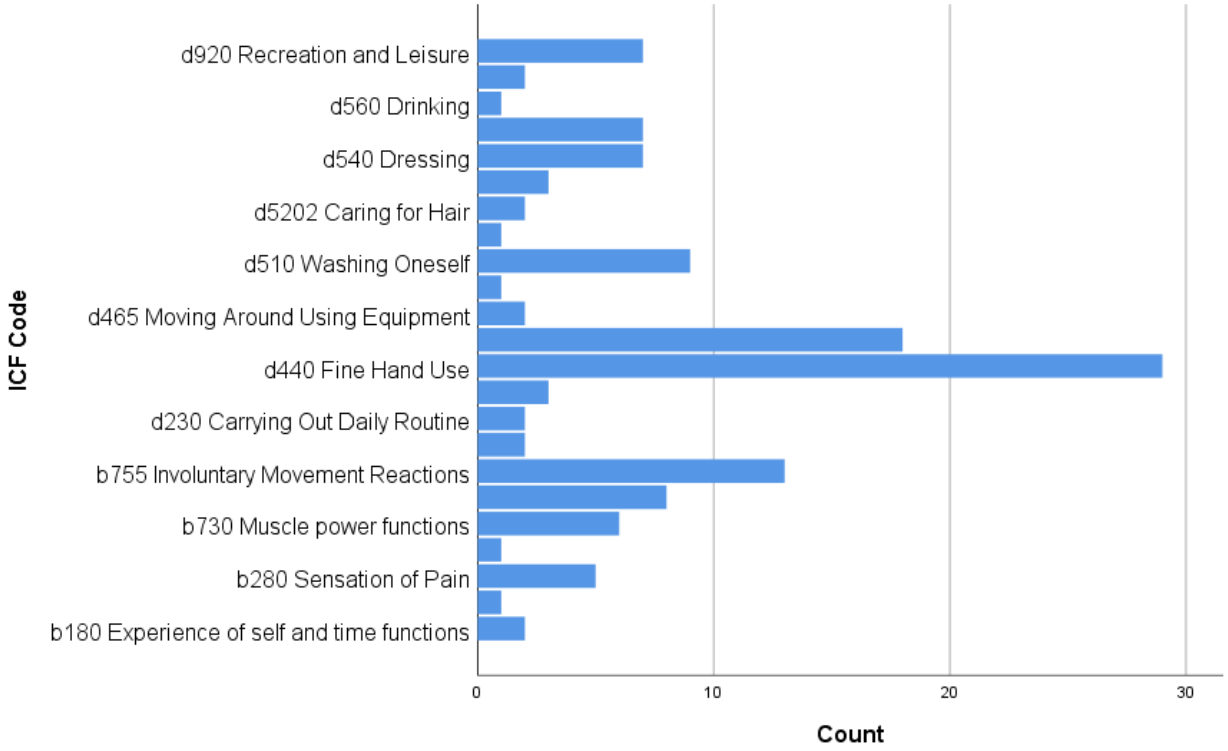
ICF Code

N	Valid	132
	Missing	0

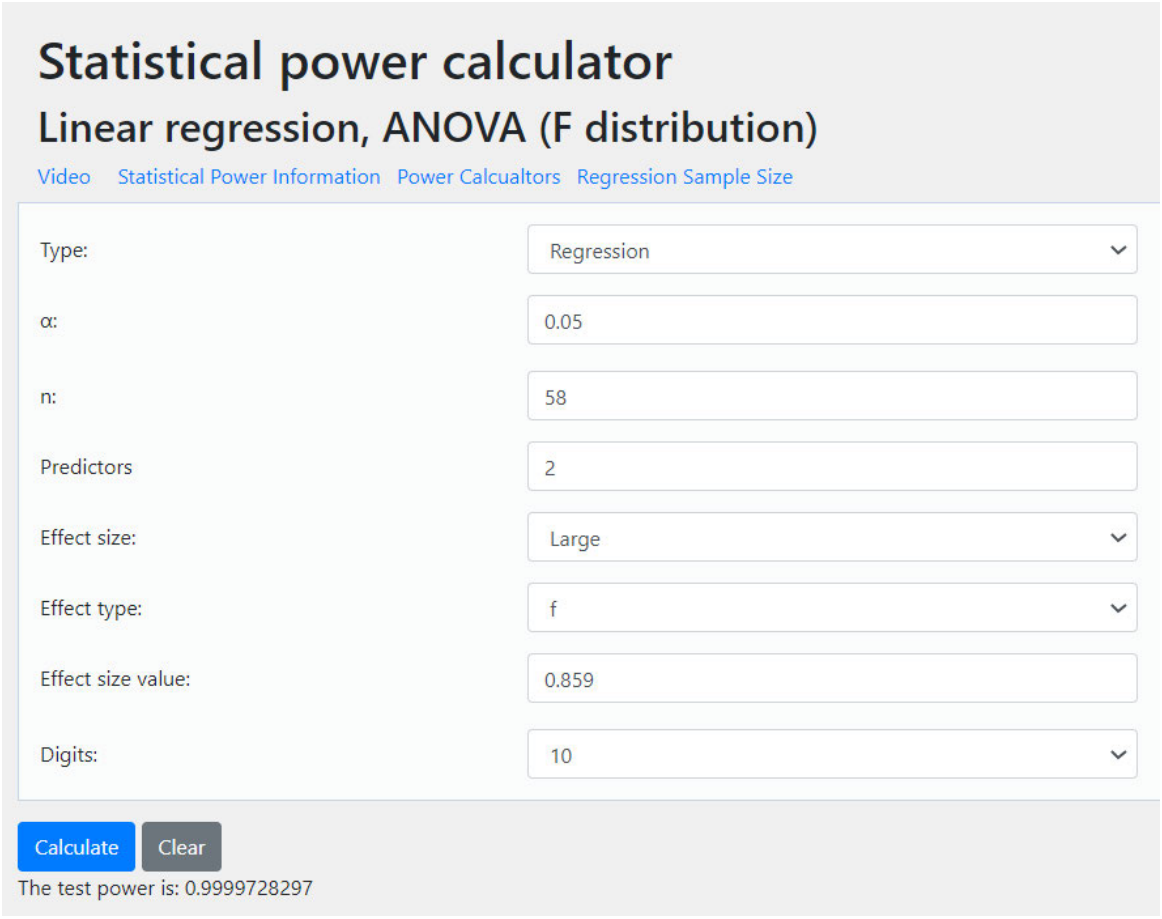
ICF Code

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	b180 Experience of self and time functions	2	1.5	1.5	1.5
	b265 Touch Function	1	.8	.8	2.3
	b280 Sensation of Pain	5	3.8	3.8	6.1
	b435 Immunological system functions	1	.8	.8	6.8
	b730 Muscle power functions	6	4.5	4.5	11.4
	b735 Muscle Tone Functions	8	6.1	6.1	17.4
	b755 Involuntary Movement Reactions	13	9.8	9.8	27.3
	b760 Control of Voluntary Movement Functions	2	1.5	1.5	28.8
	d230 Carrying Out Daily Routine	2	1.5	1.5	30.3
	d430 Lifting and Carrying Objects	3	2.3	2.3	32.6
	d440 Fine Hand Use	29	22.0	22.0	54.5
	d445 Hand and Arm Use	18	13.6	13.6	68.2
	d465 Moving Around Using Equipment	2	1.5	1.5	69.7
	5102 Drying Oneself	1	.8	.8	70.5
	d510 Washing Oneself	9	6.8	6.8	77.3
	d5201 Caring for Teeth (includes dental prosthesis or orthosis)	1	.8	.8	78.0
	d5202 Caring for Hair	2	1.5	1.5	79.5
	d5203 Caring for Fingernails	3	2.3	2.3	81.8
	d540 Dressing	7	5.3	5.3	87.1
	d550 Eating	7	5.3	5.3	92.4
	d560 Drinking	1	.8	.8	93.2
	d630 Preparing Meals	2	1.5	1.5	94.7
	d920 Recreation and Leisure	7	5.3	5.3	100.0
	Total	132	100.0	100.0	


ICF Code
ICF Code
b180 Experience of self and time functions
b265 Touch Function
b280 Sensation of Pain
b435 Immunological system functions
b730 Muscle power functions
b735 Muscle Tone Functions
b755 Involuntary Movement Reactions
b760 Control of Voluntary Movement Functions
d230 Carrying Out Daily Routine
d430 Lifting and Carrying Objects
d440 Fine Hand Use
d445 Hand and Arm Use
d465 Moving Around Using Equipment
5102 Drying Oneself
d510 Washing Oneself
d5201 Caring for Teeth (includes dental prosthesis or orthosis)
d5202 Caring for Hair
d5203 Caring for Fingernails
d540 Dressing
d550 Eating
d560 Drinking
d630 Preparing Meals


ICF Code
d920 Recreation and Leisure
d560 Drinking
d540 Dressing
d5202 Caring for Hair
d510 Washing Oneself
d465 Moving Around Using Equipment
d440 Fine Hand Use
d230 Carrying Out Daily Routine
b755 Involuntary Movement Reactions
b730 Muscle power functions
b280 Sensation of Pain
b180 Experience of self and time functions
0
10
20
30
Count

Observed power analysis



Linear regression power calculator (statskingdom.com)

Appendix H: Supplementary material from Study 5

- Participant demographic information (see Appendix G)
- Patient level of engagement in goal setting tool
- Patient satisfaction with the goal setting process tool
- Patient acceptance of GAS-Light survey
- Clinician Participant survey: GAS
- Patient participant survey: GAS-Light
- SPSS Output

Patient level of engagement in goal setting tool

University of Wollongong | La Trobe University | NSW Health Murrumbidgee Local Health District | AlfredHealth | Caulfield Hospital

Participant Number: ☐☐☐☐ Date: / /

Patient level of engagement in goal setting

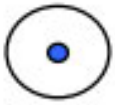
(as judged by team)

Excellent engagement
Fully independent in goal monitoring and setting their own goals

Very good engagement,
Patient takes most of responsibility for monitoring and re-setting goals

Good engagement,
But requires active support
Patient and team take 50/50 responsibility

Moderate engagement,
Patient engages to some degree, but team takes most of responsibility (>50%) for monitoring and re-setting goals



Minimal engagement,
Patient indicates general goal area, but cannot engage in goal setting to any meaningful level

Unable
Cannot engage in goal setting at any level

This is a scale to record the level of engagement of a patient in their own goal setting.
It takes into account a number of factors related to goal-setting behaviour:

- Their **cognitive ability** to be aware of themselves, their situation and their environment
- Their **communicative ability** to articulate their priorities and frame those in specific goals
- Their **adjustment to limitations** and level of **realistic expectation** for the future.
- Their **behavioural approach to rehabilitation**, including self-monitoring, motivation and ability to organize themselves

The simple scale below does not attempt to tease these out.
If the patient is at different level with respect to these factors – eg they have the cognitive ability to understand, but cannot/ will not accept the concept of goal negotiation, **score to the lower!**

Patient satisfaction with the goal setting process tool

Participant Number: ☐☐☐☐ Date: / /

Patient satisfaction with the goal setting process

(as judged by patient / family)

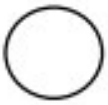
Excellent
My goals matched all my key priorities for rehab
And were entirely my own choice

Very good
My goals matched my main priorities for rehab
And I was pretty happy with my agreed goal-set

Good
My goals met most of my priorities for rehab
And I agreed with most of them

Moderate
My goals met some of my priorities for rehab
And I agreed with some of them

Poor
My goals were largely irrelevant to me
And I disagreed with most of them



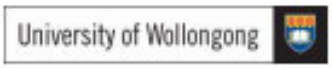
None
My goals were were completely irrelevant
And I did not agree with any of them
Or
What goals???

This is a scale to record the patient's satisfaction with goal setting.
It takes into account a number of factors related to goals:

- How well the goals **matched their priorities** for rehab
- The extent to which they **agreed** with the goals
- The extent of **choice in goal areas**
- The extent to which they felt **involved with / in charge** of the goal setting process

The simple scale below does not attempt to tease these out.
If the patient is at different level with respect to these factors – eg they has a wide choice of goals but did not agree with any of them, **score to the lower**!

Patient acceptance of GAS-Light survey



Patient Acceptance: Goal Attainment Scale-Light

This survey is completely confidential and only summary information will be reported in the study results. Thank you in advance for your help with this survey.

For each question, please mark (X) one box to the right to indicate your response		Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
1	Having the GAS-Light completed with me was not uncomfortable in any way					
2	I felt happy and/or comfortable answering all questions					
3	Considering the time cost to me, I think undergoing the GAS-Light test was worthwhile					
4	I think that information provided by the GAS-Light will be helpful in making decisions about how I am progressing towards achieving my goals.					
5	I would be happy to have the GAS-Light completed on me again should the Occupational Therapist need this information					

Demographics:

6. Gender: ☐Male ☐Female

7. Age (years only): _____

8. Primary Diagnosis/Reason for admission:_____

Office Use Only: Hospital:_____

Caseload: _____

Clinician participant survey: GAS

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Clinician Clinical Utility Survey: GAS

The Goal Attainment Scale (GAS) is a method of scoring the extent to which a patient's individual goals are achieved.

This survey is completely confidential and only summary information will be reported in the study results. Thank you in advance for your help.

1. What is your qualification

☐ Occupational Therapist ☐ Physiotherapist ☐ other, please specify_____

2. How many years have you been working in the above profession? _____

3. Have you ever used the GAS in your clinical practice (prior to your assistance in this study?)

☐ Yes, if yes, please answer questions A and B

☐ No, please go to question 4

(A) Please estimate the number of years you have been using the GAS in clinical practice:

_____

(B) Please estimate the average number of times you use the GAS in a month:

_____

4. Did you receive training on how to use the GAS tool for use within this study?

☐ Yes, please go to question 6

☐ No, please complete question 5

5. Participating in training in the GAS tool would have improved my ability to implement this tool

☐ strongly disagree ☐ disagree ☐ neither disagree nor agree ☐ agree ☐ strongly agree

6. On average, how long did it take to complete the ***GAS*** tool at ***baseline*** for upper limb spasticity management? (Includes developing the goal. Does not include the time taken to complete assessment of the participant)

_____mins

7. On average, how long did it take to complete the ***GAS*** tool at ***follow up*** for upper limb spasticity management? (does not include time taken to complete assessment of the participant)

_____mins

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Your Thoughts about the GAS

Please place an 'X' in the box which best describes your response to the following statements:
(Letss & Law)

		Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
8	The GAS is **accessible** at my place of work					
9	The format and style of the GAS is **easy to understand** and follow					
10	The time for **administration** of the GAS is appropriate for incorporation into clinical practice					
11	The time for **scoring** of the GAS is appropriate for incorporation into clinical practice					
12	The GAS assists with clinical **decision making**					
13	The GAS is **cost effective**					
14	The GAS gives **unique information** compared with other tools of goal attainment assessment that I use.					
15	The format of the GAS is **acceptable** to clients.					
16	The GAS is **responsive** to clinically meaningful change					

(Mayers, 2003, pp. 390-391)

17. Do you think that the GAS will be useful for the majority of your patients presenting for upper limb spasticity management?

☐ Yes.
☐ No. Why? _____

18. Having used the GAS, please state any important area(s) or item(s) that you think have been omitted.

19. Have you used the ***GAS-Light tool*** before?

☐ Yes, if yes, please answer the following questions

☐ No, thank you for completing the survey

20. In your experience in using the ***GAS-Light tool*** in upper limb spasticity management, on average, how long does it take to complete the GAS-Light tool at ***baseline***? (includes developing the full follow up guide. Does not include time taken to complete assessment of the patient)

_____mins

21. In your experience in using the ***GAS-Light tool*** in upper limb spasticity management, on average, how long does it take to complete the GAS-Light tool at ***follow up***? (does not include time taken to complete assessment of the patient)

_____mins

Comparing the GAS to the GAS-Light Reflecting on **both** tools (the GAS and GAS-Light), please place an 'X' in the box which best describes your response to the following statements:		Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
22	The format of the GAS is **easier to understand** and follow than the format of the GAS-Light					
23	The time for **administration** of the GAS is more appropriate for incorporation into clinical practice than the time to administer the GAS-Light					
24	The time for **scoring** of the GAS is more appropriate for incorporation into clinical practice than the time to score the GAS-Light					
25	The GAS assists with **clinical decision making** more than the GAS-Light					
26	The GAS is **more useful** for the majority of my patients presenting for upper limb spasticity management than the GAS-Light.					

Thank you for your participation.

Please return the completed questionnaire to the research team via the addressed envelope.

Shannon Pike
Ambulatory Rehabilitation Service
Wagga Wagga Base Hospital
PO Box 159
Wagga Wagga, NSW 2650

Clinician participant survey: GAS-Light

University of Wollongong | La Trobe University | NSW Health Murrumbidgee Local Health District

Clinician Clinical Utility Survey: GAS-Light

The Goal Attainment Scale-Light (GAS-Light) is a method of scoring the extent to which a patient's individual goals are achieved.

This survey is completely confidential and only summary information will be reported in the study results. Thank you in advance for your help.

1. What is your qualification

☐ Occupational Therapist ☐ Physiotherapist ☐ other, please specify_____

2. How many years have you been working in the above profession? _____

3. Have you ever used the GAS-Light in your clinical practice (prior to your assistance in this study?)

   ☐ Yes, if yes, please answer questions A and B

   ☐ No, please go to question 4

(A) Please estimate the number of years you have been using the GAS-Light in clinical practice:

_____

(B) Please estimate the average number of times you use the GAS-Light in a month:

_____

4. Did you receive training on how to use the GAS-Light tool for use within this study?

   ☐ Yes, please go to question 6

   ☐ No, please complete question 5

5. Participating in training in the GAS-Light tool would have improved my ability to implement this tool

☐ strongly disagree ☐ disagree ☐ neither disagree nor agree ☐ agree ☐ strongly agree

6. On average, how long did it take to complete the ***GAS-Light*** tool at ***baseline*** for upper limb spasticity management? (Includes developing the goal. Does not include the time taken to complete assessment of the participant)

_____mins

7. On average, how long did it take to complete the ***GAS-Light*** tool at ***follow up*** for upper limb spasticity management? (does not include time taken to complete assessment of the participant)

_____mins

University of Wollongong | LA TROBE UNIVERSITY | NSW Health Murrumbidgee Local Health District

Your Thoughts about the GAS-Light

Please place an 'X' in the box which best describes your response to the following statements:
(Letss & Law)

		Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
8	The GAS-Light is **accessible** at my place of work					
9	The format and style of the GAS-Light is **easy to understand** and follow					
10	The time for **administration** of the GAS-Light is appropriate for incorporation into clinical practice					
11	The time for **scoring** of the GAS-Light is appropriate for incorporation into clinical practice					
12	The GAS-Light assists with clinical **decision making**					
13	The GAS-Light is **cost effective**					
14	The GAS-Light gives **unique information** compared with other tools of goal attainment assessment that I use.					
15	The format of the GAS-Light is **acceptable** to clients.					
16	The GAS-Light is **responsive** to clinically meaningful change					

(Mayers, 2003, pp. 390-391)

17. Do you think that the GAS-Light will be useful for the majority of your patients presenting for upper limb spasticity management?

☐ Yes.
☐ No. Why? _____

18. Having used the GAS-Light, please state any important area(s) or item(s) that you think have been omitted.

19. Have you used the ***standard GAS tool*** before?

☐ Yes, if yes, please answer the following questions

☐ No, thank you for completing the survey

20. In your experience in using the ***standard GAS tool*** in upper limb spasticity management, on average, how long does it take to complete the standard GAS tool at ***baseline***? (includes developing the full follow up guide. Does not include time taken to complete assessment of the patient)

_____mins

21. In your experience in using the ***standard GAS tool*** in upper limb spasticity management, on average, how long does it take to complete the standard GAS tool at ***follow up***? (does not include time taken to complete assessment of the patient)

_____mins

Comparing the GAS-Light to the GAS Reflecting on **both** tools (the GAS-Light and GAS), please place an 'X' in the box which best describes your response to the following statements:		Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
22	The format of the GAS-Light is **easier to understand** and follow than the format of the GAS					
23	The time for **administration** of the GAS-Light is more appropriate for incorporation into clinical practice than the time to administer the GAS					
24	The time for **scoring** of the GAS-Light is more appropriate for incorporation into clinical practice than the time to score the GAS					
25	The GAS-Light assists with **clinical decision making** more than the GAS					
26	The GAS-Light is **more useful** for the majority of my patients presenting for upper limb spasticity management than the GAS.					

Thank you for your participation.

Please return the completed questionnaire to the research team via the addressed envelope.

Shannon Pike
Ambulatory Rehabilitation Service
Wagga Wagga Base Hospital
PO Box 159
Wagga Wagga, NSW 2650

SPSS output

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
OT years of clinical exp	15	14.00	2.00	16.00	7.7667	4.27980
PT years of clinical exp	4	17.00	2.00	19.00	8.0000	7.52773
Valid N (listwise)	4					

```
DESCRIPTIVES VARIABLES=OT_exp PT_exp All_clinicians
  /STATISTICS=MEAN STDDEV RANGE MIN MAX.
```

Descriptives

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
OT years of clinical exp	15	14.00	2.00	16.00	7.7667	4.27980
PT years of clinical exp	4	17.00	2.00	19.00	8.0000	7.52773
All clinicians yrs clinical exp	19	17.00	2.00	19.00	7.8158	4.86829
Valid N (listwise)	4					

```
DESCRIPTIVES VARIABLES=GAS_Baseline_Time GAS_followup_time
GAS_Light_baseline_time
    GAS_Light_followup_time
  /STATISTICS=MEAN STDDEV RANGE MIN MAX.
```

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
GAS baseline time	18	45	15	60	33.06	17.998
GAS followup time	19	55	5	60	16.58	17.244
GAS-Light baseline time	16	55	5	60	21.88	19.397
GAS-Light follow up time	16	38	2	40	12.44	14.085
Valid N (listwise)	16					

```
FREQUENCIES VARIABLES=GAS_Baseline_Time GAS_followup_time
GAS_Light_baseline_time
    GAS_Light_followup_time
  /STATISTICS=STDDEV RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE
  /ORDER=ANALYSIS.
```

Frequencies

Statistics

		GAS baseline time	GAS follow up time	GAS-Light baseline time	GAS-Light follow up time
N	Valid	18	19	16	16
	Missing	82	81	84	84

Mean	33.06	16.58	21.88	12.44
Median	30.00	10.00	15.00	5.00
Mode	15	5	10	5
Std. Deviation	17.998	17.244	19.397	14.085
Range	45	55	55	38
Minimum	15	5	5	2
Maximum	60	60	60	40

Frequency Table

GAS baseline time

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15	6	6.0	33.3	33.3
	20	2	2.0	11.1	44.4
	30	2	2.0	11.1	55.6
	40	3	3.0	16.7	72.2
	45	1	1.0	5.6	77.8
	60	4	4.0	22.2	100.0
	Total	18	18.0	100.0	
Missing	System	82	82.0		
Total		100	100.0		

GAS follow up time

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5	7	7.0	36.8	36.8
	10	6	6.0	31.6	68.4
	20	2	2.0	10.5	78.9
	30	2	2.0	10.5	89.5
	60	2	2.0	10.5	100.0
	Total	19	19.0	100.0	
Missing	System	81	81.0		
Total		100	100.0		

GAS-Light baseline time

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5	1	1.0	6.3	6.3
	10	6	6.0	37.5	43.8

	15	3	3.0	18.8	62.5
	20	3	3.0	18.8	81.3
	60	3	3.0	18.8	100.0
	Total	16	16.0	100.0	
Missing	System	84	84.0		
Total		100	100.0		

GAS-Light follow up time

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	3	3.0	18.8	18.8
	5	6	6.0	37.5	56.3
	8	1	1.0	6.3	62.5
	10	2	2.0	12.5	75.0
	15	1	1.0	6.3	81.3
	40	3	3.0	18.8	100.0
	Total	16	16.0	100.0	
Missing	System	84	84.0		
Total		100	100.0		

Appendix I: Published manuscripts

Study 3

Pike, S., Lannin, N. A., Cusick, A., Wales, K., Turner-Stokes, L., & Ashford, S. (2015). A systematic review protocol to evaluate the psychometric properties of measures of function within adult neuro-rehabilitation. *Systematic reviews*, 4(1), 86. doi:10.1186/s13643-015-0076-5

Pike, S., Cusick, A., Wales, K., Cameron, L., Turner-Stokes, L., Ashford, S., & Lannin, N. A. (2021) Psychometric properties of measures of upper limb activity performance in adults with and without spasticity undergoing neurorehabilitation – A systematic review. PLoS ONE 16(2): e0246288. doi.org/10.1371/journal.pone.0246288

Pike, S., Lannin, N. A., Wales, K., & Cusick, A. (2018). A systematic review of the psychometric properties of the Action Research Arm Test in neurorehabilitation. *Australian Occupational Therapy Journal, 65*(5), 449-471. doi:10.1111/1440-1630.12527

Pike et al. Systematic Reviews (2015) 4:86
DOI 10.1186/s13643-015-0076-5

SYSTEMATIC REVIEWS

PROTOCOL Open Access

A systematic review protocol to evaluate the psychometric properties of measures of function within adult neuro-rehabilitation

Shannon Pike[1,2*], Natasha Anne Lannin[1,3], Anne Cusick[4,5], Kylie Wales[6,7], Lynne Turner-Stokes[8,9] and Stephen Ashford[8,9]

Abstract

Background: Spasticity in the upper limb is common after acquired brain impairment and may have a significant impact on the ability to perform meaningful daily activities. Traditionally, outcome measurement in spasticity rehabilitation has focused on impairment, however, improvements in impairments do not necessarily translate to improvements in an individual's ability to perform activities or engage in life roles. There is an increasing need for outcome measures that capture change in activity performance and life participation.

Methods/Design: We will conduct a systematic review of the psychometric properties of instruments used to measure upper limb functional outcomes (activity performance and participation) in patients with spasticity. Assessments (n = 27) will be identified from a recently published systematic review of assessments that measure upper limb function in neurological rehabilitation for adults with focal spasticity, and a systematic review of each assessment will then be conducted. The databases MEDLINE, CINAHL and EMBASE will be searched from inception. Search strategies will include the name of the assessment and the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) published search strategy for identifying studies of measurement properties. The methodological rigour of the testing of the psychometric quality of instruments will be undertaken using the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist. International Classification of Functioning, Disability and Health (ICF) definitions of impairment, activity and participation will be used for content analysis of items to determine the extent to which assessments are valid measures of activity performance and life participation. We will present a narrative synthesis on the psychometric properties and utility of all instruments and make recommendations for assessment selection in practice.

Discussion: This systematic review will present a narrative synthesis on the psychometric properties and utility of assessments used to evaluate function in adults with upper limb focal spasticity. Recommendations for assessment selection in practice will be made which will aid clinicians, managers and funding bodies to select an instrument fit for purpose. Importantly, appropriate assessment selection will provide a mechanism for capturing how applicable to everyday life the outcomes from individualised rehabilitation programs for the upper limb really are.

Systematic review registration: PROSPERO CRD42014013190

Keywords: Botulinum toxin, Upper limb, Activity performance, Participation, Neuro-rehabilitation, Outcome measurement

* Correspondence: Shannon.pike@gsahs.health.nsw.gov.au
[1]Department of Occupational Therapy, La Trobe University, Melbourne, Australia
[2]Wagga Wagga Ambulatory Rehabilitation Service, Wagga Wagga Health Service, Wagga Wagga, Australia
Full list of author information is available at the end of the article

Background

Spasticity is most commonly accepted as being a velocity-dependent increase in stretch reflex with exaggerated tendon jerks that occurs as one part of the upper motor neurone syndrome [1]. Pandyan et al [2] more recently redefined spasticity as disordered sensorimotor control, resulting from an upper motor neurone lesion, presenting as intermittent or sustained involuntary activation of muscles. In recent years, there has been a recognition within neuro-rehabilitation that spasticity management programs must go well beyond treatment of impairments, in line with contemporary understandings of health emerging from the World Health Organisation's International Classification of Functioning, Disability and Health Framework (ICF) [3]. This framework starts with the assumption that health is not a state independent of individuals in the context of everyday life; thus spasticity, a neuro-muscular condition, cannot be considered independent of the person who has it and their daily life. This makes understanding, measuring and monitoring the impact of neuro-rehabilitation programs on function in everyday life as important as measuring and monitoring spasticity.

Contemporary neuro-rehabilitation uses the ICF as an organising framework. Apart from the environment, there are three domains in the classification—impairment, activity and participation [3]. The ICF defines "impairment" as a problem in body (physiological and psychological) function or structure (anatomical parts of the body), "activity" is the execution of a task or action by an individual and "participation" is involvement in a life situation [3]. In upper limb rehabilitation, function is an important and debated term—because impairments can affect function and it is through function that activity and participation goals can be achieved. The term function is used variably within the literature; it alludes to impairments, activity performance and/or participation in life situations in addition to associations with active task performance. Whilst concepts associated with "function" can vary, operationally, functional use of the spasticity-affected upper limb has been defined by Ashford and Turner-Stokes [4]. That is: active task performance; the affected limb actively completes the task or passive task performance; the task is completed by the affected limb with assistance from the unaffected limb or the task is assisted with or completed by a carer; a key area for spasticity interventions [4]. This three-part operational definition of upper limb function is used in the present study.

Multi-disciplinary person-centred approaches are needed to address rehabilitation needs at impairment, activity and participation levels [5–7]. Neuro-rehabilitation clinical practice guidelines recommend collaborative goal setting [7–10], so that patient preferences and priorities can inform programs. Practice guidelines also recommend the use of standardised assessments to measure impairment, activity and participation dimensions of performance relevant to everyday real life [4, 6, 10, 11]. Although most rehabilitation clinicians measure treatment outcomes [12], evidence suggests that many have limited awareness of the range of assessments available [13]. Those who do use assessment use predominantly impairment-based measures—few use measures that capture activity or participation performance [6, 14]. There are measures available. Although an earlier systematic review of functional outcome measures in the hemiparetic upper limb was conducted in 2008 [15], this study was unable to identify a single valid and reliable outcome measure that captured "real life" function. But a more recent review by Ashford and Turner-Stokes in 2013 [4] identified $n = 27$ functional assessments used in upper limb neuro-rehabilitation for people with spasticity (with and without botulinum toxin-A injection). Their inclusion criteria required the assessments to explore function in the context of everyday real life. As yet, these assessments have not been appraised in relation to the psychometric rigour or clinical utility [4, 16]. This study aims to fill that gap.

This study will use the same $n = 27$ assessments from the Ashford and Turner-Stokes study [4] to investigate the psychometric properties of each and draw conclusions regarding their relative rigour and relevance. A key focus will be the validity of these assessments in their ability to capture change in activity performance and life participation. The ICF will be used as the framework to appraise assessment content to determine the extent to which items address activity and participation domains in addition to impairment (body structures and function). Determining the content validity of items in relation to these domains is important not only to see how valid the assessment is in measuring "health" as it is defined by the ICF but also because these domains reflect common patient goals.

Common neuro-rehabilitation goals for people with upper limb spasticity include reducing pain, increasing the range of movement, preventing contractures and reducing spasticity to enable movement training, splinting or casting [6, 12, 17]. Other goals relate to increasing a person's ability to perform activities and participate in their life situation [6, 11, 16, 18–20]. To date, no systematic review has done this.

Outcomes of this review will be helpful to clinicians and researchers alike working with people who have upper limb spasticity. Right now, some attention has been given to neuro-rehabilitation for people with upper limb spasticity on function in everyday real life [11, 16], but it is a relatively new focus of program evaluation [4, 10, 11] and a challenging one [17, 20, 21]. Determining whether or not interventions impact functional outcomes in everyday life for people with upper limb spasticity has, to date, been complicated by methodological problems, not just in relation to function but also spasticity measurement [15, 22]

and the use of weak study designs [17, 21]. There is a need for more research to show that multidisciplinary upper limb spasticity management rehabilitation programs have an impact on the ability of people to perform activities in everyday real life [11, 23, 24].

The objectives of this systematic review are to locate and appraise existing evidence of the psychometric properties of the outcome measures identified by Ashford and Turner-Stokes [4] and classify those measures to conclude the best tool available for the purpose of measuring activity performance and participation outcomes following upper limb spasticity rehabilitation.

Methods/Design

This systematic review builds on the systematic search conducted by Ashford and Turner-Stokes [4] by synthesising and appraising the research of the psychometric (measurement) properties of outcome measures reported within the published paper. From the 22 studies located in the published search [4], $n = 33$ assessment approaches were identified. On review of those assessment approaches, some were in fact developed for that particular study, for example, three functional tasks (palm hygiene, cutting the fingernails, placing an arm through the sleeve), and consequently do not have published psychometric properties and were excluded from the current study. The remaining $n = 27$ outcome measures had published research investigating their psychometric (measurement) properties and will therefore form the sample for the present study. The authors acknowledge the creation of a degree of assessment selection bias due to this method.

The aims of this systematic review will be:

1. To classify the functional outcome measures reported by Ashford and Turner-Stokes [4] according to whether activity and or participation outcomes following upper limb spasticity rehabilitation are being assessed; activity performance and participation will be defined according to the ICF model [3]; and
2. To locate all of the existing evidence of the properties of the outcome measures, to evaluate the strength of this evidence and to come to a conclusion about the best measure available for the particular purpose of measuring activity and/or participation outcomes following upper limb spasticity rehabilitation.

Publication/study inclusion criteria

1. The aim of the study should be to develop or evaluate the measurement properties of a measurement instrument identified in the review published by Ashford and Turner-Stokes [4];
2. The instrument should aim to measure activity performance or participation, as defined by the ICF [3] Activity performance is defined as "the execution of a task or action by an individual" or requires assistance from or be completed by a carer for the individual. Participation is defined as "involvement in a life situation."
3. The instrument is evaluated in adult patients, over 18 years of age, with upper limb spasticity (as defined by the authors of the included studies) or patients before or after botulinum toxin injection engaging in upper limb rehabilitation programs (with or without the inclusion of botulinum toxin therapy). A rehabilitation program is one that is devised and implemented by a clinician to work towards achievement of identified goals. Participants can be engaging in the rehabilitation program whilst a hospital inpatient, transitioning to home or be community dwelling.
4. All research studies must be original research, and both conducted and published studies in English within peer-reviewed literature will be considered for this review.

Publication/study exclusion criteria

This review is concerned with outcomes of upper limb spasticity rehabilitation that identify changes in the performance of an activity or participation as defined by the ICF [3]. Studies that measure activity performance and participation will be included. Studies that measure upper limb spasticity rehabilitation outcomes through assessment of upper limb impairments only, including pain, range of movement, contracture and changes in tone, will be excluded. Outcomes that have been modified in any manner or implemented in a language other than English will be excluded.

Search methods for the identification of studies

A search will be conducted in Medical Literature Analysis and Retrieval System Online (MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL) and Excerpta Medica dataBASE (EMBASE). In MEDLINE, a validated search filter for finding studies on measurement properties will be used [25]; we report the planned MEDLINE search strategy in Additional file 1. CINAHL and EMBASE search strategies are available from the authors on request. Searches with the names of each included instruments (in the title) in combination with the terms for the study population as described in the search strategy [see Additional file 1] will be conducted until each instrument has been searched.

Screening

Once all searches have been exhausted, the abstracts will be downloaded into the reference management system EndNote and duplicates deleted. A study deemed as a

duplicate will have authors, setting and location, outcome measures implemented, date and duration of study in common.

The eligibility criteria will first be applied to the title and abstract, and if deemed relevant, the full manuscript will be retrieved to determine eligibility of potential studies. The initial screen and selection will be completed independently by the first author with the second author blindly screening a 10 % selection of articles for eligibility. Debate on the inclusion or exclusion of studies will be resolved by an independent third reviewer and discussion between all three reviewers to reach consensus.

Data management

Details on studies that were initially selected based on title and abstract, full-text articles that were retrieved and articles included in the review will be documented. Reasons for the exclusion of retrieved full-text articles, particularly in the case of doubtful articles, will also be recorded.

Data extraction

Data will be extracted from selected studies by the first author utilising a standardised data extraction form. This form will record information related to participants, study design, description of botulinum toxin therapy and rehabilitation program (s), outcome measures administered and their classification according to the ICF (activity performance and or participation focus), psychometric properties, study inclusion/exclusion criteria if available and a brief summary of the findings. The second author will crosscheck all COSMIN ratings.

Risk of bias assessment

Studies evaluating the measurement properties of an assessment require high methodological quality with a low risk bias to guarantee that appropriate conclusions are drawn about the properties of the measure [25]. Thus, it is important to evaluate those methodological qualities [26]. This review will apply the COnsensus-Based Standards for the Selection of Health Status Measurement INstruments (COSMIN) checklist with 4-point scale version [27]. This version is recommended by the COSMIN developers for use in systematic reviews of measurement properties. The checklist will be applied to assess the quality of the papers reporting on the psychometric properties of the 27 outcome measures, evaluating whether each study meets the standards for methodological quality with regard to internal consistency, reliability (test-retest, inter-rater and intra-rater reliability), measurement error, content validity (including face validity), structural validity, hypothesis testing, cross-cultural validity, criterion validity, responsiveness, interpretability and generalisability [27]. The 4-point scale will allow a methodological quality rating of either “excellent”, “good”, “fair” or ”poor” to be assigned to the study [27]. The COSMIN checklist was developed in an international Delphi study with the focus of evaluating the methodological quality of studies on measurement properties [27]. The COSMIN checklist is a modular tool, and the measurement properties evaluated in the study will determine which components or “boxes” need to be completed [27].

Data analysis

Individual assessment items within the outcome measures will be examined to extract meaningful concepts. Those concepts will then be linked to the ICF framework categories of activity performance and or participation following the linking rules suggested by Cieza et al. [28] [see Additional file 2]. This linking process will enable the extent to which outcomes are valid measures of activity performance and life participation to be determined.

The COSMIN checklist with 4-point scale version [27] as described above will be applied to the selected studies, as per COSMIN guidelines, to appraise the overall methodological quality of studies. From here, Terwee’s quality criteria for measurement properties [29] will be applied. Quality criteria for the following nine measurement properties are defined; content validity, internal consistency, criterion validity, construct validity, reproducibility, reliability, responsiveness, floor and ceiling effects and interpretability. This data analysis process will enable conclusions to be drawn regarding the strongest psychometric measure available for the particular purpose of evaluating activity and/or participation outcomes following upper limb rehabilitation. Differences in the psychometric properties of outcome measures for patients with and without upper limb spasticity will be discussed.

Discussion

This systematic review will provide a comprehensive evaluation of the measurement properties of outcome measures assessing activity performance and participation goals for adults with upper limb spasticity undergoing rehabilitation. The results of this review will provide health professionals with detailed information to guide clinical decision-making when choosing the most appropriate outcome measure for occupational performance. Rehabilitation clinicians and managers will also be provided with information to permit accurate assessment and monitoring of the relationship between rehabilitation and health outcome in these patients.

Additional files

Additional file 1: Search strategy. MEDLINE search strategy.

Additional file 2: ICF linking rules v2. Description of how to link concepts to ICF as per linking rules.

Abbreviations
CINAHL: Cumulative Index to Nursing and Allied Health Literature; COSMIN: Consensus-based Standards for the Selection of Health Measurement Instruments; EMBASE: Excerpta Medica dataBASE; ICF: International Classification of Functioning, Disability and Health; MEDLINE: Medical Literature Analysis and Retrieval System Online.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
SP led the design of the protocol and drafting of the manuscript. NL contributed to the design of the protocol and search strategy and provided critical revision of the manuscript. AC contributed to the design of the protocol and critical revision of the manuscript. KW adapted the search strategy from COSMIN for the searches of EMBASE and CINAHL databases. SA and LTS contributed to the selection of assessments for review and critical revision of the manuscript. All authors read and approved the final manuscript.

Authors' information
The study is being conducted as part of SP's program of postgraduate study through La Trobe University under the supervision of Associate Professor Natasha Lannin (La Trobe University) and Professor Anne Cusick (University of Wollongong).

Author details
[1]Department of Occupational Therapy, La Trobe University, Melbourne, Australia. [2]Wagga Wagga Ambulatory Rehabilitation Service, Wagga Wagga Health Service, Wagga Wagga, Australia. [3]Occupational Therapy Department, Alfred Health, Melbourne, Australia. [4]School of Health and Society, University of Wollongong, Wollongong, Australia. [5]Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, Australia. [6]Ageing Work and Health Research Unit, The University of Sydney, Sydney, Australia. [7]The Centre for Excellence in Population Ageing Research, The University of Sydney, Sydney, Australia. [8]Regional Rehabilitation Unit, Northwick Park Hospital, London, UK. [9]Department of Palliative Care, Policy and Rehabilitation, King's College London, School of Medicine, London, UK.

Received: 22 December 2014 Accepted: 8 June 2015
Published online: 13 June 2015

References
1. Lance JW. Symposium synopsis. In: Feldman RG, Young RR, Koella WP, editors. Spasticity: Disordered Motor Control. Miami: Symposia Specialists; 1980. p. 485–94.
2. Pandyan AD, Gregoric M, Barnes MP, Wood D, Van Wijck F, Buridge J, et al. Spasticity: clinical perception, neurological realities and meaningful measurement. Disabil Rehabil. 2005;27(1/2):2–6.
3. World Health Organisation. International Classification of Functioning, Disability and Health. Geneva: World Health Organisation; 2001.
4. Ashford S, Turner-Stokes L. Systematic review of upper-limb function measurement methods in botulinum toxin intervention for focal spasticity. Physiother Res Int. 2013;18(3):178–89. doi:10.1002/pri.1554.
5. Demetrios M, Khan F, Turner-Stokes L, Brand C, McSweeney S. Multidisciplinary rehabilitation following botulinum toxin and other focal intramuscular treatment for post-stroke spasticity (review). Cochrane Database Syst Rev. 2013;6:CD009689. doi:10.1002/14651858.CD009689.pub2.
6. Sheean G, Lannin NA, Turner-Stokes L, Rawicki B, Snow BJ. Botulinum toxin assessment, intervention and after-care for upper limb hypertonicity in adults: international consensus statement. Eur J Neurol. 2010;17 Suppl 2:74–93.
7. Royal College of Physicians, British Society of Rehabilitation Medicine, Chartered Society of Physiotherapy, Association of Chartered Physiotherapists Interested in Neurology. Spasticity in adults: management using botulinum toxin. National guidelines. London: RCP; 2009.
8. Turner-Stokes L. Goal attainment scaling (GAS) in rehabilitation: a practical guide. Clin Rehabil. 2009;23:362–70. doi:10.1177/0269215508101742.
9. Playford ED, Siegert R, Levack W, Freeman J. Areas of consensus and controversy about goal setting in rehabilitation: a conference report. Clin Rehabil. 2009;23:334–44. doi:10.1177/0269215509103506.
10. Rosewilliam S, Roskell CA, Pandyan AD. A systematic review and synthesis of the quantitative and qualitative evidence behind patient-centred goal setting in stroke rehabilitation. Clin Rehabil. 2011;25(6):501–14.
11. Turner-Stokes L, Fheodoroff K, Jacinto J, Maisonobe P. Results from the Upper Limb International Spasticity Study-II (ULISII): a large international, prospective cohort study investigating practice and goal attainment following treatment with botulinum toxin A in real-life clinical management. BMJ Open. 2013;3:e002771.
12. Ward AB, Aguilar M, De Beryl Z, Gedin S, Kanovsky P, Molteni F, et al. Use of botulinum toxin type A in management of adult spasticity—a European consensus statement. J Rehabil Med. 2003;35:98–9.
13. William G, Olver J, de Graaff S, Singer BJ. The use of botulinum toxin type A in the management of adult-onset focal spasticity: a survey of Australian allied health professionals. Aust Occup Ther J. 2012;59:257–64.
14. Bakheit AM, Zakine B, Maisonobe P, Aymard C, Fhedoroff K, Hefter H, et al. The profile of patients and current practice of treatment of upper limb muscle spasticity with botulinum toxin type A: an international survey. Int J Rehabil Res. 2010;33(3):191–204.
15. Ashford S, Slade M, Malaprade F, Turner-Stokes L. Evaluation of functional outcome measures for the hemiparetic upper limb: a systematic review. J Rehabil Med. 2008;40:787–95.
16. Intiso D, Simone V, Di Rienzo F, Santamato A, Russo M, Tolfa M, et al. Does spasticity reduction by botulinum toxin type A improve upper limb functionality in adult post-stroke patients? A systematic review of relevant studies. J Neurol Neurophysiol. 2013;4(4):167.
17. Sheean GL. Botulinum treatment of spasticity: why is it so difficult to show a functional benefit? Curr Opin Neurol. 2001;14:771–6.
18. Elia AE, Filippini G, Calandrella D, Albanese A. Botulinum neurotoxins for post-stroke spasticity in adults: a systematic review. Mov Disord. 2009;24(6):801–12.
19. Cusick A, Lannin NA, Kinnear B: Upper limb spasticity management for patients who have received botulinum toxin A injection: Australian therapy practice. Australian Occupational Therapy Journal, in press.
20. Turner-Stokes L, Baguley IJ, de Graaff S, Katrak P, Davies L, McCrory P, et al. Goal attainment scaling in the evaluation of treatment of upper limb spasticity with botulinum toxin: a secondary analysis from a double-blind placebo controlled randomised clinical trial. J Rehabil Med. 2010;42:81–9.
21. Hinderer SR, Gupta S. Functional outcome measures to assess interventions for spasticity. Arch Phys Med Rehabil. 1996;77:1083–9.
22. Platz T, Eickhof C, Nuyens G, Vuadens P. Clinical scales for the assessment of spasticity, associated phenomena, and function: a systematic review of the literature. Disabil Rehabil. 2005;27(1/2):7–18.
23. Simpson DM, Gracies JM, Graham HK, Miyasaki JM, Naumann M, Russman B, et al. Assessment: Botulinum neurotoxin for the treatment of spasticity (an evidence-based review): report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Neurology. 2008;70:1691–8.
24. Shaw LC, Price CIM, van Wijck FMJ, Shackley P, Steen N, Barnes MP, et al. Botulinum toxin for the upper limb after stroke (BoTULS) trial: effect on impairment, activity limitation, and pain. Stroke. 2011;42:1371–9.
25. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18:1115e23.
26. De Vet HC, Terwee CB, Mokkink LB, Knol DL. Systematic reviews of measurement properties. In: De Vet HCW, Terwee CB, Mokkink LB, Knol DL, editors. Measurement in Medicine: A practical guide. New York: Cambridge University Press; 2011. p. 275–314.
27. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21:651–7. doi:10.1007/s11136-011-9960-1.
28. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. ICF linking rules: an update based on lessons learned. J Rehabil Med. 2005;37:212–8.
29. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34–42.

PLOS ONE

RESEARCH ARTICLE

Psychometric properties of measures of upper limb activity performance in adults with and without spasticity undergoing neurorehabilitation–A systematic review

Shannon Pike[1,2], **Anne Cusick**[3], **Kylie Wales**[4], **Lisa Cameron**[5], **Lynne Turner-Stokes**[6,7], **Stephen Ashford**[6,7,8], **Natasha A. Lannin**[1,5,9]*

1 School of Allied Health, Human Services and Sport (Occupational Therapy), La Trobe University, Melbourne, Victoria, Australia, **2** Wagga Wagga Ambulatory Rehabilitation Service, Murrumbidgee Local Health District, Wagga Wagga, New South Wales, Australia, **3** Discipline of Occupational Therapy, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia, **4** School of Health Sciences, Faculty of Health and Medicine, The University of Newcastle, Newcastle, New South Wales, Australia, **5** Alfred Health, Melbourne, Victoria, Australia, **6** Regional Hyper-acute Rehabilitation Unit, London North West University Healthcare NHS Trust, Northwick Park Hospital, London, United Kingdom, **7** King's College London, Department of Palliative Care, Policy and Rehabilitation, London, United Kingdom, **8** Centre for Nursing, Midwifery and Allied health led Research, University College London Hospitals, National Hospital for Neurology and Neurosurgery, London, United Kingdom, **9** Department of Neurosciences, Central Clinical School, Monash University, Melbourne, Victoria, Australia

* Natasha.Lannin@monash.edu

Citation: Pike S, Cusick A, Wales K, Cameron L, Turner-Stokes L, Ashford S, et al. (2021) Psychometric properties of measures of upper limb activity performance in adults with and without spasticity undergoing neurorehabilitation–A systematic review. PLoS ONE 16(2): e0246288. https://doi.org/10.1371/journal.pone.0246288

Editor: Alessandra Solari, Foundation IRCCS Neurological Institute C. Besta, ITALY

Received: May 24, 2020

Accepted: January 15, 2021

Published: February 11, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0246288

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Abstract

Introduction

This systematic review appraises the measurement quality of tools which assess activity and/or participation in adults with upper limb spasticity arising from neurological impairment, including methodological quality of the psychometric studies. Differences in the measurement quality of the tools for adults with a neurological impairment, but without upper limb spasticity, is also presented.

Methods

29 measurement tools identified in a published review were appraised in this systematic review. For each identified tool, we searched 3 databases (Medline, Embase, CINAHL) to identify psychometric studies completed with neurorehabilitation samples. Methodological quality of instrument evaluations was assessed with use of the Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN) checklist. Synthesis of ratings allowed an overall rating of the psychometric evidence for each measurement tool to be calculated.

Results

149 articles describing the development or evaluation of psychometric properties of 22 activity and/or participation measurement tools were included. Evidence specific to tool use for

Funding: This work was supported by an Australian Government Research Training Program Scholarship (SP); NAL was supported by a Future Leader Fellowship (102055) from the National Heart Foundation of Australia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the mauscript.

Competing interests: The authors have declared that no competing interests exist.

adults with spasticity was identified within only 15 of the 149 articles and provided evidence for 9 measurement tools only. Overall, COSMIN appraisal highlighted a lack of evidence of measurement quality. Synthesis of ratings demonstrated all measures had psychometric weaknesses or gaps in evidence (particularly for use of tools with adults with spasticity).

Conclusions

The systematic search, appraisal and synthesis revealed that currently there is insufficient measurement quality evidence to recommend one tool over another. Notwithstanding this conclusion, newer tools specifically designed for use with people with neurological conditions who have upper limb spasticity, have emergent measurement properties that warrant further research.

Systematic review registration

PROSPERO CRD42014013190.

Introduction

The personal experience of a neurological condition can be profound, impacting on all areas of a person's health and wellbeing. The International Classification for Functioning Disability and Health (ICF) [1] provides a framework to consider the impact of a neurological condition on a person, highlighting both the breadth and complexity of potential issues. While the ICF can classify areas that may be impacted by neurological conditions, and some rating of impairment and limitation is possible using the ICF core sets [2, 3], precise measurement of factors known to be related to activity is essential.

Measurement is key to determining the effect of rehabilitation interventions, and therefore measurement tools used in neurorehabilitation should target all levels of functioning, disability and health–this includes activity and participation as much as impairments in body structure and function [4]. In addition to targeting all levels, measurement should also capture and reflect actual performance of everyday 'real-life' activities outside of the clinical setting [5]. Measurement of activity and participation in 'real-life' activities presents many challenges, not least of which is consistency, validity and sensitivity of 'real life' functions.

Several reviews have sought to identify and determine the most suitable measures to evaluate upper limb impairment and activity for adults with a neurological condition [5–7]. Scant evidence has been located and clear gaps have been identified in the presentation of the psychometric quality of the tools in a neurorehabilitation context. Furthermore, Alt Murphy [6], identified many of the included reviews failed to critically appraise the methodological quality of the individual studies evaluating the psychometric properties of the tools. Whilst recommendations regarding upper limb evaluation have been made, the tools identified and the evidence regarding the psychometric properties of the tools were not specifically targeted nor extracted from a sample of adults with upper limb spasticity as a result of their neurological condition.

Review work by members of this study's authorship team, Ashford and Turner-Stokes, did identify outcome measurement tools both applicable to the upper limb that assess function in the context of everyday life, and from studies including adults with upper limb spasticity [8]. They demonstrated newer upper limb measurement tools used in neurorehabilitation research

which examine activity and participation in the context of everyday real-life activities show promise [8]. There is thus a need for a comprehensive appraisal and synthesis of the psychometric properties of all these tools, to potentially recommend a tool/s for clinical and research use.

The two aims of this study, therefore, was to firstly critically appraise and summarize the quality of the psychometric properties of previously identified upper limb activity performance measurement tools [8] when used with adults with upper limb spasticity using a level of evidence approach and the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines [9–11]. Secondly, to determine if the presence of upper limb spasticity impacts on which measure should be selected based on psychometric evidence, differences in psychometric properties for the identified measurement tools for adults with a neurological impairment but without upper limb spasticity will be defined.

Method

A systematic review with COSMIN appraisal was undertaken, with PRISMA guidelines informing reporting.

Identification and selection of measurement tools

The published list of measurement tools by Ashford and Turner-Stokes [8] was used to identify and select measurement tools for appraisal. The effect of upper limb spasticity on gait is acknowledged [12]. However, we delimit this review to measurement tools that assess upper limb functional movement. As this source systematic review was published in 2013, the most recent clinical guidelines management of spasticity in the upper limb [13] was also searched so as to identify any potential tools that assess upper limb functional movement which may have been developed since 2013. One further tool, the Arm Activity Measure (ArmA), was located and subsequently included in the review.

Measurement tool inclusion criteria

To be included, measurement tools had to assess activity or performance as defined by the ICF [1], and each needed to focus on the upper limb. Activity is defined within the ICF as "the execution of a task or action by an individual" [1, p10] while participation is defined as "involvement in a life situation" [1, p10]. In the present study, the official World Health Organisation (WHO) coding of activity and participation was used, that of a single overlapping list of categories [14]; tools that only evaluate impairment/s (e.g. pain, range of movement, contracture, spasticity) were excluded.

Study search strategy

Searches were completed per protocol [15] to identify research that administered the measurement tool with adults who had neurological conditions. The search was run in Medical Literature Analysis and Retrieval System Online (MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL) and Excerpta Medical database (EMBASE) from inception to December 2016. Where able, the validated search filter for finding studies on measurement properties was used [16]; search terms are presented in S1 File. COSMIN requires information regarding the development/content validity of the measurement tools to be sought, therefore tool references were identified and obtained when not identified within the search results.

Study screening

Title and abstracts were downloaded into the reference management system EndNote™. Duplicates were removed and screened for inclusion by one reviewer. To minimize the risk of incorrect inclusion and exclusion of studies; a second reviewer screened a random 25% sample of included studies against inclusion criteria and all excluded papers were reviewed by the senior author. Disagreements were settled through independent review, followed by discussion until a consensus decision was reached. Full text papers were obtained for all included studies and checked to confirm the final inclusion/exclusion decision [15].

Study inclusion and exclusion criteria

Studies which included participants both with and without spasticity were included; to be included in the spasticity analysis, evidence of the presence of participant upper limb spasticity was required—not just the mention of 'spasticity' in text. For example, the study by Page, Levine and Hade [17] reported a Modified Ashworth Scale score of $\geq$3 as an exclusion criterion; but within the study sample there was no evidence of participants with spasticity $\leq$3. Thus, this article was deemed to be a study without upper limb spasticity. In addition, only studies which tested the measurement tool in its *original and complete* form were included. This conservative approach to study selection was taken to ensure maximum possible homogeneity in the evidence base which would be used to underpin tool recommendations for practice use. If a tool was used as a comparator to validate another tool, the study was excluded in accordance with COSMIN methodology. *Full protocol has been published elsewhere*. Inclusion criteria are detailed in Table 1.

Data analysis

Methodological quality of studies. The quality of the included studies was appraised using the COSMIN taxonomy of measurement properties and definitions for health-related patient reported outcomes [9–11] and the COSMIN Risk of Bias checklist [18] for systematic reviews of patient-reported outcome measures. The methodological quality of each study was individually assessed to evaluate whether it met the standards for measurement tool development, content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for

Table 1. Inclusion criteria.

Design
• Psychometric properties of the identified measurement tools were evaluated
• Original research
• Conducted and published in English within peer reviewed literature
Participants
• Adults (>18 years old)
• $\geq$ 90% diagnosis of a following neurological condition; Stroke, Multiple Sclerosis, Cerebral Palsy, Traumatic Brain Injury, Anoxia
• With or without upper limb spasticity
• Undergoing rehabilitation
Measurement tool
• Measured activity and/or participation
• Nil modifications
• Complete measure administered

https://doi.org/10.1371/journal.pone.0246288.t001

construct validity and responsiveness. The Risk of Bias checklist rated each measurement property as either "very good", "adequate", "doubtful" or "inadequate". As there is no accepted "gold standard" measure of upper limb activity, criterion validity was not evaluated, and construct validity and responsiveness properties were appraised within the hypothesis testing criteria of COSMIN. Where *a priori* hypotheses were not stated, studies were assigned an appropriate generic hypothesis from the list developed by the COSMIN group [18]. Information regarding the interpretability and generalizability were collected.

Quality of measurement properties. The results of individual studies reporting on the psychometric properties were then evaluated using Terwee's quality criteria for measurement properties [9], see S1 File. Results were rated as sufficient '+', indeterminant '?'or insufficient '-'.

Sample size of studies. Sample size was only assessed within individual studies evaluating the measurement properties of content validity, structural validity and cross-cultural validity as per COSMIN guidelines. Sample sizes of individual studies evaluating the remaining measurement properties were not assessed via the Risk of Bias Checklist, and sample sizes per those measurement properties were instead pooled at the synthesis stage [9].

Synthesis of best evidence. All identified evidence and results were then pooled and the modified COSMIN GRADE approach used to determine the overall quality of the evidence [9]. The modified COSMIN GRADE approach considers and downgrades the level of evidence and consequently trustworthiness of results depending on the risk of bias (methodological quality), inconsistency of results, imprecision (based on total sample size) and indirectness (evidence from different populations than the population of interest) [9, p1151]; indirectness was not applicable in this review as studies conducted in samples other than those specified in the inclusion and exclusion criteria were excluded. The synthesis determines either "high", "moderate" "low" or "very low" quality levels of 'sufficient', 'insufficient', 'inconsistent' or 'indeterminant'.

Results

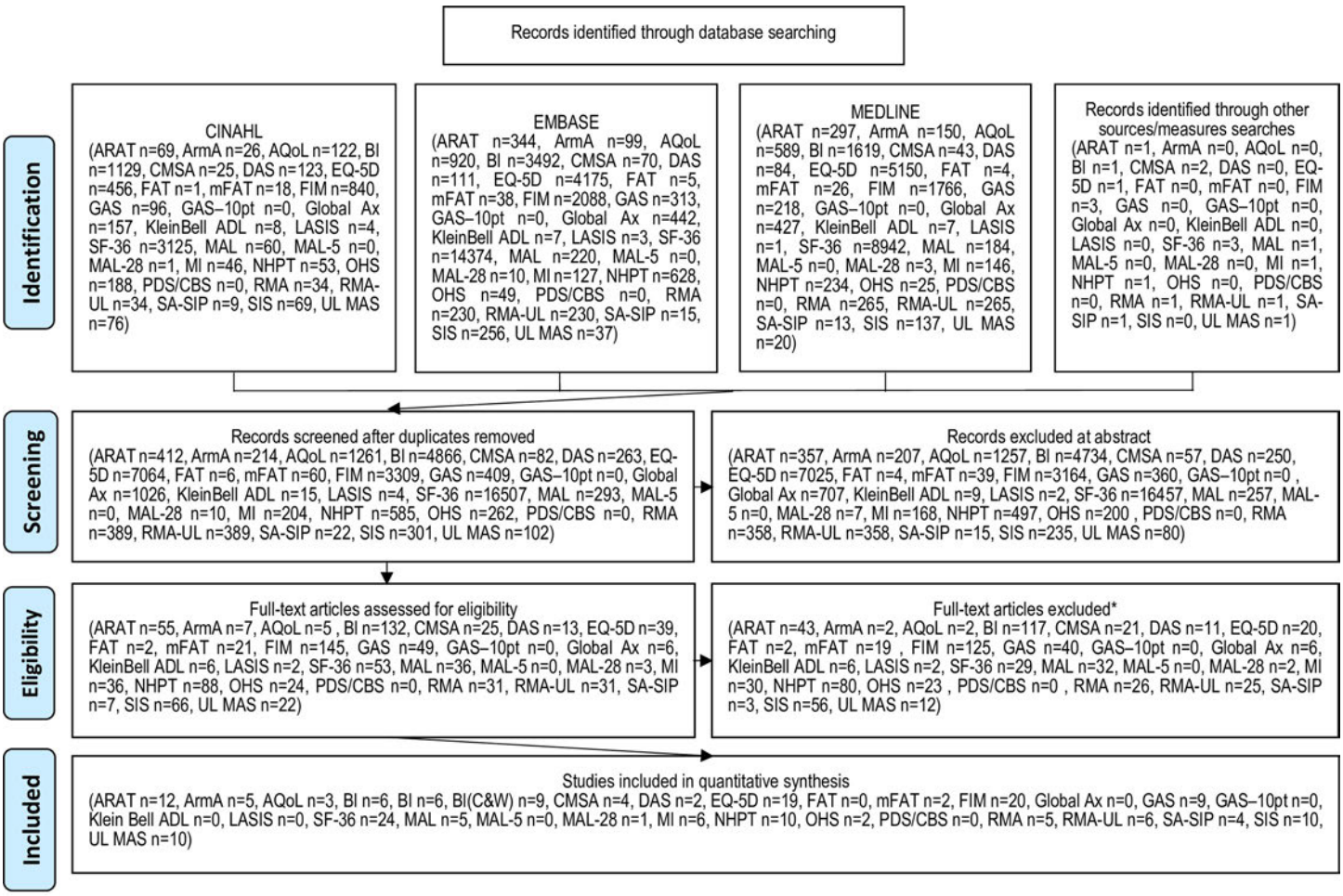
Of the 33 measurement tools identified in the Ashford and Turner-Stokes review [8], 29 measurement tools were published tools. One of the published tools, the Ten Metre Walk Test, was excluded as it does not directly assess upper limb functional movement or use. We therefore completed searches for these 28 tools plus the ArmA (which was identified in the clinical guideline review), resulting in 29 tools in total.

Flow of studies

The electronic search strategy located 55,679 studies across the individual measurement tools. After screening titles, abstracts and full text, 149 psychometric studies (some evaluating more than one included tool) were included in this systematic review. Our systematic search did not locate any studies evaluating the psychometric properties of the following: Frenchay Arm Test [19], Global Assessment Scale [20], Goal Attainment Scale– 10 point scale [21], Klein-Bell Activities of Daily Living Scale [22], Motor Activity Log-5 [23], Leeds Adult Spasticity Impact Scale [24] and Patient Disability Scale/Carer Burden Scale [24]. Fig 1 presents the flow of papers through the review.

Characteristics of the studies

The 149 included studies are outlined in Table 2. The majority of studies (n = 91, 61%) included post-stroke participants, and of these, most were greater than 6 months post-stroke. The remaining studies included diagnoses of multiple sclerosis (MS), traumatic brain injury (TBI) or mixed neurological participants. Sample characteristics varied across studies and



*See Supplementary Table 2 for reasons for full text exclusions. ARAT = Action Research Arm Test, ArmA = Arm Activity Measure, AQoL = Assessment of Quality of Life, BI = Barthel Index, CMSA = Chedoke-McMaster Stroke Assessment, DAS = Disability Assessment Scale, EQ-5D = EuroQol – 5 dimension, FAT = Frenchay Arm Test, mFAT = modified Frenchay Arm Test, FIM = Functional Independence Measure, GAS = Goal Attainment Scale, GAS – 10pt = Goal Attainment Scale – 10 point, Global Ax = Global Assessment Scale, KleinBell ADL = Klein-Bell Activities of Daily Living scale, LASIS = Leeds Adult Spasticity Impact Scale, SF-36 = Medical Outcome Study 36-Item Short-Form Health Survey, MAL = Motor Activity Log, MAL-5 = Motor Activity Log - 5, MAL-28 = Motor Activity Log - 28, MI = Motricity Index, NHPT = Nine Hole Peg Test, OHS = Oxford Handicap Scale, PDS/CBS = Patient Disability Scale / Carer Burden Scale, RMA = Rivermead Motor Assessment, RMA-UL = Rivermead Motor Assessment - Upper Limb, SA-SIP = Stroke-Adapted Version of the Sickness Impact Profile, SIS = Stroke Impact Scale, UL MAS = Upper Limb Motor Assessment Scale.

Fig 1. PRISMA flow chart.

https://doi.org/10.1371/journal.pone.0246288.g001

these are detailed in Table 2; sample sizes were commonly small (range n = 5 to n = 148,367; mean = 2335.24 (SD = 14,431.79); median = 90), with less than 100 in over half of studies (56%) and only n = 5 studies including greater than 10 000 participants. The number of studies evaluating each measurement tool varied, ranging from n = 1 study investigating the Motor Activity Log-28 (MAL-28), to n = 23 for the Medical Outcome Study 36-Item Short-Form Health Survey (SF-36). Participants with upper limb spasticity were specifically identified in n = 15 studies in total (across n = 9 of the included n = 22 measurement tools).

Characteristics of each measurement tool

The number of studies examining each measurement tool is presented, together with findings for all participants and then for participants with upper limb spasticity. The synthesis of

Table 2. Characteristics of included studies.

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Adams et al., (1997) [25]	RMA	Diagnosis = Stroke	Structural validity
	RMA-UL	Time since diagnosis *(mo)* = greater than 6	
		n = 83	
		Age *(yr)*, mean (SD) = Grp 1: 75.39 (6.41), Grp 2: 56.54 (5.73),	
		Grp 3: 56.33 (5.95)	
		Sex, number male (%) = Grp 1: (51), Grp 2: (62), Grp 3: (54)	
		Sample included people with spasticity = not reported	
Adams et al., (1997) [26]	RMA	Diagnosis = Stroke	Structural validity
	RMA-UL	Time since diagnosis *(mo)* = less than 6	
		n = 51	
		Age *(yr)*, mean (SD) = 74.37 (9.38)	
		Sex, number male (%) = 24 (47)	
		Sample included people with spasticity = not reported	
Alderman et al., (2001) [27]	EQ-5D	Diagnosis = Traumatic Brain Injury n = 29, Stroke n = 11	Construct validity
		Time since diagnosis (*mo*) = greater than 6	
		n = 11	
		Age (*yr*), mean (range) = 39 (19–66)	
		Sex, number male (%) = 42 (81)	
		Sample included people with spasticity = not reported	
Ali et al., (2013) [28]	BI	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 3787	
		Age *(yr)*, mean (median IQR) = 71 (60–78)	
		Sex, number male (%) = 2715 (55)	
		Sample included people with spasticity = not reported	
Anderson et al., (1996) [29]	SF-36	Diagnosis = Stroke	Internal consistency
		Time since diagnosis (*mo*) = greater than 6	Construct validity
		n = 90	
		Age (*yr*), mean (SD) = 72 (12)	
		Sex, number male (%) = 48 (53)	
		Sample included people with spasticity = not reported	
Ashford et al., (2015) [30]	ArmA	Diagnosis = Mixed (Stroke n = 15, TBI n = 1)	Content validity
		Time since diagnosis (mo) = greater than 6	
		n = 16	
		Age (yr), mean (SD) = 54.5 (15.7)	
		Sex number male (%) = 9 (56)	
		Sample included people with spasticity = yes	
Ashford et al., (2016) [31]	ArmA	Diagnosis = Mixed (Stroke n = 48, TBI n = 28, MS n = 6, other n = 10)	Structural validity
		Time since diagnosis *(mo)* = not reported	
		n = 92	
		Age *(yr)*, mean (SD) = 44.5 (16.7)	
		Sex number male (%) = 54 (59)	
		Sample included people with spasticity = yes	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Ashford et al., (2014) [32]	ArmA	Diagnosis = Mixed (Stroke n = 30, MS n = 4, TBI n = 22, other n = 2)	Responsiveness
		Time since diagnosis (mo) = not reported	
		n = 58	
		Age (yr), mean (SD) = 47 (17.5)	
		Sex number male (%) = 32 (55)	
		Sample included people with spasticity = yes	
Ashford et al., (2013) [33]	ArmA	Diagnosis = Stroke	Content validity
		Time since diagnosis *(mo)* = not given	
		n = 46 (clinicians), 26 (patient, carers)	
		Age *(yr)*, median (range) = 48.5 (30–64) (patients)	
		Sex, number male (%) = 8 (62) (patients)	
		Sample included people with spasticity = yes	
Ashford et al., (2013) [34]	ArmA	Diagnosis = Mixed (Stroke n = 48, TBI n = 28, MS n = 6, other n = 10)	Internal consistency
		Time since diagnosis *(mo)* = not reported	Reliability
		n = 92	Structural validity
		Age *(yr)*, mean (SD) = 44.5 (16.7)	Construct validity
		Sex, number male (%) = 54 (59)	Responsiveness
		Sample included people with spasticity = yes	Interpretability
Barer & Murphy (1993) [35]	BI (C&W)	Diagnosis = Stroke	Structural validity
		Time since diagnosis (*mo*) = less than 6	Construct validity
		n = 730	Responsiveness
		Age (*yr*), mean (SD) = 73.2 (not given)	
		Sex number male (%) = 336 (46)	
		Sample included people with spasticity = not reported	
Barton et al., (2008) [36]	EQ-5D	Diagnosis = Stroke	Construct validity
		Time since diagnosis (*mo*) = greater than 6	
		n = 62	
		Age $\geq$ 45 years	
		Sex (all sample, not only Stroke), number male (%) = 865 (46.4)	
		Sample included people with spasticity = not reported	
Barton et al., (2008) [37]	EQ-5D	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = not reported	Interpretability
		n = 57	
		Age (all sample, not only Stroke) *(yr)*, mean (range) = 64.7 (45–99)	
		Sex (all sample, not only Stroke), number male (%) = 835 (44.8)	
		Sample included people with spasticity = not reported	
Beebe & Lang (2009) [38]	ARAT	Diagnosis = Stroke	Construct validity
	NHPT	Time since diagnosis *(mo)* = less than 6	Responsiveness
		n = 33	
		Age *(yr)*, mean (SD) = 53.9 (10.2)	
		Sex, number male (%) = 19 (58)	
		Sample included people with spasticity = yes	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Benedict et al., (2011) [39]	NHPT	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = not reported	
		n = 211	
		Age *(yr)*, mean (SD) = 46.2 (8.9)	
		Sex, number male (%) = 32 (27)	
		Sample included people with spasticity = not reported	
Bohannon (1999) [40]	MI	Diagnosis = Stroke	Internal consistency
		Time since diagnosis *(mo)* = less than 6	Construct validity
		n = 10	
		Age (*yr*), mean (range) = 66.7 (46–81)	
		Sex, number male (%) = not given	
		Sample included people with spasticity = not reported	
Bovend'Eerdt et al., (2011) [41]	GAS	Diagnosis = Mixed (Stroke n = 27, TBI n = 1, MS n = 1)	Reliability
		Time since diagnosis *(mo)* = less than 6	Measurement error
		n = 29	
		Age *(yr)*, mean (SD) = 50.28 (13.88)	
		Sex, number male (%) = 18 (62)	
		Sample included people with spasticity = not reported	
Brashear et al., (2002) [42]	DAS	Diagnosis = Stroke	Reliability
		Time since diagnosis *(mo)* = greater than 6	Content validity
		n = 10 raters	
		Age *(yr)*, mean (SD) = 59.9 (16.17)	
		Sex, number male (%) = 5 (56)	
		Sample included people with spasticity = yes	
Brock et al., (2009) [43]	GAS	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 45 patients 23 carers	
		Age *(yr)*, median (range) = 66 (35–87)	
		Sex, number male (%) = (56)	
		Sample included people with spasticity = not reported	
Brown et al., (2015) [44]	FIM	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	Interpretability
		n = 148 367	
		Age *(yr)*, mean (SD) = 70.6 (13.1)	
		Sex, number male (%) = 71,726 (48)	
		Sample included people with spasticity = not reported	
Burridge et al., (2009) [45]	ARAT	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 17	
		Age *(yr)*, mean (SD) = 57 (13.4)	
		Sex, number male (%) = 11 (65)	
		Sample included people with spasticity = yes	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Carr et al., (1985) [46]	UL-MAS	Diagnosis = Stroke	Reliability
		Time since diagnosis *(mo)* = less than 6	Content validity
		n = 5	
		Age *(yr)*, mean (range) = 65 (55–78)	
		Sex, number male (%) = 1 (20)	
		Sample included people with spasticity = not reported	
Chen et al., (2012) [47]	MAL	Diagnosis = Stroke	Measurement error
		Time since diagnosis *(mo)* = 3–9	Interpretability
		n = 116	
		Age *(yr)*, range = Intervention grp 60.98 (13.47)	
		Control grp 63.26 (12.56)	
		Sex, number male (%) = Intervention grp 69 (65)	
		Control grp 73 (63)	
		Sample included people with spasticity = not reported	
Collin & Wade (1990) [48]	MI	Diagnosis = Stroke	Reliability
	RMA–UL	Time since diagnosis *(mo)* = less than 6	Construct validity
		n = 20 (reliability), n = 14 (concurrent validity)	
		Age *(yr)* mean (range) = 56.1 (15–77)	
		Sex number male (%) = 24 (67)	
		Sample included people with spasticity = not reported	
Collin et al., (1988) [49]	BI (C&W)	Diagnosis = Mixed (Stroke n = 13, Traumatic Brain Injury n = 11, other n = 1)	Reliability
		Time since diagnosis (*mo*) = less than 6	Content validity
		n = 25	
		Age (*yr*), range = 12–66	
		Sex number male (%) = 124 (52)	
		Sample included people with spasticity = not reported	
Corrigan et al., (1997) [50]	FIM	Diagnosis = Traumatic Brain Injury	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 95	
		Age *(yr)*, mean (SD) = 35.2 (not given)	
		Sex, number male (%) = 67 (70)	
		Sample included people with spasticity = not reported	
Costelloe et al., (2008) [51]	NHPT	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = not reported	Interpretability
		n = 150	
		Age *(yr)*, mean (SD) = not given	
		Sex, number male (%) = not given	
		Sample included people with spasticity = not reported	
Cullen et al., (2014) [52]	FIM	Diagnosis = Traumatic Brain Injury	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 59	
		Age *(yr)*, mean (SD) = drivers 49.77 (15.25)	
		non-driver 51.42 (15.73)	
		Sex, number male (%) = driver 28 (80) non-driver 19 (79)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Cuthbert et al., (2015) [53]	FIM	Diagnosis = Traumatic Brain Injury	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 64081	
		Age *(yr)*, mean = 76% less than 80	
		Sex, number male (%) = 41204 (64.3)	
		Sample included people with spasticity = not reported	
Dang et al., (2011) [54]	CMSA	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 74	
		Age *(yr)*, mean (SD) = 65.3 (12.4)	
		Sex, number male (%) = 48 (65)	
		Sample included people with spasticity = not reported	
Demeurisse et al., (1980) [55]	MI	Diagnosis = Stroke	Content validity
		Time since diagnosis *(mo)* = less than 6	
		n = 100	
		Age *(yr)*, mean (SD) = 69 (not reported)	
		Sex, number male (%) = 59 (59)	
		Sample included people with spasticity = not reported	
Dennis et al., (2000) [56]	BI (C&W)	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 417	
		Age *(yr)*, mean (SD) = 64.6 (not given)	
		Sex number male (%) = not reported	
		Sample included people with spasticity = not reported	
De Weerdt et al., (1985) [57]	ARAT	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	Responsiveness
		n = 53	
		Age *(yr)*, mean (SD) = 68.6 (9.3)	
		Sex, number male (%) = 25 (47)	
		Sample included people with spasticity = not reported	
Doan et al., (2012) [58]	DAS	Diagnosis = Stroke	Construct validity
	EQ-5D	Time since diagnosis *(mo)* = greater than 6	
	SA-SIP30	n = 279	
		Age *(yr)*, mean (range) = 58.2 (21–88)	
		Sex, number male (%) = 150 (54)	
		Sample included people with spasticity = yes	
Doig et al., (2010) [59]	GAS	Diagnosis = Traumatic Brain Injury	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Responsiveness
		n = 14	
		Age *(yr)*, range = 18–57	
		Sex, number male (%) = 12 (86)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Dorman et al., (1999) [60]	SF-36	Diagnosis = Stroke	Construct validity
	EQ-5D	Time since diagnosis *(mo)* = greater than 6	Interpretability
		n = 531	
		Age *(yr)*, mean (SD) = not reported	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	
Dorman et al., (1998) [61]	SF-36	Diagnosis = Stroke	Internal consistency
	EQ-5D	Time since diagnosis *(mo)* = greater than 6	Reliability
		n = 209	
		Age *(yr)*, mean = 70	
		Sex, number male (%) = 147 (54)	
		Sample included people with spasticity = not reported	
Dorman et al., (1997) [62]	EQ-5D	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = not reported	
		n = 152	
		Age % of sample by group <50 = 5%, 50–70 = 46%, >70 = 49%.	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	
Dromerick et al., (2006) [63]	ARAT	Diagnosis = Stroke	Construct validity
	MAL	Time since diagnosis *(mo)* = less than 6	Interpretability
		n = 39	
		Age *(yr)*, mean (SD) = 64.54 (14.13)	
		Sex, number male (%) = 17 (44)	
		Sample included people with spasticity = not reported	
Duncan et al., (2003) [64]	SIS	Diagnosis = Stroke	Content validity
		Time since diagnosis *(mo)* = less than 6	Structural validity
		n = 696	
		Age *(yr)*, mean (SD) = 68.6 (12.5)	
		Sex, number male (%) = 386 (55)	
		Sample included people with spasticity = not reported	
Duncan et al., (2002) [65]	SIS	Diagnosis = Stroke	Reliability
		Time since diagnosis *(mo)* = less than 6	Construct validity
		n = 287	
		Age *(yr)*, mean (SD) = 72.6 (10), 59.8 (15.5)	
		Sex, number male (%) = 135 (47), 78 27.2)	
		Sample included people with spasticity = not reported	
Duncan et al., (2005) [66]	SIS	Diagnosis = Stroke	Internal consistency
		Time since diagnosis *(mo)* = less than 6	Reliability
		n = 26	
		Age *(yr)*, mean (SD) = mail sample 68.48 (11.4)	
		telephone sample 68.84 (12.2)	
		Sex, number male (%) = mail sample 219 (97.8)	
		telephone sample 230 (98.3)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Duncan et al., (1997) [67]	SF-36	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 200	
		Age *(yr)*, mean (SD) = 63 (13)	
		Sex, number male (%) = 164 (54)	
		Sample included people with spasticity = not reported	
Duncan et al., (1999) [68]	SIS	Diagnosis = Stroke	Content validity
		Time since diagnosis *(mo)* = less than 6	
		n = 91	
		Age *(yr)*, mean (SD) = minor stroke 69.2 (10.1)	
		moderate stroke 71.9 (11.7)	
		Sex, number male (%) = 42 (46)	
		Sample included people with spasticity = not reported	
Edwards et al., (2006) [69]	SA-SIP30	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 219	
		Age *(yr)*, mean (SD) = 64.74 (15.87)	
		Sex, number male (%) = 94 (43)	
		Sample included people with spasticity = not reported	
Egan et al., (2014) [70]	FIM	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 55	
		Age *(yr)*, mean (SD) = 64.8 (13.3)	
		Sex, number male (%) = 39 (58)	
		Sample included people with spasticity = not reported	
Eriksson et al., (2013) Eriksson, Baum [71]	SIS	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Interpretability
		n = 116	
		Age *(yr)*, mean (SD) = 62.4 (12.7)	
		Sex number male (%) = 56 (48)	
		Sample included people with spasticity = not reported	
Filiatrault et al., (1991) [72]	BI	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	Responsiveness
		n = 18	
		Age *(yr)*, mean (SD) = 52.2 (13.5)	
		Sex number male (%) = 12 (67)	
		Sample included people with spasticity = not reported	
Fisk et al., (2005) [73]	EQ-5D	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis = not given	
		n = 187	
		Age *(yr)*, mean (SD) = 51 (10)	
		Sex, number male (%) = 47 (25)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Findler et al., (2001) [74]	SF-36	Diagnosis = Traumatic Brain Injury	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 326	
		Age (yr), mean (SD) = 41.7 (10.8) mild, 35.7 (9.8) moderate-severe	
		Sex, number male (%) = 130 (88)	
		Sample included people with spasticity = not reported	
Fleming et al., (2014) [75]	ARAT	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Interpretability
		n = 33	
		Age *(yr)*, mean (SD) = 61.5 (14.2)	
		Sex, number male (%) = 20 (61)	
		Sample included people with spasticity = yes	
Freeman et al., (2000) [76]	SF-36	Diagnosis = Multiple Sclerosis	Internal consistency
		Time since diagnosis *(mo)* = greater than 6	Construct validity
		n = 149	Responsiveness
		Age (yr), mean (SD) = 44.6 (10.8)	Interpretability
		Sex, number male (%) = (32)	
		Sample included people with spasticity = not reported	
Freeman et al., (1996) [77]	SF-36	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Interpretability
		n = 50	
		Age *(yr)*, mean (SD) = 44.8 (9.8)	
		Sex, number male (%) = 21 (42)	
		Sample included people with spasticity = not reported	
Gillard et al., (2015) [78]	EQ-5D	Diagnosis = Stroke	Construct validity
		Time points since diagnosis *(mo)* = greater than 6	
		n = 460	
		Age *(yr)*, mean (SD) = 67 (14)	
		Sex, number male (%) = 241 (52)	
		Sample included people with spasticity = yes	
Goodkin et al., (1988) [79]	NHPT	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Interpretability
		n = Exp 68, Control 21	
		Age *(yr)*, mean (SD) = Exp 47.16 (11.3) Control 45.24 (16.50)	
		Sex number male (%) = Exp 25 (37) Control 7 (33)	
		Sample included people with spasticity = not reported	
Gowland 1990 [80]	CMSA	Diagnosis = Stroke	Content validity
		Time since diagnosis *(mo)* = not reported	
		n = not reported	
		Age *(yr)*, mean (range) = not reported	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Gowland et al., (1993) [81]	CMSA	Diagnosis = Stroke	Reliability
		Time since diagnosis *(mo)* = less than 6	Construct validity
		n = 32	Responsiveness
		Age *(yr)*, mean (range) = 64, (18–86)	
		Sex, number male (%) = 14 (44)	
		Sample included people with spasticity = not reported	
Grant et al., (2014) [82]	FIM	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 11983	
		Age *(yr)*, median (25^{th}, 75^{th} percentile) = 72 (61, 81)	
		Sex, number male (%) = 6581 (55)	
		Sample included people with spasticity = not reported	
Green et al., (2001) [83]	BI (C&W)	Diagnosis = Stroke	Reliability
		Time since diagnosis *(mo)* = greater than 6	Measurement error
		n = 22	
		Age *(yr)*, mean (SD) = 71.6 (6.8)	
		Sex number male (%) = 16 (73)	
		Sample included people with spasticity = not reported	
Guilfoyle et al., (2010) [84]	SF-36	Diagnosis = Traumatic Brain Injury	Internal consistency
		Time since diagnosis *(mo)* = mixed, mean less than 6	Structural validity
		n = 453	Construct validity
		Age *(yr)*, mean (SD) = 36.6 (16.1)	Interpretability
		Sex, number male (%) = 392 (76.3)	
		Sample included people with spasticity = not reported	
Hagen et al., (2003) [85]	SF-36	Diagnosis = Stroke	Internal consistency
		Time since diagnosis *(mo)* = less than 6	Construct validity
		n = 136	Responsiveness
		Age *(yr)*, mean (SD) = 70 (11)	Interpretability
		Sex, number male (%) = 69 (51)	
		Sample included people with spasticity = not reported	
Hall et al., (1993) [86]	FIM	Diagnosis = Traumatic Brain Injury	Structural validity
		Time since diagnosis *(mo)* = less than 6	Construct validity
		n = 332	Interpretability
		Age *(yr)*, mean (SD) = 34.5 (16)	
		Sex, number male (%) = 259 (78)	
		Sample included people with spasticity = not reported	
Hamilton & Granger (1994) [87]	FIM	Diagnosis = Stroke	Reliability
		Time since diagnosis *(mo)* = less than 6	
		n = 1018	
		Age *(yr)*, mean (SD) = 71 (12)	
		Sex, number male (%) = 478 (47)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Harris & Eng (2007) [88]	MAL	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 93	
		Age *(yr)*, mean (SD) = 68.7 (9.4)	
		Sex, number male (%) = 61 (65)	
		Sample included people with spasticity = yes	
Hawthorne et al., (2009) [89]	AQoL	Diagnosis = Traumatic Brain Injury	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 56	
		Age *(yr)*, mean (SD) = 39 (15)	
		Sex, number male (%) = 40 (71)	
		Sample included people with spasticity = not reported	
Hawthorne et al., (1999) [90]	AQoL	Diagnosis = Mixed (medical and musculoskeletal diagnoses, healthy samples)	Content validity
		Time since diagnosis *(mo)* = less than 6	
		n = 255	
		Age *(yr)*, range = ≤29–70+	
		Sex, number male (%) = 121 (47)	
		Sample included people with spasticity = not reported	
Heinemann et al., (1997) [91]	FIM	Diagnosis = Traumatic Brain Injury	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 129	
		Age *(yr)*, mean (SD) = 37.4 (19.5)	
		Sex, number male (%) = (71)	
		Sample included people with spasticity = not reported	
Heinemann et al., (1993) [92]	FIM	Diagnosis = Mixed (Stroke n = 10092)	Structural validity
		Time since diagnosis *(mo)* = less than 6	
		n = 10092	
		Age *(yr)*, mean (SD) = 62.1 (not given) whole sample	
		Sex, number male (%) = 5349 (53) whole sample	
		Sample included people with spasticity = not reported	
Heinemann et al., (1994) [93]	FIM	Diagnosis = Mixed (Stroke n = 9961)	Structural validity
		Time since diagnosis *(mo)* = less than 6	
		n = 9961	
		Age *(yr)*, mean (SD) = 70.4 (not reported)	
		Sex, number male (%) = 4781 (48)	
		Sample included people with spasticity = not reported	
Heller et al., (1987) [94]	mFAT	Diagnosis = Stroke	Reliability
	NHPT	Time since diagnosis *(mo)* = greater than 6	
		n = 10	
		Age (yr) = not provided	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Heller et al., (1987) [94]	mFAT	Diagnosis = Stroke	Construct validity
	NHPT	Time since diagnosis *(mo)* = less than 6	Interpretability
		n = 56	
		Age (yr) = 68.1 (11.4)	
		Sex, number male (%) = 24 (43)	
		Sample included people with spasticity = not reported	
Hermann et al., (1996) [95]	SF-36	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 85	
		Age *(yr)*, mean (SD) = 44.6 ()	
		Sex, number male (%) = 20 (23)	
		Sample included people with spasticity = not reported	
Hobart et al., (2002) [96]	SF-36	Diagnosis = Stroke	Internal consistency
		Time since diagnosis *(mo)* = less than 6	Structural validity
		n = 177	Interpretability
		Age *(yr)*, mean (SD) = 62 (13)	
		Sex, number male (%) = 126 (71)	
		Sample included people with spasticity = not reported	
Houlden et al., (2006) [97]	FIM	Diagnosis = Mixed (Stroke n = 261, Traumatic Brain Injury n = 107)	Responsiveness
	BI (C&W)	Time since diagnosis (*mo*) = less than 6	Interpretability
		n = 368	
		Age (*yr*), mean (SD) = whole sample not reported	
		Sex number male (%) = 259 (63)	
		Sample included people with spasticity = not reported	
Jacob-Lloyd et al., (2005) [98]	MI	Diagnosis = Stroke	Construct validity
	NHPT	Time since diagnosis *(mo)* = less than 6	Responsiveness
		n = 58	Interpretability
		Age *(yr)* number (%) = 47 (85) older than 60	
		Sex, number male (%) = 31 (53)	
		Sample included people with spasticity = not reported	
Jenkinson et al., (2013) [99]	SIS	Diagnosis = Stroke	Internal consistency
		Time since diagnosis *(mo)* = greater than 6	Structural validity
		n = 73	
		Age *(yr)* range = 18 - >75	
		Sex, number male (%) = 88 (58)	
		Sample included people with spasticity = not reported	
Johnson & Selfe (2004) [100]	UL-MAS	Diagnosis = Stroke	Internal consistency
		Time since diagnosis *(mo)* = less than 6	
		n = 26	
		Age *(yr)* mean (SD) = 77 (9)	
		Sex, number male (%) = 13 (50)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Jones (1998) [101]	RMA	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 29	
		Age *(yr)* mean (SD) = 66 (9.4)	
		Sex, number male (%) = 13 (50)	
		Sample included people with spasticity = not reported	
Joyce et al., (1994) [102]	GAS	Diagnosis = Traumatic Brain Injury	Reliability
		Time since diagnosis *(mo)* = less than 6	Content validity
		n = 16	Construct validity
		Age *(yr)* mean (range) = 27 (17–49)	
		Sex, number male (%) = 9 (56)	
		Sample included people with spasticity = not reported	
Khan et al., (2013) [103]	UL-MAS	Diagnosis = Stroke	Structural validity
		Time since diagnosis *(mo)* = less than 6	Construct validity
		n = 481	
		Age *(yr)* range = 18–101	
		Sex, number male (%) = 255 (53)	
		Sample included people with spasticity = not reported	
Khan et al., (2008) [104]	GAS	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Responsiveness
		n = 24 (203 goals)	
		Age *(yr)* mean (SD) = 52 (8.3)	
		Sex, number male (%) = 10 (42)	
		Sample included people with spasticity = not reported	
Keith et al., (1987) [105]	FIM	Diagnosis = not reported	Content validity
		Time since diagnosis *(mo)* = not reported	
		n = not reported	
		Age *(yr)*, mean (SD) = not reported	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	
Kohn et al., (2014) [106]	EQ-5D	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Responsiveness
		n = 3044	
		Age *(yr)*, mean (SD) = 56.8 (9.9)	
		Sex, number male (%) = 600 (20)	
		Sample included people with spasticity = not reported	
Kuspinar et al (2014) [107]	EQ-5D	Diagnosis = MS	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 189	
		Age *(yr)*, mean (SD) = 43 (10)	
		Sex, number male (%) = 49 (26)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Kuspinar & Mayo (2013) [108]	EQ-5D	Diagnosis = Multiple Sclerosis	Content validity
		Time since diagnosis *(mo)* = greater than 6	Construct validity
		n = 185	
		Age *(yr)*, mean (SD) = 42.8 (10)	
		Sex, number male (%) = 48 (26)	
		Sample included people with spasticity = not reported	
Kuys et al., (2009) [109]	FIM	Diagnosis = Stroke	Construct validity
	UL-MAS	Time since diagnosis *(mo)* = less than 6	
		n = 105	
		Age *(yr)* median = 70 (13)	
		Sex, number male (%) = 64 (53)	
		Sample included people with spasticity = not reported	
Kwon et al., (2006) [110]	SIS	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	Interpretability
		n = 95	
		Age *(yr)* median = 70 (13)	
		Sex, number male (%) = 64 (53)	
		Sample included people with spasticity = not reported	
Kwon et al., (2004) [111]	BI	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 1680	
		Age *(yr)*, mean (SD) = 70 (11.4)	
		Sex number male (%) = 790 (47)	
		Sample included people with spasticity = not reported	
Lai et al., (2002) [112]	SIS	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	Interpretability
		n = 81	
		Age *(yr)*, mean (SD) = 76 (6.56)	
		Sex number male (%) = 48 (59)	
		Sample included people with spasticity = not reported	
Lang et al., (2008) [113]	ARAT	Diagnosis = Stroke	Interpretability
		Time since diagnosis *(mo)* = less than 6	
		n = 12	
		Age *(yr)*, mean (SD) = 64 (14)	
		Sex, number male (%) = 21 (40)	
		Sample included people with spasticity = not reported	
Lang et al., (2006) [114]	ARAT	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	Responsiveness
		n = 50	
		Age *(yr)*, mean (SD) = 63.7 (13.6)	
		Sex, number male (%) = 21 (42)	
		Sample included people with spasticity = yes	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Lannin (2003) [115]	GAS	Diagnosis = mixed (Stroke, Traumatic Brain Injury)	Responsiveness
		Time since diagnosis *(mo)* = greater than 6	
		n = 12	
		Age *(yr)*, mean (range) = 56.5 (26–79)	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	
Lannin (2004) [116]	UL-MAS	Diagnosis = Stroke	Internal consistency
		Time since diagnosis *(mo)* = less than 6	Structural validity
		n = 27	
		Age *(yr)*, mean (SD) = 67 (10.1)	
		Sex, number male (%) = 15 (50)	
		Sample included people with spasticity = not reported	
Lincoln & Leadbitter (1979) [117]	RMA	Diagnosis = Stroke	Content validity
		Time since diagnosis *(mo)* = not reported	
		n = 51	
		Age *(yr)*, range = 17–65	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	
Loewen & Anderson (1988) [118]	UL-MAS	Diagnosis = Stroke	Reliability
		Time since diagnosis *(mo)* = less than 6	
		n = 7	
		Age *(yr)*, mean (SD) = 73.6 (8.3)	
		Sex, number male (%) = 2 (29)	
		Sample included people with spasticity = not reported	
Loewen & Anderson (1990) [119]	UL-MAS	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 50	
		Age *(yr)*, mean (SD) = 68 (10)	
		Sex, number male (%) = 28 (56)	
		Sample included people with spasticity = not reported	
Lyle (1981) [120]	ARAT	Diagnosis = Mixed (Stroke n = unknown, Traumatic Brain Injury n = unknown)	Content validity
		Time since diagnosis *(mo)* = Greater than 6)	Structural validity
		n = 20	
		Age *(yr)*, mean (range) = 53.2 (26–72)	
		Sex, number male (%) = 13 (65)	
		Sample included people with spasticity = not reported	
Mackenzie et al., (2002) [121]	SF-36	Diagnosis = Traumatic Brain Injury	Structural validity
		Time since diagnosis *(mo)* = greater than 6	Construct validity
		n = 1197	
		Age *(yr)*, range = 18–54	
		Sex, number male (%) = 790 (66)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Madden et al., (2006) [122]	SF-36	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	Responsiveness
		n = 116	Interpretability
		Age *(yr)*, mean (range) = 70 (10)	
		Sex, number male (%) = 57 (49)	
		Sample included people with spasticity = not reported	
Mahoney & Barthel (1965) [123]	BI	Diagnosis = not given	Content validity
		Time since diagnosis *(mo)* = not given	
		n = not given	
		Age *(yr)*, mean (range) = not given	
		Sex, number male (%) = not given	
		Sample included people with spasticity = not reported	
Malec (1999) [124]	GAS	Diagnosis = Mixed (Traumatic Brain Injury n = 66, Stroke n = 15, other n = 7)	Construct validity
		Time since diagnosis *(mo)* = greater than 6 (61%)	
		n = 88	
		Age *(yr)*, mean (range) = 33.8 (18–69)	
		Sex number male (%) = 64 (72.7)	
		Sample included people with spasticity = not reported	
Malec et al., (1991) [125]	GAS	Diagnosis = Traumatic Brain Injury	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 14	
		Age *(yr)*, mean (SD) = 34.3 (12.2)	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	
Miller et al., (2010) [126]	UL-MAS	Diagnosis = Stroke	Internal consistency
		Time since diagnosis *(mo)* = less than 6	Structural validity
		n = 80	Construct validity
		Age *(yr)*, mean (SD) = 67.4 (15.6)	Interpretability
		Sex, number male (%) = 46 (58)	
		Sample included people with spasticity = not reported	
Moore et al., (2004) [127]	SF-36	Diagnosis = Multiple Sclerosis	Construct validity
	EQ-5D	Time since diagnosis *(mo)* = greater than 6	
		n = 114	
		Age *(yr)*, mean (SD) = 45 (11)	
		Sex, number male (%) = 18 (45)	
		Sample included people with spasticity = not reported	
Moreland et al., (1993) [128]	CMSA	Diagnosis = Stroke	Content validity
		Time since diagnosis *(mo)* = not reported	
		n = not reported	
		Age *(yr)*, median (range) = not reported	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Morris et al., (2013) [129]	ARAT	Diagnosis = Stroke	Construct validity
	NHPT	Time since diagnosis *(mo)* = greater than 6	Interpretability
	RMA–UL	n = 85	
		Age *(yr)*, median (range) = 69 (36–88)	
		Sex, number male (%) = 49 (58)	
		Sample included people with spasticity = not reported	
Murrell et al., (1999) [130]	SF-36	Diagnosis = Multiple Sclerosis	Reliability
		Time since diagnosis *(mo)* = greater than 6	
		n = 22	
		Age *(yr)*, mean (SD) = 52.4 (9.9)	
		Sex, number male (%) = 9 (40)	
		Sample included people with spasticity = not reported	
Nicholl et al., (2001) [131]	EQ-5D	Diagnosis = Multiple Sclerosis	Construct validity
		Time points since diagnosis *(mo)* = greater than 6	Interpretability
		n = 88	
		Age *(yr)*, mean (SD) = 48.97 (8.9)	
		Sex, number male (%) = 24 (25)	
		Sample included people with spasticity = not reported	
Oczkowski et al., (1993) [132]	FIM	Diagnosis = Stroke	Construct validity
		Time since diagnosis (*mo*) = less than 6	
		n = 113	
		Age *(yr)*, mean = 65.7 (female) 65.8 (male)	
		Sex, number male (%) = 59 (52.2)	
		Sample included people with spasticity = not reported	
O'Mahony et al., (1998) [133]	SF-36	Diagnosis = Stroke	Interpretability
		Time since diagnosis *(mo)* = not reported	
		n = 104	
		Age *(yr)*, mean (range) = > 45	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	
Ouellette et al., (2015) [134]	FIM	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 407	
		Age *(yr)*, mean (SD) = 68.2 (13.9)	
		Sex, number male (%) = not given	
		Sample included people with spasticity = not reported	
Peters et al., (2014) [135]	EQ-5D	Diagnosis = Stroke	Responsiveness
		Time since diagnosis *(mo)* = not reported	
		n = 102	
		Age *(yr)* = 78% > 55	
		Sex, number male (%) = 53 (53)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Pickard et al., (2005) [136]	EQ-5D	Diagnosis = Stroke	Responsiveness
		Time points since diagnosis *(mo)* = less than 6	Interpretability
		n = 96	
		Age *(yr)*, mean (SD) = 67 (15)	
		Sex, number male (%) = 51 (52)	
		Sample included people with spasticity = not reported	
Pickering et al., (2010) [137]	UL-MAS	Diagnosis = Stroke	Structural validity
		Time since diagnosis *(mo)* = less than 6	Interpretability
		n = 25	
		Age *(yr)*, mean (SD) = 69.96 (11.97)	
		Sex, number male (%) = 14 (56)	
		Sample included people with spasticity = not reported	
Pittock et al., (2004) [138]	SF-36	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 185	
		Age *(yr)*, mean (SD) = not given	
		Sex, number male (%) = 56 (30)	
		Sample included people with spasticity = not reported	
Poole et al., (2010) [139]	NHPT	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 56	
		Age *(yr)*, mean (SD) = 46.8 (10.48)	
		Sex, number male (%) = 11 (20)	
		Sample included people with spasticity = not reported	
Rabadi & Rabadi (2006) [140]	ARAT	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	Responsiveness
		n = 104	
		Age *(yr)*, mean (SD) = 72.0 (13)	
		Sex, number male (%) = 43 (41)	
		Sample included people with spasticity = not reported	
Rabadi & Vincent (2013) [141]	FIM	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Responsiveness
		n = 76	
		Age *(yr)*, mean (SD) = 53.6 (10.9)	
		Sex, number male (%) = 63 (83)	
		Sample included people with spasticity = yes	
Rand & Eng (2015) [142]	ARAT	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 32	
		Age *(yr)*, mean (SD) = 58.1 (12.4)	
		Sex, number male (%) = 25 (78)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Riazi et al., (2003) [143]	SF-36	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 638	
		Age *(yr)*, range = 20 - >60	
		Sex, number male (%) = 219 (35)	
		Sample included people with spasticity = not reported	
Rigby et al., (2009) [144]	OHS	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 104	
		Age *(yr)*, mean (SD) = 72.0 (13)	
		Sex, number male (%) = 43 (41)	
		Sample included people with spasticity = not reported	
Robinson et al (2009) [145]	SF-36	Diagnosis = MS	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Interpretability
		n = 249	
		Age *(yr)*, mean (range) = 39 (10.5)	
		Sex, number male (%) = 75 (30)	
		Sample included people with spasticity = not reported	
Sabari et al., (2005) [146]	UL-MAS	Diagnosis = Stroke	Structural validity
		Time since diagnosis *(mo)* = less than 6 (83%)	Interpretability
		n = 100	
		Age *(yr)*, mean (range) = 54 (18–94)	
		Sex, number male (%) = 67 (67)	
		Sample included people with spasticity = not reported	
Sackley (1990) [147]	RMA	Diagnosis = Stroke	Construct validity
	RMA-UL	Time since diagnosis *(mo)* = less than 6	
		n = 52 (R hemiparesis), 38 (L hemiparesis)	
		Age *(yr)*, mean (SD) = 63.4 (11.4) (R hemiparesis),	
		63.2 (11.9) (L hemiparesis)	
		Sex, number male (%) = 33 (64) (R hemiparesis),	
		23 (61) (L hemiparesis)	
		Sample included people with spasticity = not reported	
Salter et al., (2008) [148]	SF-36	Diagnosis = Stroke	Content validity
	EQ-5D	Time since diagnosis (*mo*) = not reported	
	SIS	n = not reported	
		Age (*yr*), mean (SD) = not reported	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	
Sarker et al., (2012) [149]	BI (C&W)	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Interpretability
		n = 238	
		Age *(yr)*, mean (SD) = 68.6 (14.2)	
		Sex number male (%) = 124 (52)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Schwid et al., (2002) [150]	NHPT	Diagnosis = Multiple Sclerosis	Measurement error
		Time since diagnosis = unknown	
		n = 27	
		Age *(yr)*, mean (SD) = 51.9 (9.0)	
		Sex, number male (%) = 16 (79)	
		Sample included people with spasticity = not reported	
Sharrack et al., (1999) [151]	BI (C&W)	Diagnosis = Multiple Sclerosis	Internal consistency
	FIM	Time since diagnosis *(mo)* = greater than 6	Reliability
		n = 25–64	Structural validity
		Age *(yr)*, median (range) = 40 (42.1–77.6)	Construct validity
		Sex, number male (%) = 22 (34)	Responsiveness
		Sample included people with spasticity = not reported	
Simon et al., (2008) [152]	OHS	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 53	
		Age *(yr)*, mean (SD) = 65.6 (12.1)	
		Sex, number male (%) = 14 (28)	
		Sample included people with spasticity = not reported	
Stineman et al., (1996) [153]	FIM	Diagnosis = mixed (Stroke = 26, 183, Traumatic Brain Injury = 3, 214)	Internal consistency
		Time since diagnosis *(mo)* = less than 6	Structural validity
		n = 29 397	
		Age *(yr)*, mean range = 41.6–71.3	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	
Stone et al., (1993) [154]	MI	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 84	
		Age *(yr)*, mean (SD) = 72.37 (12.11)	
		Sex, number male (%) = not given	
		Sample included people with spasticity = not reported	
Sturm et al., (2002) [155]	AQoL	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	Interpretability
		n = 93	
		Age *(yr)*, mean (range) = 72 (28–89)	
		Sex, number male (%) = 42 (45)	
		Sample included people with spasticity = not reported	
Turner-Stokes et al., (2010) [156]	GAS	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 90	
		Age *(yr)*, mean (SD) = 54.5 (13.2)	
		Sex, number male (%) = 54 (60)	
		Sample included people with spasticity = yes	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Uswatte & Taub (2005) [157]	MAL	Diagnosis = not reported	Content validity
		Time since diagnosis *(mo)* = not reported	
		n = not reported	
		Age *(yr)*, mean (SD) = not reported	
		Sex number male (%) = not reported	
		Sample included people with spasticity = not reported	
Uswatte et al., (2006) [158]	MAL	Diagnosis = Stroke	Internal consistency
	MAL-28	Time since diagnosis *(mo)* = greater than 6	Reliability
		n = 222	Content validity
		Age *(yr)*, mean (SD) = 62.2 (13.0)	Structural validity
		Sex number male (%) = 142 (64)	Interpretability
		Sample included people with spasticity = not reported	
Van der Putten et al., (1999) [159]	BI (C&W)	Diagnosis = Mixed (Stroke n = 82, Multiple Sclerosis n = 201)	Responsiveness
	FIM	Time since diagnosis *(mo)* = less than 6	Interpretability
		n = 283	
		Age *(yr)*, mean (SD) = 52 (16.9) (Stroke),	
		45 (11.2) (Multiple Sclerosis)	
		Sex number male (%) = 238 (84)	
		Sample included people with spasticity = not reported	
Van Straten et al (1997) [160]	SA-SIP30	Diagnosis = Stroke	Content validity
		Time since diagnosis *(mo)* = less than 6	
		n = 319	
		Age *(yr)*, mean (SD) = 69 (12.6)	
		Sex number male (%) = 175 (55)	
		Sample included people with spasticity = not reported	
Vickrey et al., (1997) [161]	SF-36	Diagnosis = Multiple Sclerosis	Internal consistency
		Time since diagnosis *(mo)* = greater than 6	Reliability
		n = 171 (internal consistency, hypothesis testing),	Construct validity
		n = 84 (reliability)	
		Age *(yr)*, mean (range) = 45 (20–67)	
		Sex, number male (%) = 123 (72)	
		Sample included people with spasticity = not reported	
Vickrey et al., (1995) [162]	SF-36	Diagnosis = Multiple Sclerosis	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 179	
		Age *(yr)*, mean (range) = 45 (20–67)	
		Sex, number male (%) = 129 (72)	
		Sample included people with spasticity = not reported	
Wade & Hewer (1987) [163]	BI (C&W)	Diagnosis = Stroke	Structural validity
	MI	Time since diagnosis *(mo)* = less than 6	Construct validity
		n = 976	
		Age *(yr)*, mean (SD) = not given	
		Sex, number male (%) = not given	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Wallace et al., (2002) [164]	BI	Diagnosis = Stroke	Responsiveness
		Time since diagnosis *(mo)* = less than 6	
		n = 372	
		Age *(yr)*, mean (SD) = 69.7 (11.6)	
		Sex number male (%) = 177 (48)	
		Sample included people with spasticity = not reported	
Ware & Sherbourne (1992) [165]	SF-36	Diagnosis = not reported	Content validity
		Time since diagnosis *(mo)* = not reported	
		n = not reported	
		Age *(yr)*, mean (SD) = not reported	
		Sex number male (%) = not reported	
		Sample included people with spasticity = not reported	
Wellwood et al., (1995) [166]	BI	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Interpretability
		n = 152	
		Age *(yr)*, mean (SD) = 73 (13.4)	
		Sex number male (%) = 68 (45)	
		Sample included people with spasticity = not reported	
Wilkinson et al., (1997) [167]	BI (C&W)	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	Interpretability
		n = 106	
		Age *(yr)*, median (range) = 71 (34–79)	
		Sex number male (%) = 57 (54)	
		Sample included people with spasticity = not reported	
Williams et al., (1999) [168]	SF-36	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = less than 6	
		n = 71	
		Age *(yr)*, mean (SD) = 61 (13)	
		Sex, number male (%) = 45 (63)	
		Sample included people with spasticity = not reported	
Williams (1990) [169]	EQ-5D	Diagnosis = not reported	Content validity
		Time since diagnosis *(mo)* = not reported	
		n = not reported	
		Age *(yr)*, mean (SD) = not reported	
		Sex, number male (%) = not reported	
		Sample included people with spasticity = not reported	
Wolf & Koster et al., (2013) [170]	SIS	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = greater than 6	
		n = 96	
		Age *(yr)*, median (range) = Grp 1 64.2 (13.4), Grp 2 60.5 (12.8)	
		Sex, number male (%) = Grp 1 28 (52), Grp 2 31 (55)	
		Sample included people with spasticity = not reported	

(*Continued*)

Table 2. (Continued)

Studies included	Measurement tool	Summary of study participants	Psychometric property tested
Xie et al., (2006) [171]	EQ-5D	Diagnosis = Stroke	Construct validity
		Time since diagnosis *(mo)* = not reported	
		n = 1040	
		Age *(yr)* = ≥18	
		Sex, number male (%) = 447 (43.9)	
		Sample included people with spasticity = not reported	
Yozbatiran et al., (2008) [172]	ARAT	Diagnosis = Stroke	Reliability
		Time since diagnosis *(mo)* = greater than 6	Construct validity
		n = 12 (validity) n = 9 (interrater reliability) n = 8 (intra rater)	
		Age *(yr)*, mean (SD) = 61.0 (15.0)	
		Sex, number male (%) = 6 (50)	
		Sample included people with spasticity = not reported	
		Rater characteristics	
		Rater n = 2 Clinical experience *(yr)* = 8	
		Observations n = 58	

RMA = Rivermead Motor Assessment, RMA-UL = Rivermead Motor Assessment–Upper Limb, BI (C&W) = Barthel Index Collin & Wade version, EQ-5D = EuroQol -5 dimension, SIS = Stroke Impact Scale, SF-36 = Medical Outcome Study 36-Item Short-Form Health Survey, ArmA = Arm Activity Measure, ARAT = Action Research Arm Test, NHPT = Nine Hole Peg Test, MI = Motricity Index, GAS = Goal Attainment Scale, DAS = Disability Assessment Scale, FIM = Functional Independence Measure, UL-MAS = Upper Limb–Motor Assessment Scale, CMSA = Chedoke-McMaster Stroke Assessment, SA-SIP30 = Stroke-Adapted Version of the Sickness Impact Profile, MAL = Motor Activity Log, BI = Barthel Index, AQoL = Assessment of Quality of Life, mFAT = modified Frenchay Arm Test, OHS = Oxford Handicap Scale, MAL-28 = Motor Activity log– 28.

https://doi.org/10.1371/journal.pone.0246288.t002

evidence for each measurement tools is presented in Table 3. Due to the volume of data, summaries of individual study results and psychometric properties tested are tabulated within S2 and S3 Tables. The following summarizes the appraisal of each tool. *These have been placed in alphabetical order*.

Action Research Arm Test. The Action Research Arm Test (ARAT) [173] is an obervational performance test that evaluates a person's ability to use their upper limb to handle objects using grasp, grip, pinch and gross motor movements. Twelve studies evaluated the psychometric properties of the ARAT [38, 45, 57, 63, 75, 113, 114, 120, 129, 140, 142, 172], four of those studies specifically identified participants with upper limb spasticity [38, 45, 75, 114]. The majority of studies included participants post-stroke with a single study including a mixed sample, post-stroke and TBI [120].

Content validity. The Upper Extremity Function Test (UEFT) [174] was modified by Lyle [173] to produce the ARAT. No further content validity studies were identified. The ARAT was found to have sufficient relevance, but indeterminant ratings for comprehensiveness and comprehensibility and no participants were interviewed regarding those properties.

Results for whole sample. Research supports hierarchical ordering of items [173] and reliability within (ICC = 0.99) and between raters (ICC 0.99) [172]. The ARAT was found to correlate highly with other like-tests of activity and dexterity (r = 0.65–0.95) [57, 63, 129, 140, 142, 172] and weak to moderately with the Functional Independence Measure (FIM), a more global measure of function (r = 0.47) [140]. ARAT scores were not, however, a predictor of overall quality of life [129]. The ARAT was found to be responsive over time in acute as well as chronic stroke and TBI samples [38, 57, 114, 140]. ARAT was found to be equally sensitive to

Table 3. Synthesis of evidence.

Measurement tool	Sample	Content validity	Structural validity	Internal consistency	Cross cultural validity	Reliability			Measurement error	Construct validity	Responsiveness
						Inter	Intra	Retest			
ARAT	*Spasticity*									*Moderate*	*Low*
	n = 4									*- (13/21)*	*+ (4/4)*
	Whole sample	Very Low	Very Low			Very Low	Very Low			Moderate	Moderate
	n = 12		+			+	+			- (19/30)	+ (6/6)
ArmA	*Spasticity*	*High*	*High*	*Moderate*				*Low*		*Very Low*	*Moderate*
	n = 5		+	+				+		+	*+ (4/4)*
	Whole sample	High	High	Moderate				Low		Very Low	Moderate
	n = 5		+	+				+		+	+ (4/4)
AQoL	*Spasticity*										
	n = 0										
	Whole sample	Very Low								High	
	n = 3									+ (3/3)	
BI	*Spasticity*										
	n = 0										
	Whole sample	Very Low								High	Very Low
	n = 6									+ (5/6)	- (0/1)
BI (C&W)	*Spasticity*										
	n = 0										
	Whole sample	Very Low	Low			Very Low		Very Low	Very Low	Moderate	Low
	n = 9		+			?		?	+	+	- (2/3)
CMSA	*Spasticity*										
	n = 0										
	Whole sample n = 4	Very Low				Moderate +	Moderate	Low		Moderate	Very Low
						Low +*	+	+		+ (5/6)	+ (1/1)
DAS	*Spasticity*	*Very Low*				*Low*	*Low*			*Moderate*	
	n = 2					?	-			*+ (2/2)*	
	Whole sample	Very Low				Low	Low			Moderate	
	n = 2					?	-			+ (2/2)	
EQ-5D	*Spasticity*									*High*	
	n = 2									*+ (3/3)*	
	Whole sample	Moderate						Moderate +^		Moderate	Low
	n = 19	?						Very Low - ^^		+ (24/34)	- (11/15)
FAT	*Spasticity*										
	n = 0										
	Whole sample										
	n = 0										

(*Continued*)

Table 3. (Continued)

Measurement tool	Sample	Content validity	Structural validity	Internal consistency	Cross cultural validity	Reliability			Measurement error	Construct validity	Responsiveness
						Inter	Intra	Retest			
mFAT	*Spasticity*										
	n = 0										
	Whole sample					Very Low		Very Low		Very Low	
	n = 2					?		?		- (0/1)	
FIM	*Spasticity*									*Moderate*	*Very Low*
	n = 1									*+ (1/1)*	*+ (1/1)*
	Whole sample	Very Low	High	High		Moderate	Low			High	Moderate
	n = 20		+	+		+	+			+ (23/29)	- (5/7)
Global Ax	*Spasticity*										
	n = 0										
	Whole sample										
	n = 0										
GAS	*Spasticity*									*Very Low*	
	n = 1									*– (3/7)*	
	Whole sample					Low			Low	Moderate	Low
	n = 9					-			?	– (14/23)	+ (4/4)
GAS-10pt	*Spasticity*										
	n = 0										
	Whole sample										
	n = 0										
Klein-Bell	*Spasticity*										
	n = 0										
	Whole sample										
	n = 0										
LASIS	*Spasticity*										
	n = 0										
	Whole sample										
	n = 0										
MAL	*Spasticity*									*Low*	
	n = 1									*- (3/7)*	
	Whole sample	Very Low	Very Low						Low	Moderate	
	n = 5		?						?	- (4/9)	
MAL-5	*Spasticity*										
	n = 0										
	Whole sample										
	n = 0										

(*Continued*)

Table 3. (Continued)

Measurement tool	Sample	Content validity	Structural validity	Internal consistency	Cross cultural validity	Reliability			Measurement error	Construct validity	Responsiveness
						Inter	Intra	Retest			
MAL-28	*Spasticity*										
	n = 0										
	Whole sample	Very Low	Very Low	Very Low				Moderate +^		Very Low + (3/4)^	
	n = 1		?	+**				Low -^^		Very Low–(2/4)^^	
MI	*Spasticity*										
	n = 0										
	Whole sample	Very Low		Very Low		Very Low				Moderate	Very Low
	n = 6			?		?				- (4/6)	- (0/1)
NHPT	*Spasticity*									*Very Low*	*Very Low*
	n = 1									*- (3/5)*	*+ (2/2)*
	Whole sample					Very Low		Very Low	Very Low	Moderate	Low
	n = 10					?		?	+	- (21/32)	+ (3/3)
OHS	*Spasticity*										
	n = 0										
	Whole sample									Low	
	n = 2									- (2/3)	
PDS / CBS	*Spasticity*										
	n = 0										
	Whole sample										
	n = 0										
RMA	*Spasticity*										
	n = 0										
	Whole sample	Very Low	Very Low							High	
	n = 5		-							+ (2/2)	
RMA–UL	*Spasticity*										
	n = 0										
	Whole sample	Very Low	Very Low							High	
	n = 6		+, - ^^^							+ (3/4)	
SF-36	*Spasticity*										
	n = 0										
	Whole sample	Very Low	Moderate	High				Moderate +^		Moderate–(25/44)	Very Low–(0/4)
	n = 24		?	+				Low -^^			
SA-SIP	*Spasticity*									*Moderate*	
	n = 1									*+ (1/1)*	
	Whole sample	Moderate								High	
	n = 4									+ (3/3)	

(*Continued*)

Table 3. (Continued)

Measurement tool	Sample	Content validity	Structural validity	Internal consistency	Cross cultural validity	Reliability			Measurement error	Construct validity	Responsiveness
						Inter	Intra	Retest			
SIS	*Spasticity*										
	n = 0										
	Whole sample n = 10	Moderate	High	Moderate		Low		Low		High	
			+	+		?		+		+ (18/19)	
UL-MAS	*Spasticity*										
	n = 0										
	Whole sample n = 10	Very Low	Moderate	Moderate		Low	Low			Moderate	
			+	+**		?	?			- (3/8)	

High = Very confident that the true measurement property lies close to that of the estimate of the measurement property. *Moderate* = Moderate confidence in the measurement property estimate. *Low* = Limited confidence in the measurement property estimate. *Very low* = Little confidence in the measurement property estimate, full definition of ratings reported in [9]. + = sufficient,—insufficient,? indeterminant [9].

*Moderate + Impairment Inventory, Low + Activity Inventory

**Internal consistency evidence strength cannot exceed structural validity as per COSMIN guidelines and has been reduced accordingly.

^Patients reports

^^ proxy reports

^^^ '+' acute sample, '-' subacute sample.

ARAT = Action Research Arm Test, ArmA = Arm Activity Measure, AQoL = Assessment of Quality of Life, BI = Barthel Index, BI (C&W) = Barthel Index—Collin & Wade version, CMSA = Chedoke-McMaster Stroke Assessment, DAS = Disability Assessment Scale, EQ-5D = EuroQol– 5 dimension, FAT = Frenchay Arm Test, mFAT = modified Frenchay Arm Test, FIM = Functional Independence Measure, GAS = Goal Attainment Scale, GAS– 10pt = Goal Attainment Scale– 10 point, Global Ax = Global Assessment Scale, KleinBell ADL = Klein-Bell Activities of Daily Living scale, LASIS = Leeds Adult Spasticity Impact Scale, SF-36 = Medical Outcome Study 36-Item Short-Form Health Survey, MAL = Motor Activity Log, MAL-5 = Motor Activity Log—5, MAL-28 = Motor Activity Log—28, MI = Motricity Index, NHPT = Nine Hole Peg Test, OHS = Oxford Handicap Scale, PDS/CBS = Patient Disability Scale / Carer Burden Scale, RMA = Rivermead Motor Assessment, RMA-UL = Rivermead Motor Assessment—Upper Limb, SA-SIP = Stroke-Adapted Version of the Sickness Impact Profile, SIS = Stroke Impact Scale, UL MAS = Upper Limb Motor Assessment Scale.

https://doi.org/10.1371/journal.pone.0246288.t003

change as like measures when used with participants less than 6 months post-stroke [57, 140]. Mixed results have been reported with respect to ceiling effect in stroke populations [63, 75] and there is one study which has reported a minimal, clinically important change of 12 points (dominant) and 17 (non-dominant) [113].

Results pertaining to sample with upper limb spasticity. The ARAT correlated strongly with like measures of activity and dexterity (r = 0.69–0.95) [38] and less with a global measure of function (Functional Independence Measure (FIM) r = 0.2–0.6) [114] and impairments, including grip and pinch strength, spasticity and AROM (r = - 0.28–0.86) [38, 45, 114]. The ARAT was moderate to highly responsive to capture change in participants less than 6 months post-stroke (ES = 0.55–1.018) [38, 114], being as equally responsive as like measures (NHPT and Jebsen-Taylor test of hand function), more responsive than measures of impairment (pinch and grip strength), but less responsive than the SIS-Hand (ES = 0.55–1.018) [38]. Neither a floor nor ceiling effects were found in a sample of participants greater than 6 months post-stroke [75].

Arm Activity measure. The Arm Activity measure (ArmA) is a 20-item self-report tool which includes 7 passive and 13 active items to capture real arm activity in neurological populations [33]. Five studies [30–34] evaluated the psychometric properties of the ArmA, the

majority of studies included a mixed sample including participants post-stroke, TBI and MS. All included studies specifically identified participants with upper limb spasticity.

Content validity. The ArmA was developed based on goal analysis, systematic literature review and a modified Delphi survey which demonstrated relevance, comprehensiveness and comprehensibility [30, 33].

Results pertaining to sample with upper limb spasticity. The ArmA subscales demonstrated internal consistency (passive subscale $\alpha = 0.85$, active subscale $\alpha = 0.96$) and retest reliability (quadratic weight kappa 0.90 (CI 0.68–1.12), active subscale 0.93 (CI 0.71–1.15)) in a sample with upper limb spasticity [34]. The ArmA demonstrated convergent and divergent validity with passive and active items of the Leeds Adult Spasticity Scale (LASIS) and Disabilities of Arm Shoulder and Hand (DASH) (convergent: Rho 0.48; p = 0.01 to 0.63; p = 0.01; divergent: Rho 0.02; p = 0.9 to 0.23; p = 0.078) [34] and was found to be responsive [32, 34]. Preliminary analysis suggests clinically meaningful change is indicated by 2.5 or 3 point improvement (passive subscale) and 1.1 or 2.5 point improvement (active subscale) [34]. The ArmA active function subscale suffered a ceiling effect (37%), however no floor effect was observed for either subscale [34].

Assessment of Quality of Life. The Assessment of Quality of Life (AQoL) is a generic HRQoL measure that assesses independent living, social relationships, physical senses, psychological wellbeing and illness [90]. Three studies evaluated the psychometric properties of the AQoL, one included participants greater than 6 months post TBI [89] and two less than 6 months post-stroke [90, 154]. Neither study specifically identified participants with upper limb spasticity.

Content validity. Development research underpinning the AQoL [90] demonstrated sufficient relevance, but indeterminant ratings for comprehensiveness and comprehensibility. No other content validity studies conducted in a neurological sample were identified.

Results for whole sample. The AQoL discriminated between participants with and without TBI (effect size (ES) = 0.80), with participants post TBI scoring 2.0 utilities lower than participants without [89]. The AQoL correlated more strongly with measures of handicap (London Handicap Scale (LHS) r = 0.83) than disability (Barthel Index (BI) r = 0.77) or impairment (National Institute of Health Stroke Scale (NIHSS) r = -0.69) in the first 6 months post-stroke and was a significant predictors of death or institutionalization at 12 months [155]. No floor or ceiling effects (1–2%) were found in a stroke population [155].

Barthel Index. The Barthel Index (BI) was initially developed to score the abilities of participants to care for themselves [123]. The BI evaluates 10 activity areas, with a maximum score of 100 indicating independence in all included areas. Six studies evaluated the psychometric properties of the BI [28, 72, 111, 123, 164, 166]. Five studies were completed with participants post-stroke, 4 included participants less than 6 months post-stroke [28, 72, 111, 164], 1 greater than 6 months post-stroke [166] and 1 discussed tool development with a non-specific sample [123]. No included studies specifically identified participants with upper limb spasticity.

Content validity. No research on the development of the BI was located.

Results for whole sample. The BI correlated moderately with measures of upper limb function (Fugl-Meyer Rho = 0.60) (Functional Test for the Hemiplegic/Paretic Upper Limb Rho = 0.61) [72] and global measures function (FIM $r_s = 0.95$, $p<0.0001$; Modified Rankin Scale (MRS) $r_s = 0.89$, $p<0.0001$; Office of Population Censuses and Surveys (OPCS) disability instrument r = 0.73, $p<0.001$) [111, 166]. The BI was equally responsive to change within the first three months post-stroke as like global measures (FIM) [164] and a measure of motor function (Fugl-Meyer Test) [72], however determined responsiveness was low. Evidence of a ceiling effect was found in a sample greater than 6 months post-stroke [166].

Barthel Index (Collin & Wade). The Barthel Activities of Daily Living Index (BI C&W) [49] is a modification of the original BI measurement tool, with all 10 areas of activity included but is scored in increments of 1 rather than 5 as per the original BI [123]. Nine studies evaluated the psychometric properties of the BI(C&W) [35, 49, 56, 83, 97, 149, 159, 163, 167], 6 studies included participants post-stroke [35, 56, 83, 149, 163, 167] and 3 included mixed samples (stroke, MS, TBI) [49, 97, 159]. No studies specifically identified participants with upper limb spasticity.

Content validity. No information presenting the methodology used to revise the original BI was found, only justification from revised test authors who felt the original five-point incremental scoring was misleading in accuracy [49].

Results for whole sample. Research supports use of a summed BI(C&W) score due to a single factor (68% of variance) underlying the scale [163]. While the hierarchical nature of the BI (C&W) was supported by Wade and Hewer [163], Barer and Murphy [35] reported a failure to meet Guttman scaling criteria. Test-retest reliability results appear mixed, with high agreement (75%) between scores but variations in kappa (-0.99 to 0.81) [83]. Inter-rater reliability between self-report, family, nursing staff and skilled observers was acceptable (agreement within 2 points or less for 72% of participants) [49]. The BI(C&W) was strongly associated with measures of upper limb activity (r = 0.729–0.826) (Motricity Index Upper Limb (MI UL) and Motricity Index (MI) total, Frenchay Activity Index (FAI)), complex daily activities ($r \geq 0.80$), and disability ($r_s = 0.726–0.80$) (London Handicap Scale, Modified Rankin Scale (MRS)), and less with measures of psychological wellbeing and impairments (depression, anxiety, pain) (r = 0.2–0.423) [56, 149, 163, 167]. Research suggests that BI(C&W) is at least equally responsive to FIM [97, 159]. However, BI(C&W) suffered from floor and ceiling effects across the acute through to community continuum in a mixed neurorehabilitation sample [97, 149, 159, 167].

Chedoke-McMaster Stroke Assessment. The Chedoke-McMaster Stroke assessment (CMSA) is comprised of two parts; the impairment inventory and the activity inventory (formerly known as the disability inventory) [81]. The CMSA impairment inventory classifies participants into subgroups based on the stages of motor recovery, while the CMSA activity inventory provides a measure of activity performance. Four studies evaluated the psychometric properties of the CMSA, two included participants less than 6 months post-stroke[54, 81], two did not report on the length of time post-stroke for participants [80, 128] and no study specifically identified participants with upper limb spasticity.

Content validity. Evidence located for the development of the CMSA [80, 128], did not indicate participants were consulted on the comprehensiveness or comprehensibility of included items. Relevance of items for the intended purpose of assessment of stroke clients within rehabilitation setting was sufficient, however further content validity studies were not identified.

Results for whole sample. Evidence supports the reliability of the CMSA; inter-rater (ICC 0.88 (95% CI 0.76–0.94) to 0.99 (95%CI 0.98–1.00)), intra-rater (ICC 0.93 (95% CI 0.85–0.96) to 0.98 (95% CI 0.95–0.99)), test retest (ICC 0.98 (95% CI 0.95–0.99)) [81]. Consistent with the definition of the CMSA, strong correlations with both subscales and total scores for like measures of upper limb activity performance (Fugl-Meyer r = 0.95, $p<0.001$) and global measures of function (FIM r = 0.79, $p<0.05$) were demonstrated [81]. The predictive validity through use of the Gowland's predictive equations, however, were not supported due to large error associated with the predicted value [54]. The CMSA was found to be more responsive than the FIM when used with participants less than 6 months post-stroke [81].

Disability Assessment Scale. The Disability Assessment Scale (DAS) is a brief measure of functional disability [42]. Two studies were included, both identified participants with upper limb spasticity [42, 58].

Content validity. Brashear and colleagues [42] reported the development of the DAS to fill the identified gap within the evaluation of functional impairment commonly seen in participants with post-stroke upper limb spasticity (i.e. dressing, hygiene, limb position, pain). No additional research underpinning measurement tool development was reported.

Results pertaining to sample with upper limb spasticity identified. Good to excellent intra-rater reliability (78% of evaluations weighted kappa $\geq$ .4) and good inter-rater reliability (Kendall W 0.49 (95% CI 0.30–1.00, $p < .001$) to 0.77 (95% CI 0.37–1.00, $p < .001$) was reported when used by professionals (neurologists, physiatrists, occupational therapists and physical therapists) with a mean of 6 years clinical experience [42]. Greater DAS scores were found to be associated with Stroke-Adapted Version of the Sickness Impact Scale (SA-SIP) scores ($P < .05$), reduced quality of life and caregiver burden ($P < .05$) [58, 175].

EuroQol-5 dimension. The EuroQol-5 dimension (EQ-5D) is a generic measure of health-related quality of life [73, 78, 169]. Nineteen studies evaluated the psychometric properties of the EQ-5D, including participants with MS (n = 6), [73, 106–108, 127, 131] a mixed neurological sample (n = 1) [27] and post-stroke (n = 12) [36, 37, 58, 60–62, 78, 135, 136, 148, 171]. Two studies specifically identified participants with upper limb spasticity [58, 78].

Content validity. During the development of the EQ-5D there is no evidence that participants were consulted on the comprehensiveness or comprehensibility of included items. Relevance of items for the intended purpose was sufficient [108]. The EQ-5D contains 6 of 9 recommended dimensions for patient-based, health related quality of life measures and is less comprehensive than the Stroke Impact Scale (SIS) [148].

Results for whole sample. Test-retest reliability of the patient-reported EQ-5D was moderate to good for VAS and the mobility domain (ICC $\geq$0.70) [61, 73], test-retest reliability was lower in proxy-reported scores [61]. The EQ-5D correlated moderately with global measures of function such as the EDSS (r = -0.66) [73], but was less sensitive than disease-specific quality of life scales and the generic SF-36 when used with participants with MS [131]. A single study found a moderate inverse relationship between the EQ-5D and the Nine Hole Peg Test, a specific measure of upper limb use (r = -0.56) [73]. When used with participants post-stroke, the EQ-5D correlated with global measures of function including the SF-6D, a classification for describing health from a selection of SF-36 items (r = 0.77) [37] and the SF-36 (r = 0.57–0.63) [60]. Evidence of the discriminant ability was found between participants post-stroke and those who had not suffered a stroke [36, 171], between stroke type and severity [62], and between participants with and without spasticity [78]. The EQ-5D Index had the greatest change score when compared to like generic HRQoL measures less than 6 months post-stroke [136], was more responsive to changes in disability (MRS r = -0.36) and daily activities (BI r = 0.57) in comparison to the EQ-5D VAS [136]. Contrarily, neither the EQ-5D Index or VAS was responsive to change over a one year period post-stroke despite 23.8% of participants reporting improvement and 23.2% deterioration [135]. The EQ-5D did not demonstrate either floor and ceiling effects when used with acute participants post-stroke [136].

Results pertaining to sample with upper limb spasticity identified. The EQ-5D index scores were found to correlate with measures of disability ($p < .002$) and carer burden ($p < .05$) [58] and to distinguish between participants with and without upper limb spasticity post-stroke, with mean differences (-0.07, 95% CI -0.12 to -0.33) equivalent to the MCID established for the EQ-5D for other health conditions (MCID is yet to be established for post-stroke populations) [78].

Modified Frenchay Arm Test. The modified Frenchay Arm Test (mFAT), reduces the 25 clinical tests to 5 so as to measure arm function after stroke [94]. Two studies evaluated the psychometric properties of the mFAT [94]; no studies specifically identified participants with upper limb spasticity.

Content validity. No studies were identified providing information targeting measurement tool development and/or content validity.

Results for whole sample. There was evidence for the reliability of the mFAT (inter-rater (Rho = 0.75–0.99), test-retest (Rho = 0.68–0.90 and 0.83–0.99)) when administered to participants 18 months post-stroke [94]. The mFAT was found to be less sensitive than the NHPT in participants less than 6 months post-stroke with mild impairments [94]. Floor effects (30%) and ceiling effects (34%) were evident within acute stroke [94].

Functional independence measure. A total of 20 studies evaluated the psychometric properties, in participants post-stroke (n = 9) [44, 70, 82, 87, 92, 93, 111, 132, 134], TBI (n = 5) [50, 52, 53, 86, 91], MS (n = 2) [141, 151] and a mixed neurological sample (n = 3) [97, 153, 159]. One study specifically identified participants with upper limb spasticity in a sample with MS [141].

Content validity. The FIM was found to have sufficient relevance, but indeterminant ratings for comprehensiveness and comprehensibility during development, as nil information was located to determine if participants were interviewed regarding those properties [105].

Results pertaining to whole sample. A two factor structure was identified for the FIM by a number of researchers, with separate motor and cognitive domains accounting for 89.4 to 97.9% of variance [86, 92, 93, 151]. Evidence for internal consistency has been reported across a number of sample populations (complete FIM α = 0.94–0.98, FIM motor α = 0.93–0.97 and FIM cognitive α = 0.93–0.94 for stroke, MS, traumatic and non-traumatic samples [151, 153]). And between-rater reliability has been demonstrated for both the motor and cognitive domains of the FIM in acute stroke (ICC 0.96, 0.91) respectively [87] and with participants with MS (FIM total inter-rater ICC = 0.99, FIM total intra-rater ICC = 0.94) [151]. Predictive associations between FIM scores and length of stay, discharge destination, minutes of assistance and supervision required on discharge and return to driving were identified [44, 50, 52, 82, 91, 132, 134]. When used with participants with MS, FIM was found to be a valid measure of disability [141], strongly correlating with like global measures (BI r = 0.88), activity measures (Ambulation Index r = - 0.73) and moderate to strongly with specific activity measures including housework (r = 0.64, $p<0.001$), work (r = -0.59 $p<0.001$), independence (r = -0.44, p = 0.001), and disability r = -0.96, $p< 0.001$) [151]. The FIM total score was at best only moderately responsive to change in a neurorehabilitation sample (ES 0.52–0.72), but the FIM cognitive was not (ES = 0.35–0.43) [97]. In comparison to other measures, the FIM was found to be less responsive than the original BI, equally responsive to BI(C&W) in stroke and more responsive than EDSS in MS, yet still only weak to moderately responsive to change (FIM ES = 0.46, FIM SRM 0.53, EDSS 0.15) [141, 151, 159]. Evidence of floor and ceiling effects for FIM were also found [44, 151, 159].

Results pertaining to sample with upper limb spasticity identified. FIM scores correlated with a measures of disability (Kurtkze Expanded Disability Status Scale (EDSS) r_s = -0.69) [141] and was found to be responsive when capturing change in participants with MS (SRM = 0.53) [141].

Goal Attainment Scaling. Goal Attainment Scaling (GAS) was first introduced by Kirusek and Sherman [176] and provides a structured approach to defining and measuring individualized patient centered and/or program based goals. A total of 9 studies evaluated the psychometric properties, in post-stroke (n = 2) [43, 156], MS (n = 1) [104], TBI (n = 3) [59, 120, 125]and mixed ABI (n = 3) samples [41, 115, 124]. Only one study met inclusion criteria that specifically identified participants with upper limb spasticity (in a sample greater than 6 months post-stroke) [156].

Content validity. Not assessed, as GAS identifies goal content particular to individual participants and programs (i.e. high face validity).

Results for whole sample. There were conflicting results in inter-rater reliability within a mixed neurological sample, while Joyce, Rockwood and Mate-Kole [102] report high reliability (r = 0.92, r = 0.94) between an individual rater familiar with GAS and the treating team, Bovend'Eerdt, Dawes, Izadi and Wade [41] found a fair level ($ICC_{A,k}$ 0.478) and low agreement (LOA -1.52 ± 25.54) between a therapist and masked assessor. When used with participants with MS, GAS change score correlated weakly with the BI (r_s = -0.25) and FIM (r_s = -0.6) [104]. In a sample of participants with ABI secondary to trauma and stroke, GAS also correlated strongly with global clinical impressions (r = 0.81) [104], weak to strongly with measures of daily activity, participation, disability, vocational outcome and quality of life (r = 0.34–0.81) but not with length of stay [102, 124, 125]. In the same sample, GAS at 2 months predicted final GAS scores at the completion of a rehabilitation program ranging from 7 to 42 weeks [125]. Ratings between participants and significant others agreed on 70% of occasions [59]. GAS was more responsive than the FIM and BI (ES 9.0 SRM: 2.4 t value 10.0 z value 1.4) in MS [104] and was responsive to patient centred outcomes and program change in a mixed neurological sample [115].

Results pertaining to sample with upper limb spasticity. GAS was found to have moderate correlations with self-reported benefit (rho = 0.46, $p < .001$), low correlations with quality of life (rho = 0.07, p = 0.52), disability (rho = 0.19, p = 0.08), carer burden (rho = 0.14, p = 0.26), measures of pain (rho = 0.03, p = 0.77), mood (rho = 0.06, p = 0.61) and spasticity (rho = 0.35, p = 0.001 [156].

Medical Outcome Study 36-Item Short-Form Health Survey. The Medical Outcome Study 36-Item Short-Form Health Survey (SF-36) is a global scale assessing eight health concepts [165, 177]. A total of 24 studies investigated the psychometric properties of the SF-36, 10 included participants with MS [76, 77, 95, 127, 130, 138, 143, 145, 161, 162], 10 post-stroke [29, 60, 61, 67, 85, 96, 122, 133, 148, 168], 3 post TBI [74, 84, 123] and 1 discussed tool development with nil specific sample [165]. No studies specifically identified participants with upper limb spasticity.

Content validity. The development of the SF-36 [165] did not appear to consult participants on the comprehensiveness or comprehensibility of included items [165]. Relevance of items for the intended purpose was sufficient. The SF-36 contains 6 of 9 recommended dimensions for patient-based, health related quality of life, less comprehensive than the SIS [148].

Results for whole sample. The SF-36 was found to have a two-factor structure; with the eight dimensions falling within the two constructs of physical and mental health [177]. Mixed results were found for the use of the domain scores, with scaling assumptions met in the TBI population [84] but only 6 of 8 scales meeting the scaling assumptions in stroke [96]. Evidence for internal consistency of the 8 dimensions, Cronbach alpha >0.70 in majority of studies [29, 61, 74, 76, 84, 161], however dimensions of vitality and general health did not meet this criteria (α = 0.68, α = 0.66–0.68) [85, 96]. Test-retest reliability varied; higher for patient reported scores (ICC = 0.30–0.81) than proxy reported scores (ICC = 0.25 to 0.76) [61, 130, 162]. Individual domains of the SF-36 correlated with like subscales of global measures (all r = $\geq$ 0.50) post-stroke (EQ-5D) [60] post TBI (Symptom Checklist, Health Problem List, Beck Depression Inventory) [74] and with participants with MS (LHS, FIM, general health questionnaire) [76]. Correlations, however, were not as strong as hypothesized between individual domains and like dimensions for the BI, CNS and FIM post stroke [85, 122] nor with the MSFC in a MS population (r = 0.16–0.51) [145]. The SF-36 physical and mental summary scores had weak to moderate correlations with participants rating of severity of symptoms (r = 0.38, r = 0.18) and quality of life (r = 0.47, r = 0.29) [127, 168]. The ability to discriminate between subgroups of participants with varying levels of function across post-stroke, TBI and MS populations was demonstrated [95, 138, 145, 161, 162]. The SF-36 was more responsive in the first three

months post-stroke [85] but less responsive in comparison to other tools measuring associated constructs in MS (ES = 0.01–0.30) [76]. SF-36 did not correlate with FIM change scores, suggesting the change captured within a HRQoL measure was not reflected in a global measure of activity [122]. There was evidence of significant floor and ceiling effects within MS [76, 77] and TBI [84], and varied reports post-stroke [60, 85, 96, 122, 133]. The minimal important clinical change varied across dimensions, reported to be 4–9 points within physical functioning, 6–8 within role physical, 6–7 social functioning and 6 points within the physical summary score [145].

Motor Activity Log. The Motor Activity Log (MAL) is a structured interview designed to capture use of the affected upper limb on two scales, Amount of Use (AOU) and Quality of Movement (QOM) [158]. Five studies evaluated the psychometric properties of MAL; all involved participants post-stroke [47, 63, 88, 157, 158], and one specifically identified participants with upper limb spasticity [88].

Content validity. The MAL was developed based on the non-use model to capture real-world arm function [157]. Item analysis suggests 2 items (put on makeup and write on paper) had greater than 20% missing data, with participants rating as not applicable, and had lower item-total correlations and reliability coefficients [158].

Results for the whole sample. The self-reported QOM scale correlated with performance based measures (ARAT r = 0.61, WMFT r = 0.65) with the AOU scale correlating less strongly with the WMFT r = 0.40 [63, 158]. The minimal detectable change was defined as 16.8% for the AOU and 15.3% for the QOM scales, but the minimal important change was not defined [47].

Results pertaining to sample with upper limb spasticity. The MAL correlated strongly with measures of activity (Chedoke Arm and Hand Activity Inventory (CAHAI) r = 0.82 $p<0.01$), weakly with measures of participation (Reintegration to Normal Living Index (RNL) r = 0.23 $p<0.05$) and of varying strengths (weak to moderate) with impairments, stronger than expected (spasticity r = -0.71, strength r = 0.61 to 0.84, pain r = -0.06, sensation r = -0.43, all $p<0.01$) [88].

Motor Activity Log-28. The Motor Activity Log-28 (MAL-28) is a revision of the MAL-30 with removal of redundant items 'write on paper' and 'put makeup/shaving cream on face' [158]. A single study evaluated the psychometric properties of this measurement tool involving participants greater than 6 months post-stroke, and without any participants with upper limb spasticity [158].

Content validity. Content analysis indicated appropriate range of items to cover basic (63%) and instrumental (41%) daily activities in addition to items that require finger movement, bimanual and unimanual tasks [158].

Results for the whole sample. Item analysis indicated that 98% of participants encountered included items in daily life [158]. There was evidence for internal consistency (α = 0.94–0.95) and increased test-retest reliability with self-ratings rather than proxy [158]. The MAL-28 held convergent validity with real life measure of hand performance and less with overall physical activity, patient ratings stronger than proxy [158].

Motricity Index. The Motricity Index (MI) is a brief scale of motor recovery [55]. Six studies evaluated the psychometric properties of MI [40, 48, 55, 98, 154, 163]; all involved participants post-stroke, and none specifically identified participants with upper limb spasticity.

Content validity. Demeurisse, Demol and Robaye [55] detailed the development of the MI with mixed results regarding its relevance and no evidence supporting either comprehensiveness nor comprehensibility.

Results for whole sample. There was evidence of the internal consistency of this tool (α = 0.97) [40] and high inter-rater reliability between an experienced and junior doctor

(rho = 0.88) rating 20 participants six weeks post-stroke [48]. The Upper Limb MI (UL MI) correlated strongly with like measures of upper limb activity (RMA arm r = 0.73–0.76) [48] and with global measures of activity (BI r = 0.77) [163] whilst correlating moderately with measures of dexterity (NHPT r = 0.36–0.56) [98]. The UL MI correlated strongly with impairments also, including grip strength (r = 0.74–0.94) [40]. The MI, when combined with the visual neglect recovery index and age at 2–3 days post-stroke was a significant predictor of independence at 3 months ($\beta = 0.042$, $p < .001$) and 6 months ($\beta = 0.038$, $p < .001$) [154]. Evidence of a ceiling effect was noted, with 18% of the sample scoring the maximum score within the UL component of the MI on discharge from a rehabilitation ward post-stroke [98]. There was no evidence of a floor effect.

Nine-Hole Peg Test. The Nine-Hole Peg Test (NHPT) is a timed measure of unilateral upper limb dexterity through the placing and removal of nine pegs in/out of a board [178]. Ten studies evaluated the psychometric properties; 5 post-stroke [38, 94, 98, 129] and 5 included participants with MS [39, 51, 79, 139, 150]. One study specifically identified participants with upper limb spasticity [38].

Content validity. The NHPT was first discussed as being used in a study in 1985 [179]; no information was reported to inform the development nor content validity of the NHPT.

Results for whole sample. The NHPT when used with participants post-stroke correlated with both observed (r = 0.36–0.95) [38, 79, 94, 98, 139] and self-reported measures of activity and hand use (r = 0.53–0.66) [98], was more sensitive than the FAT [94], had poor predictive validity in comparison to like measures, and did not predict HRQoL [129]. The NHPT correlated highly with measures of tremor and dexterity in MS, common activity limitation features (r = -0.62 - -0.87 $p<0.005$) [180]. There was evidence for the reliability of the NHPT (inter-rater Rho = 0.75–0.99 and test-retest Rho = 0.68–0.90 and 0.83–0.99) when administered to participants 18 months post-stroke [94]. The NHPT was moderate to highly responsive within the first 6 months post-stroke (ES = 0.52–0.66) [38, 98], was more responsive than the upper limb MI [98] and measures of strength, equally responsive to the ARAT, Jebsen-Taylor test of hand function and less responsive than the SIS-hand [38]. True change was indicated by a change of 20% when administered to participants with MS [150]. There were no floor or ceiling effects found in the MS population.

Results pertaining to sample with upper limb spasticity identified. Strong correlations with measures of hand use, grip and dexterity were reported in stroke populations ($r_s = 0.61–0.95$) and with measures of strength ($r_s = 0.61–0.82$) [38] despite the NHPT being a simulated task performance measure. The NHPT was found to be equally responsive as like measures of upper limb activity performance (ARAT and Jebsen-Taylor test of hand function) (ES 0.52–0.66), more responsive than measures of impairment (pinch and grip strength) but less responsive than the SIS-Hand (ES = 0.55–1.018) in the first 6 months post-stroke [38].

Oxford Handicap Scale. The Oxford Handicap Scale (OHS) is a simple tool modified from the Rankin Scale to grade the ability of a person and the level of daily assistance required to live independently [181]. Two studies evaluated the psychometric properties of the OHS, both including participants less than 6 months post-stroke [144, 152]. Neither study specifically identified participants to have upper limb spasticity.

Content validity. No published information regarding the development nor content validity of the OHS was located.

Results for whole sample. The OHS was not a predictor of caregiver burden [144] but was found to predict both the number of services and amount of time required from services on discharge [152].

Rivermead Motor Assessment. The Rivermead Motor Assessment (RMA) [117] is comprised of three sections; for this review studies were separated into two categories 1) 'RMA' all

three sections (upper limb, trunk and leg) administered and reported and 2) 'RMA UL' upper limb section of the RMA only administered and reported. A total of 7 studies were included [25, 26, 48, 117, 129, 147], all studies included participants post-stroke, 4 of the 7 studies included participants less than 6 months post-stroke [26, 48, 101, 147]. When separated into the two categories, evidence for the 'complete RMA' was drawn from 5 studies [25, 26, 101, 117, 147] and evidence for the 'RMA UL' section was drawn from 6 studies [25, 26, 48, 117, 129, 147].

Content validity. Test authors Lincoln and Leadbitter [117] detail the measurement tool development. This was completed via selecting a preliminary series of items ranging widely in difficulty ordered into the three sections; gross, leg and trunk and arm. All individual sections were found to have mixed results regarding relevance, reduced due to methods used to create items and nil information regarding comprehensiveness nor comprehensibility.

Results for whole sample. The hierarchical scale of the RMA in an acute and non-acute stroke sample found varying results. Evidence to support the scalability of the RMA was found for the gross function and arm section in acute stroke only [26]. Scalability was supported in the gross function section only, when used with participants 6 and 12 months post-stroke [25, 147]. The RMA correlated with ADL performance (r = 0.51) and balance (r = -0.45) [147], a related construct. Agreement between clinician and participants predicted scores with achieved scores was found (clinician ICC 0.965 Bland Altman 96.6; participants ICC 0.908 Bland Altman 79.3) [101]. The hierarchical scale of the RMA UL section was supported only when administered to participants in the acute phase post-stroke (Guttman scaling criteria met) [26], the scalability criteria was not met when used with participants 6 and 12 months post-stroke [25]. The UL section of the RMA was found to correlate strongly with measures of upper limb activity at 6, 12 and 18 weeks post stroke (r = Rho 0.73–0.76) [48] and greater than six months post stroke (r = - 0.80) [129]. The RMA UL correlated moderately with perceived physical activity (r = -0.47) and did not predict overall HRQoL [129].

Stroke-Adapted Version of the Sickness Impact Profile. The Stroke-Adapted Version of the Sickness Impact Profile (SA-SIP30) was derived from the original Sickness Impact Profile and contains the following 8 subscales: body care and movement, mobility, ambulation, social interaction, emotional behavior, alertness behavior, communication and household management [160]. Four studies evaluated the psychometric properties of the SA-SIP30 [58, 69, 148, 160], all involved participants post-stroke, and only one study specifically identified participants with upper limb spasticity [58].

Content validity. Test authors detailed the methodology applied to create the SA-SIP, based on statistical relevancy and homogeneity [160]. The scale was found to be relevant, however to lack comprehensiveness (as only 5 of 9 recommended dimensions for patient-based, health related quality of life measures were included) [148]. No information regarding comprehensibility was provided.

Results for whole sample. The SA-SIP accounted for 53% of variance in predicting participation ($R^2 = 0.63$, $P<0.001$) and was more sensitive to detecting stroke related changes impacting on independence at 6 months post-stroke [69].

Results pertaining to sample with upper limb spasticity. The SA-SIP30 was significantly associated with greater disability in hygiene, dressing, limb posture and pain ($P < .05$) [58].

Stroke Impact Scale. The Stroke Impact Scale (SIS) is a stroke-specific measure of global health outcome [64] and comprises of eight domains: strength, hand function, activities of daily living, instrumental activities of daily living, mobility, communication, emotion, memory and thinking, and participation. The SIS was found to be reported as either individual or collective domains which are administered and reported separately. To maintain consistency across all measures within this review, the SIS was required to be administered in full and in

the form of version 3 to meet inclusion criteria. Ten studies evaluated the psychometric properties of version 3 of the SIS [64–66, 68, 71, 99, 110, 112, 148, 170], all included participants post-stroke and none specifically identified participants with upper limb spasticity.

Content validity. The SIS was originally developed following a comprehensive iterative process with the use of participants, caregivers and standardized instrument development guidelines implemented but specific details are not available (unpublished information) [68]. Rasch analysis led to revision of the measure [64] demonstrating comprehensiveness (containing 7 of 9 recommended dimensions for patient-based, health related quality of life) and to be more comprehensive than EQ-5D, SA-SIP and SF-36 [148].

Results for whole sample. Rasch analysis refined the SIS into version 3 producing unidimensional domains ranging in item difficulty and with the ability to discriminate [64]. A single index was proposed, aggregated from the 8 domains ($\alpha = 0.93$) accounting for 68.76% of the variance [99]. These 8 domains were each found to be internally consistent ($\alpha \geq 0.86$–0.96) [66, 99], suggesting possible item redundancy and further investigations of shorter forms. Agreement between patient and proxy ratings were fair to excellent, being stronger in the observable physical domains (ICC 0.50 to 0.83) [65]. The tool was reliable between testing sessions when administered via mail (ICC 0.77–0.99) and telephone modes (ICC 0.90–0.99) [66]. The individual and related domains of the SIS were found to correlate with global measures of independence, activity and participation, both patient and proxy reported, ($r = 0.69$–0.78) [65, 110, 170]. The SIS was able to discriminate between participants deemed recovered by the BI [112] and held superior ability to discriminate between varying levels of disability compared to the FIM and SF-36V (modified version of the SF-36) when tools were administered via phone [110]. Floor and ceiling effects were varied ranging from nil floor effect and 0–32% ceiling effect [71, 110].

Upper-Limb Motor Assessment Scale. The Upper Limb -Motor Assessment Scale (UL-MAS) is a subscale of items 6, 7 and 8 of the Motor Assessment Scale, and it provides a task orientated performance-based measure of upper limb activity [46]. Ten studies evaluating the psychometric properties of the UL-MAS were included [46, 100, 103, 109, 116, 118, 119, 126, 137, 146], all involved participants less than 6 months post-stroke, and no studies specifically identified participants with upper limb spasticity.

Content validity. Evidence located for the development of the MAS and subsequent UL-MAS did not indicate participants were consulted on the comprehensiveness or comprehensibility of included items [46]. Relevance of items for the intended purpose was sufficient.

Results for whole sample. There was evidence to support the production of a single composite score from the UL-MAS items, which may be interpreted as a total score for UL function [116]. Inconsistencies were identified within the hierarchical scoring [126, 137, 146] with clinical recommendations to attempt and score every item [126]. Furthermore, task 2 within the Hand Movements item may not be indicative of upper limb motor recovery in adults aged 65 years and older [126]. The UL-MAS is a unidimensional scale measuring a single construct, upper limb motor performance, ($\alpha = 0.83$ to 0.95, and with removal of wrist deviation 0.93) [100, 116, 126]. It was reliable between (Kendall Tau = 0.74–1.00) and amongst assessors (kappa 0.93–1.0, 88–85% agreement) [46, 118]. The UL-MAS was able to discriminate between differing levels of motor recovery both in the acute and subacute phase, with Rasch based scoring more precise [103]. Varying levels of floor and ceiling effects have been reported for the UL-MAS (floor effect 0–38%, ceiling effect 0–67%) [126, 137, 146].

Discussion

This systematic review located, appraised and synthesized the body of literature investigating the psychometric properties of measurement tools which assess upper limb function in the

context of everyday activities. Across the included 29 measurement tools, there was wide variability in the quality of evidence in relation to participants with neurological conditions, but overall, tools with the greatest number of psychometric publications demonstrated the strongest evidence. While the FIM™ had the highest quality evidence supporting its validity and reliability, it suffered from both floor and ceiling effects. On consideration of specific constructs measured by the tools, wide variability across quality of evidence remained. Both patient-reported measures, the ArmA and DAS, and performance-based measures, the UL-MAS and ARAT, demonstrated evidence within the measures specifically targeting upper limb activity. Evidence supported use regardless of whether upper limb spasticity was present or not, except for the UL-MAS, which is replaced with the MAL for patients with identified upper limb spasticity. Despite the BI and BI(C&W) holding high to moderate levels of evidence for construct validity, the FIM held the strongest level of evidence for global measures of activity, regardless of whether or not upper limb spasticity was present. The SIS, a patient-reported measure, held the strongest level of evidence across a greater number of properties and demonstrated higher correlations with measures of upper limb performance and activity of the global health-related quality of life measures. The EQ-5D and SA-SIP were the only health-related quality of life measures with evidence supporting construct validity for participants with upper limb spasticity. In light of mixed findings without a clearly superior measurement tool, findings highlights the need for further research into the psychometric properties of measurement tools which capture upper limb activity and/or participation performance.

The search yielded psychometric studies primarily conducted between 2000 and 2010, with an even split of additional evidence located in the 10 years either side of that decade. It was interesting that few papers have been published in the more recent years–this may reflect publication preferences of journals in rehabilitation or a potential assumption by clinicians that the psychometric properties have been well established. Most studies were completed with participants post-stroke in the acute to subacute phase, and as such, findings from these studies may not apply to a more chronic population or a group of neurological clients who have not suffered a stroke. Individual study sample sizes were commonly small (less than n = 100 in over half (56%) of studies), which is a common limitation highlighted by other reviews of functional measurement tools [182, 183]. This finding strengthens earlier calls for continued investment in appropriately powered psychometric studies, inclusion of psychometric evaluation in both routine data collection and longitudinal studies, and a need for scientific journals or outcome tool publishers to publish such research.

The construct validity and responsiveness, followed by reliability properties of measurement tools, were most commonly evaluated across the different tools, but rarely was content validity or measurement error tested. The methodological quality of included studies was wide ranging, from 'inadequate' to 'very good', suggesting that making decisions between measures may be difficult, since there was little consistent data to guide decisions. Detailed data was often lacking within studies such as those reporting on the reliability of tools where information failed to describe testing conditions, stability of patients between sessions and evidence for systematic change occurrence. The COSMIN process recommends that an 'a priori' hypothesis be developed when evaluating construct validity and responsiveness, however in our review only a very small number of studies clearly defined hypotheses about the expected results. The majority of studies were found to report generic hypotheses, where hypotheses were assigned based on interpretations by the authors. Furthermore, the quality of statistical approaches used were low, for example often reporting on statistical significance of findings rather than expected strengths and direction of correlations. Consistent with Zaki and colleagues [184], our review also suggests that the quality of research in psychometrics is unlikely to improve without education and clear guidelines on analysis. The COSMIN checklist may

provide such guidance; the COSMIN process separates the statistical methods based on Classical Test Theory (CTT) or on Item Response Theory (IRT) and an understanding of these methods is likely key to improving the psychometrics of scales where multiple items contribute to an overall score.

The review identified very limited evidence useful for the clinical selection of a single tool to evaluate upper limb activity when upper limb spasticity is present. Inadequate representation of the intended population within the sample of a psychometric study can lead to erroneous assumptions about the psychometrics of a tool [185]. In the context of instrument development, internal and external validity are important for application of an instrument in assessing new target populations (in this case, adults with upper limb spasticity). The DAS, EQ-5D, FIM™, NHPT and SA-SIP had evidence supporting both internal and external validity and responsiveness, however no single measurement tool had identified psychometric evidence for all properties in a sample of participants with upper limb spasticity. This gap in available research is acknowledged, and is both a limitation to this systematic review and a recommendation for further research. The evidence located to guide selection for the broader neurorehabilitation sample was larger in comparison primarily due to additional numbers of contributing studies. However, despite large numbers of contributing studies, we could still not conclude that any of the identified measurement tools from the Ashford and Turner-Stokes [8] review have published psychometric evidence for all relevant psychometric properties.

In this review, despite selecting the most recent and comprehensive set of tools at the time of registering our protocol, we acknowledge a potential limitation in range of tools included and that other existing tools had not been used in clinical trials or cohort studies of patients with spasticity, and therefore were not synthesized in the Ashford and Turner-Stokes [8] review. The limited psychometric testing of the tools that were included was a further limitation, making it difficult to compare the psychometric properties of tools across different pathologies. This may mean that the preferred assessments of a reader does not appear in this extensive review, and where included, it may have only been tested in a single diagnostic population. Only one additional measurement tool beyond the initial systematic review was recommended in the recent national guidelines [13], that tool being the Arm Activity Measure (ArmA). Psychometric studies not published in English were also excluded for pragmatic reasons; formal translations have not yet occurred in many of the measurement tools (e.g. ARAT and UL-MAS) and therefore studies conducted in languages other than English were excluded as per COSMIN guidelines.

Conclusions

This systematic review provides a comprehensive synthesis of the psychometric properties of the upper extremity measurement tools used to evaluate the dimensions of activity and/or participation. The findings may provide guidance for clinicians on evidence-based measurement tool selection, however further psychometric evaluation of tools is recommended. Together, 29 measurement tools met the inclusion criteria and of these, 8 demonstrated at least a moderate level of confidence in the measurement property estimate in two or more standards. While no tool had at least moderate estimates for all standards (i.e. content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for construct validity and responsiveness), the review was able to suggest which measurement tools should continue to be researched and refined for use. Future research needs to investigate the psychometric properties of these

measurement tools, across a range of neurological populations as well as with a subsample with spasticity in the upper limb.

Supporting information

S1 Checklist. PRISMA checklist.
(DOC)

S1 File. Search strategy and search terms. MEDLINE search strategy and terms used in search.
(DOCX)

S1 Table. Full text exclusion reasons (PRISMA). This file details reasons for and numbers of studies excluded.
(DOCX)

S2 Table. Methodological quality and quality criteria ratings. This file lists all included studies and methodological quality and quality criteria ratings.
(DOCX)

S3 Table. Summary of results. This file provides a summary of results for all included studies.
(DOCX)

S4 Table. Terwee quality criteria and guide for strength of correlations.
(DOCX)

Acknowledgments

Many thanks to Jenny Price, Murrumbidgee Local Health District librarian who assisted with sourcing published studies.

Author Contributions

Conceptualization: Shannon Pike, Anne Cusick, Natasha A. Lannin.

Formal analysis: Shannon Pike, Anne Cusick, Kylie Wales, Natasha A. Lannin.

Investigation: Shannon Pike, Anne Cusick, Kylie Wales, Lisa Cameron, Natasha A. Lannin.

Methodology: Shannon Pike, Anne Cusick, Kylie Wales, Lynne Turner-Stokes, Stephen Ashford, Natasha A. Lannin.

Supervision: Anne Cusick, Natasha A. Lannin.

Writing – original draft: Shannon Pike, Anne Cusick, Natasha A. Lannin.

Writing – review & editing: Shannon Pike, Anne Cusick, Kylie Wales, Lisa Cameron, Lynne Turner-Stokes, Stephen Ashford, Natasha A. Lannin.

References

1. World Health Organisation. International Classification of Functioning, Disability and Health. Geneva: World Health Organisation; 2001.
2. Geyh S, Cieza A, Schouten J, Dickson H, Frommelt P, Omar Z, et al. ICF Core Sets for stroke. Journal of rehabilitation medicine. 2004(44 Suppl):135–41. https://doi.org/10.1080/16501960410016776 PMID: 15370761
3. Zhang T, Liu L, Xie R, Peng Y, Wang H, Chen Z, et al. Value of using the international classification of functioning, disability, and health for stroke rehabilitation assessment: A multicenter clinical study.

Medicine (Baltimore). 2018; 97(42):e12802–e. https://doi.org/10.1097/MD.0000000000012802 PMID: 30334972

4. Lohmann S, Decker J, Müller M, Strobl R, Grill E. The ICF Forms a Useful Framework for Classifying Individual Patient Goals in Post-acute Rehabilitation. Journal of rehabilitation medicine. 2011; 43 (2):151–5. https://doi.org/10.2340/16501977-0657 PMID: 21234515
5. Ashford S, Slade M, Malaprade F, Turner-Stokes L. Evaluation of functional outcome measures for the hemiparetic upper limb: A systematic review. Journal of Rehabilitation Medicine. 2008; 40:787–95. https://doi.org/10.2340/16501977-0276 PMID: 19242614
6. Alt Murphy M, Resteghini C, Feys P, Lamers I. An overview of systematic reviews on upper extremity outcome measures after stroke. BMC Neurology. 2015; 15(1). https://doi.org/10.1186/s12883-015-0292-6 PMID: 25880033
7. Lamers I, Kelchtermans S, Baert I, Feys P. Upper Limb Assessment in Multiple Sclerosis: A Systematic Review of Outcome Measures and their Psychometric Properties. Archives of physical medicine and rehabilitation. 2014; 95(6):1184–200. https://doi.org/10.1016/j.apmr.2014.02.023 PMID: 24631802
8. Ashford S, Turner-Stokes L. Systematic review of upper-limb function measurement methods in botulinum toxin intervention for focal spasticity. Physiotherapy Research International. 2013; 18(3):178–89. https://doi.org/10.1002/pri.1554 PMID: 23630050
9. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research. 2018; 27 (5):1147–57. https://doi.org/10.1007/s11136-018-1798-3 PMID: 29435801
10. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of Life Research. 2012; 21(4):651–7. https://doi.org/10.1007/s11136-011-9960-1 PMID: 21732199
11. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Quality of Life Research. 2018; 27(5):1159–70. https://doi.org/10.1007/s11136-018-1829-0 PMID: 29550964
12. AlHakeem N, Ouellette EA, Travascio F, Asfour S. Surgical Intervention for Spastic Upper Extremity Improves Lower Extremity Kinematics in Spastic Adults: A Collection of Case Studies. Front Bioeng Biotechnol. 2020; 8:116–. https://doi.org/10.3389/fbioe.2020.00116 PMID: 32154240
13. Royal College of Physicians, British Society of Rehabilitation Medicine, The Chartered Society of Physiotherapy, Association of Chartered Physiotherapists in Neurology, Royal College of Occupational Therapists. Spasticity in adults: management using botulinum toxin. National guidelines. London: RCP; 2018.
14. World Health Organisation. How to use the ICF: A practice manual for using the International Classification of Functioning Disability and Health (ICF). Exposure draft for comment. Geneva: WHO; 2013.
15. Pike S, Lannin NA, Cusick A, Wales K, Turner-Stokes L, Ashford S. A systematic review protocol to evaluate the psychometric properties of measures of function within adult neuro-rehabilition. Systematic Reviews. 2015; 4(86).
16. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Quality of Life Research. 2009; 18(8):1115–23. https://doi.org/10.1007/s11136-009-9528-5 PMID: 19711195
17. Page SJ, Levine P, Hade E. Psychometric properties and administration of the Wrist/Hand Subscales of the Fugl-Meyer assessment in minimally impaired upper extremity hemiparesis in stroke. Archives of Physical Medicine and Rehabilitation. 2012; 93:2373–6. https://doi.org/10.1016/j.apmr.2012.06.017 PMID: 22759831
18. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias Checklist for systematic reviews of Patient-Reported Outcome Measures. Quality of Life Outcomes. 2018; 27:1171–9. https://doi.org/10.1007/s11136-017-1765-4 PMID: 29260445
19. De Souza LH, Langton-Hewer R, Miller S. Assessment of recovery of arm control in hemiplegic stroke patients. 1. Arm function tests. International Rehabilitation Medicine. 1980; 2(1):3–9. https://doi.org/10.3109/09638288009163947 PMID: 7440085
20. Smith SJ, Ellis E, White S, Moore AP. A double-blind placebo-controlled study of botulinum toxin in upper limb spasticity after stroke or head injury. Clinical Rehabilitation. 2000; 14(1):5–13. https://doi.org/10.1191/026921500666642221 PMID: 10688339
21. Bhakta BB, O'Connor RJ, Cozens AJ. Associated Reactions After Stroke: A Randomized Controlled Trial of the Effect of Botulinum Toxin Type A. Journal of Rehabilitation Medicine. 2008; 40(1):36–41. https://doi.org/10.2340/16501977-0120 PMID: 18176735

22. Klein RM, Bell B. Self-care skills: Behavioural measurement with the Klein-Bell ADL Scale. Archives of Physical Medicine and Rehabilitation. 1982; 63:335–8. PMID: 7092535

23. Chang C-L, Munin MC, Skidmore ER, Niyonkuru C, Huber LM, Weber DJ. Effect of Baseline Spastic Hemiparesis on Recovery of Upper-Limb Function Following Botulinum Toxin Type A Injections and Postinjection Therapy. Archives of Physical Medicine and Rehabilitation. 2009; 90(9):1462–8. https://doi.org/10.1016/j.apmr.2009.03.008 PMID: 19735772

24. Bhakta BB, Cozens JA, Chamberlain MA, Bamford JM. Impact of botulinum toxin type A on disability and carer burden due to arm spasticity after stroke: a randomsed double blind placebo controlled trial. Journal of Neurology, Neurosurgery & Psychiatry. 2000; 69:217–21.

25. Adams SA, Pickering RM, Ashburn A, Lincoln NB. The scalability of the Rivermead Motor Assessment in nonacute stroke patients. Clinical Rehabilitation. 1997; 11:52–9. https://doi.org/10.1177/026921559701100108 PMID: 9065360

26. Adams SA, Pickering RM, Taylor D. The scalability of the Rivermead Motor Assessment in acute stroke patients. Clinical Rehabilitation. 1997; 11:42–51. https://doi.org/10.1177/026921559701100107 PMID: 9065359

27. Alderman N, Dawson K, Rutterford NA, Reynolds PJ. A comparison of the validity of self-report measures amongst people with acquired brain injury: A preliminary study of the usefulness of EuroQol-5D. Neuropsychological Rehabilitation 2001; 11(5):529–37.

28. Ali M, Fulton R, Quinn T, Brady M, VISTA. How well do standard stroke outcome measures reflect quality of life? A retrospective analysis of clinical trial data. Stroke. 2013; 44:3161–5. https://doi.org/10.1161/STROKEAHA.113.001126 PMID: 24052510

29. Anderson C, Laubscher SD, Burns R. Validation of the Short Form 36 (SF-36) Health Survey Questionnaire Among Stroke Patients. Stroke A Journal of Cerebral Circulation. 1996; 27(10):1812–6. https://doi.org/10.1161/01.str.27.10.1812 PMID: 8841336

30. Ashford S, Jackson D, Turner-Stokes L. Goal setting, using goal attainment scaling, as a method to identify patient selected items for measuring arm function. Physiotherapy (United Kingdom). 2015; 101(1):88–94. https://doi.org/10.1016/j.physio.2014.04.001 PMID: 24954806

31. Ashford S, Siegert RJ, Alexandrescu R. Rasch measurement: the Arm Activity measure (ArmA) passive function sub-scale. Disability & Rehabilitation. 2016; 38(4):384–90. https://doi.org/10.3109/09638288.2015.1041613 PMID: 25918961

32. Ashford S, Slade M, Nair A, Turner-Stokes L. Arm Activity measure (ArmA) application for recording functional gain following focal spasticity treatment. . .including commentary by Balakatounis K. International Journal of Therapy & Rehabilitation. 2014; 21(1):10–7.

33. Ashford S, Slade M, Turner-Stokes L. Conceptualisation and development of the arm activity measure (ArmA) for assessment of activity in the hemiparetic arm. Disability & Rehabilitation. 2013; 35(18):1513–8. https://doi.org/10.3109/09638288.2012.743602 PMID: 23294435

34. Ashford S, Turner-Stokes L, Siegert R, Slade M. Initial psychometric evaluation of the Arm Activity Measure (ArmA): a measure of activity in the hemiparetic arm. Clinical Rehabilitation. 2013; 27(8):728–40. https://doi.org/10.1177/0269215512474942 PMID: 23426566

35. Barer DH, Murphy JJ. Scaling the Barthel: a 10-point hierarchical version of the activities of daily living index for use with stroke patients. Clinical Rehabilitation. 1993; 7(4):271–7.

36. Barton GR, Sach TH, Avery AJ, Jenkinson C, Doherty M, Whynes DK, et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged $\geq$ 45 years. Health Economics. 2008; 17:815–32. https://doi.org/10.1002/hec.1298 PMID: 17893863

37. Barton GR, Sach TH, Doherty M, Avery AJ, Jenkinson C, Muir KR. An assessment of the discriminative ability of the EQ-5D$_{index,}$ SF-6D, and EQ VAS, using sociodemographic factors and clinical conditions. European Journal of Health Economics. 2008; 9:237–49.

38. Beebe JA, Lang CE. Relationships and Responsiveness of Six Upper Extremity Function Tests During the First Six Months of Recovery After Stroke. Journal of Neurologic Physical Therapy. 2009; 33(2):96–103. https://doi.org/10.1097/NPT.0b013e3181a33638 PMID: 19556918

39. Benedict HB, Holtzer R, Motl RW, Foley FW, Kaur S, Hojnacki D, et al. Upper and lower extremity motor function and cognitive impairment in multiple sclerosis. Journal of the International Neuropsychological Society. 2011; 17:643–53. https://doi.org/10.1017/S1355617711000403 PMID: 21486517

40. Bohannon RW. Motricity index scores are valid indicators of paretic upper extremity strength following stroke. Journal of Physical Therapy Science. 1999; 11:59–61.

41. Bovend'Eerdt TJH, Dawes H, Izadi H, Wade DT. Agreement between two different scoring procedures for goal attainment scaling is low. Journal of Rehabilitation Medicine. 2011; 43:46–9. https://doi.org/10.2340/16501977-0624 PMID: 21042701

42. Brashear A, Zafonte R, Corcoran M, Galvez-Jimenez N, Gracies JM, Gordon MF, et al. Inter- and Intrarater reliability of the Ashworth scale and the Disability Assessment Scale in patients with upper-limb poststroke spasticity Archives of Physical Medicine and Rehabilitation. 2002; 83:1349–54. https://doi.org/10.1053/apmr.2002.35474 PMID: 12370866

43. Brock K, Black S, Cotton S, Kennedy G, Wilson S, Sutton E. Goal achievement in the six months after inpatient rehabilitation after stroke. Disability and Rehabilitation. 2009; 31(11):880–6. https://doi.org/10.1080/09638280802356179 PMID: 19037772

44. Brown AW, Therneau TM, Schultz BA, Niewczyk PM, Granger CV. Measure of functional independence dominates discharge outcome prediction after inpatient rehabilitation for stroke. Stroke. 2015; 46:1038–44. https://doi.org/10.1161/STROKEAHA.114.007392 PMID: 25712941

45. Burridge JH, Turk R, Notley SV, Pickering RM, Simpson DM. The relationship between upper limb activity and impairment in post-stroke hemiplegia. Disability and Rehabilitation. 2009; 31(2):109–17. https://doi.org/10.1080/09638280701824699 PMID: 18608395

46. Carr JH, Shephard RB, Nordholm L, Lynne D. Investigation of a new motor assessment scale for stroke patients. Physical Therapy. 1985; 65:175–80. https://doi.org/10.1093/ptj/65.2.175 PMID: 3969398

47. Chen S, Wolf SL, Zhang Q, Thompson PA, Winstein CJ. Minimal detectable change of the actual amount of use test and the motor activity log: The EXCITE trial. Neurorehabilitation and Neural Repair. 2012; 26(5):507–14. https://doi.org/10.1177/1545968311425048 PMID: 22275157

48. Collin C, Wade D. Assessing motor impairment after stroke: a pilot reliability study. Journal of Neurology, Neurosurgery, and Psychiatry. 1990; 53(7):576–9. https://doi.org/10.1136/jnnp.53.7.576 PMID: 2391521

49. Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: A reliability study. International Disability Studies. 1988; 10(2):61–3. https://doi.org/10.3109/09638288809164103 PMID: 3403500

50. Corrigan JD, Smith-Knapp K, Granger CV. Validity of the functional independence measure for persons with traumatic brain injury. Archives of Physical Medicine and Rehabilitation. 1997; 78(8):828–34. https://doi.org/10.1016/s0003-9993(97)90195-7 PMID: 9344301

51. Costelloe L, O'Rourke K, McGuigan C, Walsh C, Tubridy N, Hutchinson M. The longitudinal relationship between the patient-reported Multiple Sclerosis Impact Scale and the clinician-assessed Multiple Sclerosis Functional Composite. Multiple Sclerosis. 2008; 14:255–8. https://doi.org/10.1177/1352458507081274 PMID: 17942522

52. Cullen N, Krakowski A, Taggart C. Functional independence measure at rehabilitation admission as a predictor of return to driving after traumatic brain injury. Brain Injury. 2014; 28(2):189–95. https://doi.org/10.3109/02699052.2013.862738 PMID: 24456058

53. Cuthbert JP, Harrison-Felixx C, Corrigan JD, Bell JM, Haarbauer-Krupa JK, Miller C. Unemployment in the United States after traumatic brain injury for working-age individuals: prevalence and associated factors 2 years postinjury. Journal of Head Trauma Rehabilitation. 2015; 30(3):160–74. https://doi.org/10.1097/HTR.0000000000000090 PMID: 25955703

54. Dang M, Ramsaran KD, Street ME, Syed SN, Barclay-Goddard R, Stratford PW, et al. Estimating the accuracy of the Chedoke-McMaster Stroke Assessment predictive equations for stroke rehabilitation. Physiotherapy Canada. 2011; 63(3):334–41. https://doi.org/10.3138/ptc.2010-17 PMID: 22654239

55. Demeurisse G, Demol O, Robaye E. Motor evaluation in vascular hemiplegia. European Neurology. 1980; 19:382–9. https://doi.org/10.1159/000115178 PMID: 7439211

56. Dennis M, O'Rourke S, Lewis S, Sharpe M, Warlow C. Emotional outcomes after stroke: factors associated with poor outcome. Journal of Neurology, Neurosurgery & Psychiatry 2000; 68:47–52. https://doi.org/10.1136/jnnp.68.1.47 PMID: 10601401

57. De Weerdt WJG, Harrison MA. Measuring recovery of arm-hand function in stroke patients: A comparison of the Brunnstrom-Fugl-Meyer test and the Action Research Arm test. Physiotherapy Canada. 1985; 37(2):65–70.

58. Doan QV, Brashear A, Gillard PJ, Varon SF, Vandenburgh AM, Turkel CC, et al. Relationship between disability and health-related quality of life and caregiver burden in patients with upper limb poststroke spasticity. PM&R. 2012; 4(1):4–10. https://doi.org/10.1016/j.pmrj.2011.10.001 PMID: 22200567

59. Doig E, Fleming J, Kuipers P, Cornwell PL. Clinical utility of the combined use of the Canadian Occupational Performance Measurer and Goal Attainment Scaling. American Journal of Occupational Therapy. 2010; 64(6):904–14.

60. Dorman PJ, Dennis M, Sandercock P. How do scores on the EuroQol relate to scores on the SF-36 after stroke? Stroke. 1999; 30:2146–51. https://doi.org/10.1161/01.str.30.10.2146 PMID: 10512920

61. Dorman PJ, Slattery J, Farrell B, Dennis M, Sandercock P. Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 Questionnaires after stroke. Stroke. 1998; 29:63–8. https://doi.org/10.1161/01.str.29.1.63 PMID: 9445330

62. Dorman PJ, Waddell F, Slattery J, Dennis M, Sandercock P. Is the EuroQol a valid measure of health related quality of life after stroke? Stroke. 1997; 28:1876–82. https://doi.org/10.1161/01.str.28.10.1876 PMID: 9341688

63. Dromerick AW, Lang CE, Birkenmeier R, Hahn MG, Sahrmann SA, Edwards DF. Relationships between upper-limb functional limitation and self-reported disability 3 months after stroke. Journal of Rehabilitation and Development. 2006; 43(3):401–8. https://doi.org/10.1682/jrrd.2005.04.0075 PMID: 17041825

64. Duncan PW, Bode RK, Min Lai S, Perera S, Investigators GAiNA. Rasch analysis of a new stroke-specific outcome scale: The Stroke Impact Scale. Archives of Physical Medicine and Rehabilitation. 2003; 84:950–63. https://doi.org/10.1016/s0003-9993(03)00035-2 PMID: 12881816

65. Duncan PW, Min Lai S, Tyler D, Perera S, Reker DM, Studenski S. Evaluation of proxy responses to the Stroke Impact Scale. Stroke. 2002; 33:2593–9. https://doi.org/10.1161/01.str.0000034395.06874.3e PMID: 12411648

66. Duncan PW, Reker D, Kwon S, Lai S-M, Studenski S, Perera S, et al. Measuring stroke impact with the Stroke Impact Scale: telephone versus mail administration in veterans with stroke. Medical Care. 2005; 43(5):507–15. https://doi.org/10.1097/01.mlr.0000160421.42858.de PMID: 15838417

67. Duncan PW, Samsa GP, Weinberger M, Goldtein LB, Bonito A, Witter DM, et al. Health status of individuals with mild stroke. Stroke. 1997; 28:740–5. https://doi.org/10.1161/01.str.28.4.740 PMID: 9099189

68. Duncan PW, Wallace D, Min Lai S, Johnson D, Embretson S, Laster LJ. The Stroke Impact Scale Version 2.0: Evaluation of reliability, validity, and sensitivity to change. Stroke. 1999; 30(10):2131–40. https://doi.org/10.1161/01.str.30.10.2131 PMID: 10512918

69. Edwards DF, Hahn M, Baum C, Dromerick AW. The impact of mild stroke on meaningful activity and life satisfaction. Journal of stroke and cerebrovascular diseases. 2006; 15(4):151–7. https://doi.org/10.1016/j.jstrokecerebrovasdis.2006.04.001 PMID: 17904068

70. Egan M, Davis CG, Dubouloz C-J, Kessler D, Kubina L-A. Participation and well-being poststroke: Evidence of reciprocal effects. Archives of Physical Medicine and Rehabilitation. 2014; 95:262–8. https://doi.org/10.1016/j.apmr.2013.08.013 PMID: 24001446

71. Eriksson G, Baum MC, Wolf TJ, Connor LT. Perceived participation after stroke: The influence of activity retention, reintegration, and perceived recovery. The American Journal of Occupational Therapy. 2013; 67:e13–e8.

72. Filiatrault J, Bertrand Arsenault A, Dutil E, Bourbonnais D. Motor function and activities of daily living assessments: a study of three tests for persons with hemiplegia. The American Journal of Occupational Therapy. 1991; 45(9):806–10. https://doi.org/10.5014/ajot.45.9.806 PMID: 1928288

73. Fisk JD, Brown MG, Sketris IS, Metz LM, Murray TJ, Stadnyk KJ. A comparison of health utility measures for the evaluation of multiple sclerosis treatments. Journal of Neurology, Neurosurgery & Psychiatry. 2005; 76:58–63. https://doi.org/10.1136/jnnp.2003.017897 PMID: 15607996

74. Findler M, Cantor J, Haddad L, Gordon W, Ashman T. The reliability and validity of the SF-36 health survey questionnaire for use with individuals with traumatic brain injury. Brain Injury. 2001; 15(8):715–23. https://doi.org/10.1080/02699050010013941 PMID: 11485611

75. Fleming MK, Newham DJ, Roberts-Lewis SF, Sorinola IO. Self-perceived utilization of the paretic arm in chronic stroke requires high upper limb functional mobility. Archives of Physical Medicine and Rehabilitation. 2014; 95:918–24. https://doi.org/10.1016/j.apmr.2014.01.009 PMID: 24480335

76. Freeman JA, Hobart JC, Langdon DW, Thompson AJ. Clinical appropriateness: a key factor in outcome measure selection: the 36 item short form health survey in multiple sclerosis. Journal of Neurology, Neurosurgery & Psychiatry. 2000; 68:150–6. https://doi.org/10.1136/jnnp.68.2.150 PMID: 10644779

77. Freeman JA, Langdon DW, Hobart JC, Thompson AJ. Health-Related Quality if Life in people with Multiple Sclerosis undergoing inpatient rehabilitation. Journal of Neurological Rehabilitation. 1996; 10 (3):185–94.

78. Gillard PJ, Sucharew H, Kleindorfer D, Belagaje S, varon SF, Alwell K, et al. The negative impact of spasticity on the health-related quality of life of stroke survivors: a longitudinal cohort study. Health and Quality of Life Outcomes. 2015; 13. https://doi.org/10.1186/s12955-015-0340-3 PMID: 26415945

79. Goodkin DE, Hertsgaard D, Seminary J. Upper extremity function in multiple sclerosis: Improving assessment sensitivity with Box-and-Block and Nine-Hole Peg Tests. Archives of Physical Medicine and Rehabilitation. 1988; 69:850–4. PMID: 3178453

80. Gowland AC. Staging Motor Impairment After Stroke. Stroke. 1990; 21(9 Suppl II):II-19–II-21. PMID: 2399544

81. Gowland C, Stratford P, Ward M, Moreland J, Torresin W, Van Hullenaar S, et al. Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment. Stroke. 1993; 24(1):58–63. https://doi.org/10.1161/01.str.24.1.58 PMID: 8418551

82. Grant C, Goldsmith CH, Anton HA. Inpatient stroke rehabilitation lengths of stay in Canada derived from the national rehabilitation reporting system, 2008 and 2009. Archives of Physical Medicine and Rehabilitation. 2014; 95:74–8. https://doi.org/10.1016/j.apmr.2013.08.014 PMID: 24001444

83. Green J, Forster A, Young J. A test-retest reliability study of the Barthel Index, the Rivermead Mobility Index, the Nottingham extended Activities of Daily Living Scale and the Frenchay Activities Index in stroke patients. Disability and Rehabilitation. 2001; 23(15):670–6. https://doi.org/10.1080/09638280110045382 PMID: 11720117

84. Guilfoyle MR, Seeley HM, Corteen E, Harkin C, Richards H, Menon DK, et al. Assessing quality of life after traumatic brain in jury: examination of the Short Form 36 Health Survey. Journal of Neurotrauma. 2010; 27:2173–81. https://doi.org/10.1089/neu.2010.1353 PMID: 20939701

85. Hagen S, Bugge C, Alexander H. Psychometric properties of the SF-36 in the elderly post-stroke phase. Journal of Advanced Nursing. 2003; 44(5):461–8. https://doi.org/10.1046/j.0309-2402.2003.02829.x PMID: 14651694

86. Hall KM, Hamilton BB, Gordon WA, Zasler ND. Characteristics and comparisons of functional assessment indices: Disability rating scale, functional independence measure and functional assessment measure. Journal of Head Trauma Rehabilitation. 1993; 8(2):60–74.

87. Hamilton BB, Granger CV. Disability outcomes following inpatient rehabilitation for stroke. Physical Therapy. 1994; 74(5):494–503. https://doi.org/10.1093/ptj/74.5.494 PMID: 8171110

88. Harris JE, Eng JJ. Paretic upper-limb strength best explains arm activity in people with stroke. Physical Therapy. 2007; 87(1):88–97. https://doi.org/10.2522/ptj.20060065 PMID: 17179441

89. Hawthorne G, Gruen RL, Kaye AH. Traumatic brain injury and long-term quality of life: Findings from an Australian study. Journal of Neurotrauma. 2009; 26:1623–33. https://doi.org/10.1089/neu.2008.0735 PMID: 19317590

90. Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of Health-Related Quality of Life. Quality of Life Research 1999; 8(3):209–24. https://doi.org/10.1023/a:1008815005736 PMID: 10472152

91. Heinemann AW, Kirk P, Hastie BA, Semik P, Hamilton BB, Linacre JM, et al. Relationships between disability measures and nursing effort during medical rehabilitation for patients with traumatic brain and spinal cord injury. Archives of Physical Medicine and Rehabilitation. 1997; 78:143–9. https://doi.org/10.1016/s0003-9993(97)90255-0 PMID: 9041894

92. Heinemann AW, Linacre JM, Wright BD, Hamilton BB, Granger C. Relationships between impairment and physical disability as measured by the Functional Independence Measure. Archives of Physical Medicine and Rehabilitation. 1993; 74:566–73. https://doi.org/10.1016/0003-9993(93)90153-2 PMID: 8503745

93. Heinemann AW, Linacre JM, Wright BD, Hamilton BB, Granger C. Measurement characteristics of the Functional Independence Measure. Topics in Stroke Rehabilitation. 1994; 1(3):1–15. https://doi.org/10.1080/10749357.1994.11754030 PMID: 27680951

94. Heller A, Wade DT, Wood VA, Sunderland A, Hewer RL, Ward E. Arm function after stroke: measurement and recovery over the first three months. Journal of Neurology, Neurosurgery, and Psychiatry. 1987; 50(6):714–9. https://doi.org/10.1136/jnnp.50.6.714 PMID: 3612152

95. Herrmann BP, Vickrey B, Hays RD, Cramer J, Devinsky O, Meador K, et al. A comparison of health-related quality of life in patients with epilepsy, diabetes and multiple sclerosis Epilepsy Research. 1996; 25(2):113–8. https://doi.org/10.1016/0920-1211(96)00024-1 PMID: 8884169

96. Hobart JC, Williams LS, Moran K, Thompson AJ. Quality of life measurement after stroke: Uses and abuses of the SF-36. Stroke. 2002; 33:1348–56. https://doi.org/10.1161/01.str.0000015030.59594.b3 PMID: 11988614

97. Houlden H, Edwards M, McNeil J, Greenwood R. Use of the Barthel Index and Functional Independence Measure during early inpatient rehabilitation after single incident brain injury. Clinical Rehabilitation. 2006; 20:153–9. https://doi.org/10.1191/0269215506cr917oa PMID: 16541936

98. Jacob-Lloyd HA, Dunn OM, Brain ND, Lamb SE. Effective measurement of the functional progress of stroke clients. British Journal of Occupational Therapy. 2005; 68(6):253–9.

99. Jenkinson C, Fitzpatrick R, Crocker H, Peters M. The Stroke Impact Scale: Validation in a UK setting and development of a SIS Short Form and SIS Index,. Stroke. 2013; 44:2532–5. https://doi.org/10.1161/STROKEAHA.113.001847 PMID: 23868278

100. Johnson L, Selfe J. Measurement of mobility following stroke: a comparison of the Modified Rivermead Mobility Index and the Motor Assessment Scale. Physiotherapy. 2004; 90(3):132–8.

101. Jones F. The accuracy of predicting functional recovery in patients following a stroke by physiotherapists and patients. Physiotherapy Research International. 1998; 3(4):244–56. https://doi.org/10.1002/pri.148 PMID: 9859133

102. Joyce BM, Rockwood KJ, Mate-Kole CC. Use of goal attainment scaling in brain injury in a rehabilitation hospital. American Journal of Physical Medicine and Rehabilitation. 1994; 73(1):10–4. PMID: 8305175

103. Khan A, Chien CW, Brauer SG. Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. journal of Clinical Epidemiology. 2013; 66:681–7. https://doi.org/10.1016/j.jclinepi.2012.12.014 PMID: 23523550

104. Khan F, Pallant J, Turner-Stokes L. Use of goal attainment scaling in inpatient rehabilitation for persons with multiple sclerosis. Archives of Physical Medicine and Rehabilitation. 2008; 89:652–9. https://doi.org/10.1016/j.apmr.2007.09.049 PMID: 18373995

105. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. Adv Clin Rehabil. 1987; 1:6–18. PMID: 3503663

106. Kohn CG, Sidovar MF, Kaur K, Zhu Y, Coleman CI. Estimating a minimal clinically important difference for the EuroQol 5-dimension health status index in persons with multiple sclerosis. Health and Quality of Life Outcomes. 2014; 12(66). https://doi.org/10.1186/1477-7525-12-66 PMID: 24886430

107. Kuspinar A, Finch LE, Pickard S, Mayo NE. Using existing data to identify candidate items for a health state classification system in multiple sclerosis. Quality of Life Research. 2014; 23:1445–57. https://doi.org/10.1007/s11136-013-0604-5 PMID: 24338161

108. Kuspinar A, Mayo NE. Do generic utility measures capture what is important to the quality of life of people with multiple sclerosis? Health and Quality of Life Outcomes. 2013; 11(71). https://doi.org/10.1186/1477-7525-11-71 PMID: 23618072

109. Kuys SS, Bew PG, Lynch MR, Morrison G, Brauer SG. Measures of activity limitation on admission to rehabilitation after stroke predict walking speed at discharge: an observational study. Australian Journal of Physiotherapy. 2009; 55(4):265–8. https://doi.org/10.1016/s0004-9514(09)70006-6 PMID: 19929769

110. Kwon S, Duncan P, Studenski S, Perera S, Lai SM, Reker D. Measuring stroke impact with SIS: Construct validity of SIS telephone administration. Quality of Life Research. 2006; 15(3):367–76. https://doi.org/10.1007/s11136-005-2292-2 PMID: 16547774

111. Kwon S, Hartzema AG, Duncan PW, Min Lai S. Disability measures in stroke: Relationship among the Barthel Index, the Functional Independence Measure, and the Modified Rankin Scale. Stroke. 2004; 35(4):918–23. https://doi.org/10.1161/01.STR.0000119385.56094.32 PMID: 14976324

112. Lai SM, Studenski S, Duncan PW, Perera S. Persisting consequences of stroke measured by the Stroke Impact Scale. Stroke. 2002; 33(7):1840–18440000019289.15440.F2. https://doi.org/10.1161/01.str.0000019289.15440.f2 PMID: 12105363

113. Lang CE, Edwards DF, Birkenmeier RL, Dromerick AW. Estimating minimal clinically important differences of upper-extremity measures early after stroke. Archives of Physical Medicine and Rehabilitation. 2008; 89(9):1693–700. https://doi.org/10.1016/j.apmr.2008.02.022 PMID: 18760153

114. Lang CE, Wagner JM, Dromerick AW, Edwards DF. Measurement of upper-extremity function early after stroke: Properties of the Action Research Arm Test. Archives of Physical Medicine and Rehabilitation. 2006; 87(12):1605–10. https://doi.org/10.1016/j.apmr.2006.09.003 PMID: 17141640

115. Lannin N. Goal attainment scaling allows program evaluation of a home-based occupational therapy program. Occupational Therapy in Health Care. 2003; 17(1):43–54. https://doi.org/10.1080/J003v17n01_04 PMID: 23941188

116. Lannin NA. Reliability, validity and factor structure of the upper limb subscale of the Motor Assessment Scale (UL-MAS) in adults following stroke. Disability and Rehabilitation. 2004; 26(2):109–15. https://doi.org/10.1080/0963828032000157970 PMID: 14668148

117. Lincoln N, Leadbitter D. Assessment of Motor Function in Stroke Patients. Physiotherapy. 1979; 65 (2):48–51. PMID: 441189

118. Loewen SC, Anderson BA. Reliability of the Modified Motor Assessment Scale and the Barthel Index. Physical Therapy. 1988; 68(7):1077–81. https://doi.org/10.1093/ptj/68.7.1077 PMID: 3387463

119. Loewen SC, Anderson BA. Predictors of stroke outcome using objective measurement scales. Stroke. 1990; 21(1):78–81. https://doi.org/10.1161/01.str.21.1.78 PMID: 2300994

120. Lyle CR. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. International Journal of Rehabilitation Research. 1981; 4(4):483–92. https://doi.org/10.1097/00004356-198112000-00001 PMID: 7333761

121. Mackenzie JE, McCarthy LM, Ditunno FJ, Forrester-Staz SC, Gruen WG, Marion CD, et al. Using the SF-36 for Characterizing Outcome after Multiple Trauma Involving Head Injury. The Journal of Trauma: Injury, Infection, and Critical Care. 2002; 52(3):527–34. https://doi.org/10.1097/00005373-200203000-00018 PMID: 11901330

122. Madden S, Hopman WM, Bagg S, Verner J, O'Callaghan CJ. Functional status and health-related quality of life during inpatient stroke rehabilitation. American Journal of Physical Medicine and Rehabilitation. 2006; 85(10):831–8. https://doi.org/10.1097/01.phm.0000240666.24142.f7 PMID: 16998430

123. Mahoney FI, Barthel DW. Functional evaluation: The Barthel Index. Maryland State Medical Journal. 1965; 14:61–5. PMID: 14258950

124. Malec JF. Goal attainment scaling in rehabilitation. Neuropsychological Rehabilitation. 1999; 9(3/4):253–75.

125. Malec JF, Smigielski JS, DePompolo RW. Goal attainment scaling and outcome measurement in post-acute brain injury rehabilitation. Archives of Physical Medicine and Rehabilitation. 1991; 72(2):138–43. PMID: 1991015

126. Miller KJ, Slade AL, Pallant JF, Galea MP. Evaluation of the psychometric properties of the upper limb subscales of the motor assessment scale using a rasch analysis model. Journal of Rehabilitation Medicine. 2010; 42(4):315–22. https://doi.org/10.2340/16501977-0519 PMID: 20461333

127. Moore F, Wolfson C, Alexandrov L, Lapierre Y. Do general and multiple sclerosis-specific quality of life instruments differ? Canadian Journal of Neurological Sciences. 2004; 31(1):64–71. https://doi.org/10.1017/s0317167100002857 PMID: 15038473

128. Moreland J, Gowland C, Van Hullenaar S, Huijbregts M. Theoretical basis of the Chedoke-McMaster Stroke Assessment. Physiotherapy Canada. 1993; 45(4):231–8.

129. Morris JH, van Wijck F, Joice S, Donaghy M. Predicting health related quality of life 6 months after stroke: the role of anxiety and upper limb dysfunction. Disability and Rehabilitation. 2013; 35(4):291–9. https://doi.org/10.3109/09638288.2012.691942 PMID: 22691224

130. Murrell RC, Kenealy PM, Graham Beaumont J, Lintern TC. Assessing quality of life in persons with severe neurological disability associated with multiple sclerosis: The psychometric evaluation of two quality of life. measures. British Journal of Health Psychology. 1999; 4(4):349–62.

131. Nicholl CR, Lincoln NB, Francis VM, Stephan TF. Assessing quality of life in people with multiple sclerosis. Disability and Rehabilitation. 2001; 23(14):597–603. https://doi.org/10.1080/09638280110043933 PMID: 11697457

132. Oczkowski W, Barreca S. The functional independence measure: Its use to identify rehabilitation needs in stroke survivors. Archives of Physical Medicine and Rehabilitation. 1993; 74(12):1291–4. https://doi.org/10.1016/0003-9993(93)90081-k PMID: 8259894

133. O'Mahony PG, Rodgers H, Thomson RG, Dobson R, James OFW. Is the SF-36 suitable for assessing health status of older stroke patients? Age and Ageing. 1998; 27(1):19–22. https://doi.org/10.1093/ageing/27.1.19 PMID: 9504362

134. Ouellette DS, Timple C, Kaplan SE, Rosenberg SS, Rosario ER. Predicting discharge destination with admission outcome scores in stroke patients. NeuroRehabilitation. 2015; 37(2):173–9. https://doi.org/10.3233/NRE-151250 PMID: 26484509

135. Peters M, Crocker H, Dummett S, Jenkinson C, Doll H, Fitzpatrick R. Change in health status in long-term conditions over a one year period: a cohort survey using patient-reported outcome measures. Quality of Life Outcomes. 2014; 12(1).

136. Pickard AS, Johnson JA, Feeny DH. Responsiveness of generic health-related quality of life measures in stroke. Quality of Life Research. 2005; 14(1):207–19. https://doi.org/10.1007/s11136-004-3928-3 PMID: 15789955

137. Pickering RL, Hubbard IJ, Baker KG, Parsons MW. Assessment of the upper limb in acute stroke: The validity of hierarchical scoring for the Motor Assessment Scale. Australian Occupational Therapy Journal. 2010; 57(3):174–82. https://doi.org/10.1111/j.1440-1630.2009.00810.x PMID: 20854586

138. Pittock SJ, Mayr WT, McClelland RL, Jorgensen NW, Weigand SD, Noseworthy JH, et al. Quality of life is favorable for most patients with multiple sclerosis. Archives of Neurology. 2004; 61(5):679–86. https://doi.org/10.1001/archneur.61.5.679 PMID: 15148144

139. Poole JL, Nakamoto TN, McNulty T, Montoya JR, Weill D, Dieruf K, et al. Dexterity, visual perception, and activities of daily living in persons with multiple sclerosis. Occupational Therapy in Health Care. 2010; 24(2):159–70. https://doi.org/10.3109/07380571003681202 PMID: 23898901

140. Rabadi MH, Rabadi FM. Comparison of the Action Research Arm Test and the Fugl-Meyer Assessment as measures of upper-extremity motor weakness after stroke Archives of Physical Medicine and Rehabilitation. 2006; 87(7):962–6. https://doi.org/10.1016/j.apmr.2006.02.036 PMID: 16813784

141. Rabadi MH, Vincent AS. Comparison of the Kurtkze Expanded Disability Status Scale and the Functional Independence Status Scale and the Functional Independence Measure: measures of multiple sclerosis-related disability. Disability and Rehabilitation. 2013; 35(22):1877–84. https://doi.org/10.3109/09638288.2013.766269 PMID: 23600712

142. Rand D, Eng JJ. Predicting daily use of the affected upper extremity 1 year after stroke. Journal of Stroke and Cerebrovascular Diseases. 2015; 24(2):274–83. https://doi.org/10.1016/j.jstrokecerebrovasdis.2014.07.039 PMID: 25533758

143. Riazi A, Hobart JC, Lamping DL, Fitzpatrick R, Freeman JA, Jenkinson C, et al. Using the SF-36 measure to compare the health impact of multiple sclerosis and Parkinson's disease with normal population health profiles. Journal of Neurology, Neurosurgery & Psychiatry. 2003; 74(6):710–4. https://doi.org/10.1136/jnnp.74.6.710 PMID: 12754336

144. Rigby H, Gubitz G, Eskes G, Reidy Y, Christian C, Grover V, et al. Caring for stroke survivors: baseline and 1-year determinants of caregiver burden. International Journal of Stroke. 2009; 4(3):152–8. https://doi.org/10.1111/j.1747-4949.2009.00287.x PMID: 19659814

145. Robinson D Jr, Zhao N, Gathany T, Kim L-L, Cella D, Revicki D. Health perceptions and clinical characteristics of relapsing-remitting multiple sclerosis patients: baseline data from an international clinical trial. Current Medical Research and Opinion. 2009; 25(5):1121–30. https://doi.org/10.1185/03007990902797675 PMID: 19317608

146. Sabari JS, Lim AL, Velozo CA, Lehman L, Kieran O, Lai JS. Assessing arm and hand function after stroke: a validity test of the hierarchical scoring system used in the motor assessment scale for stroke. Archives of Physical Medicine and Rehabilitation. 2005; 86(8):1609–15. https://doi.org/10.1016/j.apmr.2004.12.028 PMID: 16084815

147. Sackley CM. The relationship between weight-bearing asymmetry after stroke, motor function and activities of daily living. Physiotherapy Theory and Practice. 1990; 6(4):179–85.

148. Salter KL, Moses MB, Foley NC, Teasell RW. Health-related quality of life after stroke: what are we measuring? International Journal of Rehabilitation Research. 2008; 31(2):111–7. https://doi.org/10.1097/MRR.0b013e3282fc0f33 PMID: 18467925

149. Sarker S-J, Rudd AG, Douiri A, Wolfe CDA. Comparison of 2 extended activities of daily living scales with the Barthel Index and predictors of their outcomes: Cohort study within the South London Stroke Register (SLSR). Stroke. 2012; 43(5):1362–9. https://doi.org/10.1161/STROKEAHA.111.645234 PMID: 22461336

150. Schwid SR, Goodman AD, McDermott MP, Bever CF, Cook SD. Quantitative functional measures in MS: What is a reliable change? Neurology. 2002; 58(8):1294–6. https://doi.org/10.1212/wnl.58.8.1294 PMID: 11971105

151. Sharrack B, Hughes RAC, Soudain S, Dunn G. The psychometric properties of clinical rating scales used in multiple sclerosis. Brain. 1999; 122(1):141–59. https://doi.org/10.1093/brain/122.1.141 PMID: 10050902

152. Simon C, Kumar S, Kendrick T. Formal support of stroke survivors and their informal carers in the community: a cohort study. Health and Social Care in the Community. 2008; 16(6):582–92. https://doi.org/10.1111/j.1365-2524.2008.00782.x PMID: 18371168

153. Stineman MG, Shea JA, Jette A, Tassoni CJ, Ottenbacher KJ, Fiedler R, et al. The functional independence measure: Test of scaling assumptions, structure, and reliability across 20 diverse impairment categories. Archives of Physical Medicine and Rehabilitation. 1996; 77(11):1101–8. https://doi.org/10.1016/s0003-9993(96)90130-6 PMID: 8931518

154. Stone SP, Patel P, Greenwood RJ. Selection of acute stroke patients for treatment of visual neglect. Journal of Neurology, Neurosurgery & Psychiatry. 1993; 56(5):463–6. https://doi.org/10.1136/jnnp.56.5.463 PMID: 8505635

155. Sturm JW, Osborne RH, Dewey HM, Donnan GA, Macdonell RAL, Thrift AG. Brief comprehensive quality of life assessment after stroke: The assessment of quality of life instrument in the north east Melbourne stroke incidence study (NEMESIS). Stroke. 2002; 33(12):2888–94. https://doi.org/10.1161/01.str.0000040407.44712.c7 PMID: 12468787

156. Turner-Stokes L, Baguley IJ, De Graaff S, Katrak P, Davies L, McCrory P, et al. Goal attainment scaling in the evaluation of treatment of upper limb spasticity with botulinum toxin: a secondary analysis from a double-blind placebo-controlled randomized clinical trial. Journal of Rehabilitation Medicine. 2010; 42(1):81–9. https://doi.org/10.2340/16501977-0474 PMID: 20111849

157. Uswatte G, Taub E. Implications of the learned nonuse formulation for measuring rehabilitation outcomes: Lessons from constraint-induced movement therapy. Rehabilitation Psychology. 2005; 50(1):34–42.

158. Uswatte G, Taub E, Morris D, Light K, Thompson PA. The motor activity log-28: Assessing daily use of the hemiparetic arm after stroke. Neurology. 2006; 67(7):1189–94. https://doi.org/10.1212/01.wnl.0000238164.90657.c2 PMID: 17030751

159. van der Putten JJMF, Hobart JC, Freeman JA, Thompson AJ. Measuring change in disability after inpatient rehabilitation: comparison of thew responsiveness of the Barthel Index and the Functional Independence Measure. Journal of Neurology, Neurosurgery & Psychiatry. 1999; 66(4):480–4.

160. van Straten A, de Haan RJ, Limburg M, Shuling J, Bossuyt PM, van den Bos GAM. A Stroke-adapted 30-item version of the sickness impact profile to assess quality of life (SA-SIP30). A Journal of Cerebral Circulation. 1997; 28(11):2155–61. https://doi.org/10.1161/01.str.28.11.2155 PMID: 9368557

161. Vickrey BG, Hays RD, Genovese BJ, Myers LW, Ellison GW. Comparison of a generic to disease-targeted health-related quality-of-life measures for multiple sclerosis. Journal of Clinical Epidemiology. 1997; 50(5):557–69. https://doi.org/10.1016/s0895-4356(97)00001-2 PMID: 9180648

162. Vickrey BG, Hays RD, Harooni R, Myers LW, Ellison GW. A health-related quality of life measure for multiple sclerosis. Quality of Life Research. 1995; 4(3):187–206. https://doi.org/10.1007/BF02260859 PMID: 7613530

163. Wade DT, Hewer RL. Functional abilities after stroke: measurement, natural history and prognosis. Journal of Neurology, Neurosurgery & Psychiatry. 1987; 50(2):177–82. https://doi.org/10.1136/jnnp.50.2.177 PMID: 3572432

164. Wallace D, Duncan PW, Min Lai S. Comparison of the responsiveness of the Barthel Index and the Motor component of the Functional Independence Measure in stroke: The impact of using different methods for measuring responsiveness. Journal of Clinical Epidemiology. 2002; 55(9):922–8. https://doi.org/10.1016/s0895-4356(02)00410-9 PMID: 12393081

165. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36) I. Conceptual framework and item selection. Medical Care. 1992; 30(6):473–83. PMID: 1593914

166. Wellwood I, Dennis MS, Warlow CP. A comparison of the Barthel Index and the OPCS Disability Instrument used to measure outcome after acute stroke. Age and Ageing. 1995; 24(1):54–7. https://doi.org/10.1093/ageing/24.1.54 PMID: 7762463

167. Wilkinson PR, Wolfe CDA, Warburton FG, Rudd AG, Howard RS, Ross-Russell RW, et al. Longer term quality of life and outcome in stroke patients: is the Barthel index alone an adequate measure of outcome. Quality of Life Outcomes. 1997; 6(3):125–30.

168. Williams LS, Weinberger M, Harris LE, Biller J. Measuring quality of life in a way that is meaningful to stroke patients. Neurology. 1999; 53(8):1839–43. https://doi.org/10.1212/wnl.53.8.1839 PMID: 10563636

169. Williams A. EuroQol—a new facility for the measurement of health-related quality of life. Health policy. 1990; 16(3):199–208. https://doi.org/10.1016/0168-8510(90)90421-9 PMID: 10109801

170. Wolf T, Koster J. Perceived recovery as a predictor of physical activity participation after mild stroke. Disability and Rehabilitation. 2013; 35(14):1143–8. https://doi.org/10.3109/09638288.2012.720635 PMID: 23013280

171. Xie J, Wu EQ, Zheng ZJ, Croft JB, Greenlund KJ, Mensah GA, et al. Impact of stroke on health-related quality of life in the noninstitutionalized population in the United States. Stroke. 2006; 37(10):2567–72. https://doi.org/10.1161/01.STR.0000240506.34616.10 PMID: 16946158

172. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. Neurorehabilitation and Neural Repair. 2008; 22(1):78–90. https://doi.org/10.1177/1545968307305353 PMID: 17704352

173. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. International Journal of Rehabilitation Research. 1981; 4(4):483–92. https://doi.org/10.1097/00004356-198112000-00001 PMID: 7333761

174. Carroll D. A quantitative test of upper extremity function. Journal of Chronic Disability. 1965; 18:479–91. https://doi.org/10.1016/0021-9681(65)90030-5 PMID: 14293031

175. Brashear A, Gordon MF, Elovic E, Kassicieh VD, Marciniak C, Do M, et al. Intramuscular injection of Botulinum Toxin for the treatment of wrist and finger spasticity after a stroke. The New England Journal of Medicine. 2002; 347:395–400. https://doi.org/10.1056/NEJMoa011892 PMID: 12167681

176. Kiresuk TJ, Sherman RE. Goal attainment scaling: a general method for evaluating comprehensive community mental health programs. Community Mental Health Journal. 1968; 4(6):443–53. https://doi.org/10.1007/BF01530764 PMID: 24185570

177. MacKenzie EJ, McCarthy ML, Ditunno JF, Forrester-Staz C, Gruen GS, Marion DW, et al. Using the SF-36 for characterizing outcome after multiple trauma involving head injury. The Journal of Trauma. 2002; 52(3):527–34. https://doi.org/10.1097/00005373-200203000-00018 PMID: 11901330

178. Kellor M, Frost J, Silberberg N, Iversen I, Cummings R. Norms for clinical use: Hand strength and dexterity. The American Journal of Occupational Therapy. 1971; 25(2):77–83. PMID: 5551515

179. Mathiowetz V, Weber K, Kashman N, Volland G. Adult norms for the nine hole peg test of finger dexterity. The Occupational Therapy Journal of Research. 1985; 5(1):24–38.

180. Alusi SH, Worthington J, Glickman S, Findley LJ, Bain PG. Evaluation of three different ways of assessing tremor in multiple sclerosis. Journal of Neurology, Neurosurgery & Psychiatry. 2000; 68:756–60. https://doi.org/10.1136/jnnp.68.6.756 PMID: 10811700

181. Bamford JM, Sandercock PA, Warlow CP, Slattery J. Interobserver agreement for the assessment of handicap in stroke patients. Stroke. 1989; 20(6):828. https://doi.org/10.1161/01.str.20.6.828 PMID: 2728057

182. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthritis and Cartilage. 2012; 20(12):1548–62. https://doi.org/10.1016/j.joca.2012.08.015 PMID: 22944525

183. Wales K, Clemson L, Lannin N, Cameron I. Functional Assessments Used by Occupational Therapists with Older Adults at Risk of Activity and Participation Limitations: A Systematic Review. PLOS ONE. 2016; 11(2):e0147980. https://doi.org/10.1371/journal.pone.0147980 PMID: 26859678

184. Zaki R, Bulgiba A, Nordin N, Ismail N. A Systematic Review of Statistical Methods Used to Test for Reliability of Medical Instruments Measuring Continuous Variables. Iranian Journal of Basic Medical Sciences. 2013; 16(6):803–7. PMID: 23997908

185. Sikorskii A, Noble PC. Statistical considerations in the psychometric validation of outcome measures. Clin Orthop Relat Res. 2013; 471(11):3489–95. https://doi.org/10.1007/s11999-013-3028-1 PMID: 23645337

Australian Occupational Therapy Journal

Australian Occupational Therapy Journal (2018) **65**, 449–471 doi: 10.1111/1440-1630.12527

Special Issue Article

A systematic review of the psychometric properties of the Action Research Arm Test in neurorehabilitation

Shannon Pike,[1,2] Natasha A. Lannin,[1,3] Kylie Wales[4] and Anne Cusick[5]

[1]*School of Allied Health (Occupational Therapy), La Trobe University,* [3]*Occupational Therapy Department, Alfred Health, Melbourne,* [2]*Wagga Wagga Ambulatory Rehabilitation Service, Murrumbidgee Local Health District, Wagga Wagga,* [4]*School of Health Sciences, Faculty of Health and Medicine, The University of Newcastle, Callaghan, New South Wales, and* [5]*Discipline of Occupational Therapy, The University of Sydney, New South Wales, Australia*

***Background/aim*:** *The Action Research Arm Test (ARAT) measures upper limb activity limitations in people with acquired brain injuries. Evidence relating to the use of this test in neurorehabilitation is scattered. This review identifies, rates and synthesises evidence on the original 1981 ARAT use within neurorehabilitation. Psychometric properties are reviewed, including specific examination of participants with upper limb spasticity.*

***Methods*:** *Systematic review of published articles describing psychometric properties and/or use of the original version of the ARAT in neurorehabilitation. **CO**nsensus-based **S**tandards for the selection of health **M**easurement **IN**struments (COSMIN) search strategy, reporting and methodological checklist with criterion-based appraisal of quality criteria for good measurement properties were applied. A best evidence synthesis for each psychometric property was completed.*

***Results*:** *In 28 included studies, participants had suffered a stroke or traumatic brain injury, with 46% >6 months post-injury. Six studies identified participants with upper limb spasticity. Methodological quality of psychometric properties ranged from poor to excellent. Best evidence synthesis determined moderate positive evidence for using the ARAT with people without limb spasticity: intra-rater reliability (ICC 0.71 (95% CI 0.53–0.89) to 0.99 (95% CI 0.98, 0.99)); responsiveness (ROC curve 0.72–0.88, SRM 0.89); and regarding construct validity, it is a valid measure of activity limitation. Limited evidence for psychometric properties of the ARAT were found when used with people with upper limb spasticity for construct validity and responsiveness (ES 0.55–0.78). Gaps in evidence were found for inter and test–retest reliability, measurement error, content validity, structural validity, floor and ceiling effects.*

***Conclusions*:** *The ARAT is an appropriate measure of activity limitation post-stroke and should be considered for use with people with TBI; evidence for people with upper limb spasticity is limited. Gaps and mixed limited to moderate evidence for psychometric properties in neurorehabilitation mean further research is required.*

KEY WORDS *activities of daily living, neurological rehabilitation, outcome assessment (health care), psychometrics, upper limb.*

Introduction

Neurorehabilitation outcome measurement is complicated by highly variable clinical presentations, along with diverse individualised and person-centred intervention goals. Occupational therapists seek outcome measures that can accommodate the diverse and individualised nature of neurorehabilitation at the same time as providing meaningful, sensitive and reliable data on which to base decisions and plans. Occupational therapists working in neurorehabilitation are particularly interested in the effect of interventions on attainment of "real life" activity and participation goals. Outcome measures thus need to capture the complexity of factors contributing to performance of everyday tasks and participation in "real life" contexts.

The Action Research Arm Test (ARAT) (Lyle, 1981) is a standardised observational performance measure that evaluates a person's ability to use their upper limb to

Shannon Pike BHlthSc(OT); PhD Candidate. **Natasha A. Lannin** PhD; Associate Professor in Occupational Therapy. **Kylie Wales** PhD; Lecturer; Occupational Therapy. **Anne Cusick** PhD; Professor and Chair – Discipline of Occupational Therapy.

Correspondence: Shannon Pike,Wagga Wagga Ambulatory Rehabilitation Service, Murrumbidgee Local Health District, PO Box 159 Wagga Wagga, NSW 2650, Australia. Email: shannon.pike@health.nsw.gov.au

Conflict of interest
All authors declare that they have no conflicts of interest.

Accepted for publication 6 August 2018.

handle objects using grasp, grip, pinch and gross motor movements. These movements are needed to perform many everyday tasks. For this reason, an inability to perform test items is thus proposed to be a valid indicator of upper limb activity limitation (Kwakkel *et al*., 2017). The ARAT has been demonstrated to be unidimensional, measuring the single construct of upper limb function related to everyday activities (Koh *et al*., 2006; Van der Lee, Roorda, Beckerman, Lankhorst & Bouter, 2002). Furthermore, it has been found to hold concurrent validity with other tests of activity limitation, including the Wolf Motor Function Test (WMFT), Motor Activity Log (MAL), Stroke Impact Scale (SIS) hand function items (Lin, Chuang, Wu, Hsieh & Chang, 2010; Lin *et al*., 2009). To date, no review has used consensus standards to examine existing evidence regarding psychometric properties of the ARAT when used with adults who have neurological conditions and are undergoing neurorehabilitation and experience spasticity. This study aims to fill this information gap.

Development of the ARAT

The ARAT was developed by Ronald Lyle in 1981 and is based on Carroll's Upper Extremity Function Test (UEFT) (Carroll, 1965). Theoretically, the UEFT assumed that "upper extremity movements used in daily activities can be reduced to certain patterns", and that observation of these patterned movements could provide information to monitor upper extremity function related to everyday activities (Carroll, p. 479). The patterns in the UEFT were grouped as grip; lateral prehension; pinch; placing; pronation and supination; and writing. In the ARAT, Lyle sought to adapt the UEFT to shorten administration from 60 minutes. Lyle questioned Carroll's grouping of items into five patterns and performed correlational analysis, item reduction and hierarchical scoring order to produce what became known as the ARAT – a "rapid yet reliable" measurement tool to measure "changes in upper limb function" (Lyle, 1981, p. 483).

Description of the ARAT

There are a number of adaptations of the original (1981) ARAT. This study of psychometric properties examines only the original version, hereby called "original ARAT" in the remainder of the paper. This is a 19-item performance test where participants are asked to complete movements and handle objects with each upper limb, starting with the less affected limb. Item performance is not timed. The items are organised into four subscales:

- Grasp (of differing sized blocks, a ball and stone);
- Grip (pouring water between glasses, moving differing diameter tubes placed vertically, placing washer over bolt);
- Pinch (of various sized marbles and ball bearing between thumb and individual digit combinations and place on shelf); and
- Gross movement (of hand behind and on top of head and to mouth without objects).

The test is standardised through the description of the size and nature of each object to be handled and the action to be performed. The test can be administered anywhere by anyone and no test certification is required. The test kit can be purchased through the test site (http://www.aratest.eu/) or self-assembled using guidelines.

ARAT scoring

Original ARAT items are scored on observation of movement performance using an ordinal four-level scale ranging from 0 to 3. Each limb is scored separately. A score of 0 indicates that no part of the test can be performed; 1 indicates partial test performance; 2 indicates with abnormally long time or great difficulty; 3 indicates normal item performance. Possible total scores range from 0 to 57.

ARAT structure

Items in each subtest are ordered hierarchically, with the most difficult item presented first. If the highest score is obtained for the first item (score of 3), it is inferred that all items less difficult in the subscale could be completed and the person is not required to attempt remaining items and moves to the next subtest. If <3 is obtained in a subtest, the second and easiest item is attempted. If 0 is scored, the person is deemed to score 0 on all other subtest items. If 1 or 2 is scored, they must attempt all remaining subtest items. This hierarchical nature of the original ARAT speeds administration for people with high or low performance in subtests.

ARAT implementation

The original ARAT provided clinicians and researchers with instructions that were limited in detail giving few specifications in what to observe in administration and how to score in a consistent and standardised manner. This resulted in variable in administration, scoring and interpretation of results across clinicians, researchers and sites. Subsequent guidelines and manuals included additional operational definitions to increase standardisation in administration and scoring (e.g., defining "abnormally long" into specific time frames) (Hsueh, Lee & Hsieh, 2002; Platz, Pinkowski, van Wijck & Johnson, 2005; Platz, Pinkowski, van Wijck, Kim, *et al*., 2005; Wagenaar *et al*., 1990; Yozbatiran, Der-Yeghiaian & Cramer, 2008). None of these post-1981 ARAT versions have been identified as "gold-standard"; thus, only studies using the original ARAT are included in this review.

ARAT uptake

The original ARAT and adaptations are commonly cited as a primary outcome measure within clinical trials and

intervention studies (Santisteban *et al*., 2016). Apart from its popularity, the robust and sensitive nature of the ARAT are indicated by its use as a criterion measure in validation studies for new, existing and/or modified assessments (Barreca, Stratford, Masters, Lambert & Griffiths, 2006; Blennerhassett, Avery & Carey, 2010; Edwards, Lang, Wagner, Birkenmeier & Dromerick, 2012; Page, Hade & Persch, 2015). The ARAT has also been included in clinical practice guidelines and consensus statements relevant to neurorehabilitation, stroke (Kwakkel *et al*., 2017; Sullivan *et al*., 2013) and upper limb spasticity (Sheean, Lannin, Turner-Stokes, Rawicki & Snow, 2010).

Internationally, the ARAT has widespread practice use, including translated versions, however, like many other outcome measures, uptake as measured by published intervention studies particular to stroke using this tool varies from country to country (Santisteban *et al*., 2016). Its published use is most common in versions presented in English, with the United Kingdom and Australia being highest (Santisteban *et al*.). When the ARAT is used in translation, these versions lack validation and translation technique information; thus, only studies using the original ARAT citing Lyle (1981) and in the English language are included in this review.

ARAT sensitivity in practice

An important milestone in research and clinical application of the original ARAT in neurorehabilitation came with the identification of a minimum clinically important change (MCD), although it was specific only to acute and chronic post-stroke samples (Lang, Wagner, Dromerick & Edwards, 2006; Van der Lee *et al*., 2001). The score was 5.7 points or 10% of the total score. The threshold score gave a sound base to inform research and practice decisions regarding interpretation of intervention impacts.

Psychometric properties of the ARAT

Psychometric properties of the original ARAT have been evaluated in studies with patients post-stroke. Many of these studies are included in results of this review and are thus not cited here. Studies excluded on the basis of language but relevant to this introduction show that the original ARAT is unidimensional, measuring the single construct of upper limb function related to everyday activities (Koh *et al*., 2006; Van der Lee *et al*., 2002). Furthermore, it has been demonstrated to hold predictive validity and concurrent validity with similar tests of activity limitation, including the WMFT, MAL and SIS hand function (Lin *et al*., 2009; Lin *et al*., 2010). The ARAT has been demonstrated to be a reliable and responsive measure (Chen, Lin, Wu & Chen, 2012; Hsueh & Hsieh, 2002; Rabadi & Rabadi, 2006).

While a handful of assessment systematic reviews with various foci have included the ARAT, no systematic review has yet synthesised the psychometric properties of any version of the ARAT or the original ARAT specifically using the COnsensus-based **S**tandards for the selection of health **M**easurement **IN**struments (COSMIN) approach to evaluate the methodological quality of included studies or their conformity with consensus-based measurement standards. There has not been a synthesis of the evidence of psychometric properties of the original ARAT in neurorehabilitation. Furthermore, despite the high incidence of patients in neurorehabilitation with upper limb spasticity, there is limited guidance on the use of original ARAT with this clinical population.

Aim

This review will identify and synthesise existing evidence regarding psychometric properties of the original ARAT when used with adults who have a neurological condition and are undergoing neurorehabilitation.

The research questions are:

1 What are the psychometric properties; internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing (construct validity), cross-cultural validity, responsiveness and interpretability of the original ARAT when used with adults with a neurological condition undergoing neurorehabilitation?
2 Are psychometric properties of the original ARAT different when the presence of upper limb spasticity is apparent in the study samples?

Method

This systematic review applied the COSMIN methodological approach (Mokkink *et al*., 2010) supplemented by a quality appraisal proposed by Terwee (Terwee *et al*., 2007, 2017).

Identification and selection of studies

To identify relevant articles, searches were conducted of the following from inception until December 2017: Medical Literature Analysis and Retrieval System Online (MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL) and Excerpta Medica dataBASE (EMBASE). The search strategy is reported in a larger evaluation of upper limb function assessments (Pike *et al*., 2015). Studies were included: if the original version of the ARAT was used with no modifications, in full, with all items administered and data reported; if the study was conducted and reported in English; if any of the psychometric properties defined by COSMIN were investigated; if it was an original study that collected data; if the original ARAT was either the primary outcome measure or was used in such a way that its psychometric properties were still evaluated and reported; and if reports were in peer-reviewed literature. Studies needed to

have participants who were adults (>18 years), who were undergoing rehabilitation with a study sample where there was no less than 90% of participants with a neurological condition diagnosis of stroke, multiple sclerosis, cerebral palsy, traumatic brain injury or anoxia.

Studies reviewed in the subset which included upper limb spasticity were identified by explicit documentation of the presence of upper limb spasticity in participants whose data were reported. For example, Page, Levine and Hade (2012) reported: (a) a Modified Ashworth Scale score ≥3 as an exclusion criterion and (b) nil report of participants with spasticity ≤3 in the study. Page *et al.* were thus deemed a study with nil upper limb spasticity participants present.

One investigator screened all titles and abstracts; potential exclusions were examined by one of two co-investigators and those with agreement were excluded. Others required inspection of the full text by two of three investigators and a consensus decision on inclusion or exclusion was made. Full text was obtained for all included papers, and following full text inspection, final exclusion decisions were made by consensus.

Data collection

An author developed data extraction form in Excel™ was used to record information. The first author examined full text and entered data into the form, referring uncertain aspects for consensus decision by two and/or three of the investigators, whereupon the data were entered into the form. Aspects recorded were study design; participants; description of rehabilitation programmes; outcome measures used; the ICF classification; psychometric properties; and inclusion/exclusion decisions.

Data analysis

The quality of included papers was evaluated using the COSMIN checklist with 4-point scale. This checklist was applied to determine whether each study met the standards for methodological quality with regard to internal consistency, reliability (test–retest, inter-rater and intra-rater), measurement error, content validity (including face validity), structural validity, hypothesis testing, cross-cultural validity, criterion validity, responsiveness, interpretability and generalisability (Pike *et al.*, 2015; Terwee *et al.*, 2012). A rating of "excellent", "good", "fair" or "poor" was assigned for each measurement property.

Terwee's (Terwee *et al.*, 2007, 2017) quality criteria were applied to individual studies to analyse the measurement properties of the original ARAT. The study design, methods and outcomes, for content validity, internal consistency, construct validity, structural validity, IRT/Rasch analysis, reliability, responsiveness, measurement error, floor and ceiling effects and interpretability were evaluated. Criterion validity was considered in relation to the measurement of upper limb activity performance – the decision was made not to evaluate this as no agreed "gold standard" measure of upper limb activity performance exists. Measurement properties were rated as positive '+', indeterminant '?', negative '−' or no information '0'.

The sample size of individual studies was not assessed within the COSMIN data extraction and rating phase, rather sample size was considered in the "evidence synthesis" stage by combining sample sizes from included individual studies – in line with Dobson *et al.*'s (2012) approach.

A best evidence synthesis was completed for each psychometric property based on: methodological quality of reporting studies (COSMIN); rating and consistency of the rating assigned for measurement properties meeting the quality criteria (the "Terwee" criteria); and overall sample size (evidence was assigned 'strong' when total sample size of combined eligible studies was ≥100, 'moderate' with total samples between 50 and 99, 'limited' with total samples between 25 and 49 and 'unknown' with total sample less than 25 (Dobson *et al.*, 2012). The synthesis of best evidence approach was based on that applied by Wales, Clemson, Lannin and Cameron (2016) and Dobson *et al.* and adapted from Terwee *et al.* (2007). Studies with poor methodological quality were *not* included in the best evidence synthesis.

Results

The search strategy identified 711 studies (excluding duplicates). After screening titles, abstracts and full text, 28 of these 711 (4%) were deemed eligible and included for appraisal. Figure 1 presents the flow of papers through the review. The included studies are detailed in Table 1. A summary of study results is detailed in Table 2; the synthesis of best evidence for psychometric properties is within Table 3; and COSMIN and Terwee ratings are outlined in supplementary data within Tables S1 and S2.

Study participants

Twenty-five studies included only post-stroke participants and three were mixed samples, including post-stroke and traumatic brain injury (TBI). There was a total of 1005 participants, 985 with stroke, 15 TBI and 20 participants who could not be differentiated as either stroke or TBI.

Chronicity post brain injury was extracted from studies and assigned >6 or <6 months post brain injury. The split was relatively even, with 46% of the studies, including participants >6 months post their initial brain injury. This percentage included all studies with both mixed sample and TBI diagnoses.

Six of the included studies specifically identified 199 participants with upper limb spasticity; MAS scores ranged from 1 to 3 (182 post-stroke; 15 TBI).

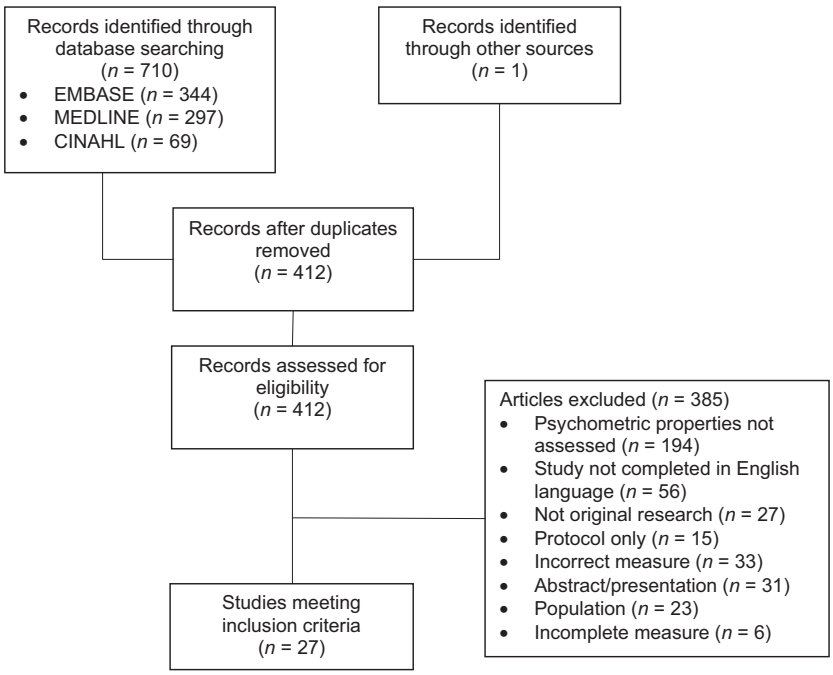


FIGURE 1: *Study inclusion–exclusion process.*

Measurement properties and synthesis of best evidence

Information regarding original ARAT measurement properties was extracted and appraised. This included reliability, measurement error, validity (both structural and construct), responsiveness, interpretability and floor–ceiling effects. Interpretability was examined according to COSMIN guidelines whereby data are extracted only if the study explicitly aimed to assess interpretability through floor and ceiling effects, minimal important change (MIC) and distribution of scores in subgroups. No COSMIN score is assigned for interpretability. Terwee's quality criteria were thus applied to consider interpretability as the qualitative meaning of quantitative scores and floor and ceiling effects (see COSMIN and Terwee analysis for individual studies in supplementary data Tables S1 and S2).

Reliability

Five studies evaluating ARAT reliability were located.

Retest reliability of the original ARAT was examined in one study (McDonnell, Hillier, Ridding & Miles, 2006). This showed that it was reliable between testing sessions ($ICC_{(3,1)} = 0.93 \pm 0.05$) in 17 participants, 2–7 months post-stroke. This study was only of fair methodological quality due to limited methodological detail being available and thus uncertainty regarding independent administrations of repeat measures. *The evidence synthesis resulted in an unknown level of retest reliability* due to a sample size <25 in participants post-stroke with no evidence located for people with a neurological condition with upper limb spasticity.

Three studies evaluated the intra-rater reliability (Page *et al.*, 2012); all included participants >6 months post-stroke with nil identified upper limb spasticity. The studies found a high level of reliability, ICC ranging from 0.71 (95% CI 0.53–0.89) to 0.99 (95% CI 0.98, 0.99) within raters. The methodological quality of ranged from fair to excellent with two studies receiving final ratings of good and all three had positive quality criteria. *This synthesis of evidence found moderate positive evidence to support intra-rater reliability* when used with people >6 months post-stroke. There was no evidence located to support or refute intra-rater reliability when the original ARAT was used with people with upper limb spasticity.

One study examined inter-rater reliability (Yozbatiran *et al.*, 2008). This was high when two blinded raters scored the original ARAT within the same session with nine participants who had a mean 34 months post-stroke (ICC 0.96). The methodological rating of this paper was good with a positive Terwee rating; ratings were reduced due to limited methodological detail provided. *This synthesis of evidence found unknown evidence for inter-rater reliability* due to the small sample size of the single study.

Responsiveness

A total of 10 studies were appraised with only nine considered in the best evidence synthesis stage of this study. These studies had a methodological quality ranging from fair to excellent. Ratings were reduced due to

TABLE 1: *Characteristics of included studies*

Studies included	Summary of study participants	Psychometric property tested
Barden *et al.* (2014)	Diagnosis = Mixed (Stroke *n* = 22, Traumatic Brain Injury *n* = 6) Time since diagnosis *(months)* = greater than 6 *n* = 28 Age *(year)*, mean (SD) = 51 (17) Sex, number male (%) = 15 (54) Sample included people with spasticity = yes	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity ✓ Responsiveness ✓ Interpretability
Barden *et al.* (2012)	Diagnosis = Mixed (Stroke *n* = 29, Traumatic Brain Injury *n* = 9) Time since diagnosis *(months)* = greater than 6 *n* = 38 Age *(year)*, median (range) = 50 (18–81) Sex, number male (%) = 22 (58) Sample included people with spasticity = yes	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross cultural validity Responsiveness ✓ Interpretability
Barreca *et al.* (2005)	Diagnosis = Stroke Time since diagnosis *(months)* = less than 6 *n* = 39 Age *(year)*, mean (SD) = acute grp 71.4 (50.9–90.0) chronic grp 64.0 (44.7–76.6) Sex, number male (%) = 20 (51) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity ✓ Responsiveness Interpretability
Barreca, Stratford, Masters, Lambert & Griffiths (2006)	Diagnosis = Stroke Time since diagnosis *(months)* = less than 6 *n* = 105 Age *(year)*, quartiles = mild-mod impairment 66, 76, 81, severe impairment 59, 69, 77 Sex, number male (%) = 54 (51) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity ✓ Responsiveness Interpretability
Barreca, Stratford, Masters, Lambert, Griffiths & McBay (2006)	Diagnosis = Stroke Time since diagnosis *(months)* = mixed (62% less than 6) *n* = 39 Age *(year)*, median (1st, 3rd quartiles) = acute grp 71 (51, 90) chronic grp 64 (45, 77) Sex, number male (%) = 20 (51) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity Responsiveness Interpretability

TABLE 1: (*Continued*)

Studies included	Summary of study participants	Psychometric property tested
Beebe and Lang (2009)	Diagnosis = Stroke Time since diagnosis *(months)* = less than 6 *n* = 33 Age *(year)*, mean (SD) = 53.9 (10.2) Sex, number male (%) = 19 (58) Sample included people with spasticity = yes	Internal consistency Reliability Measurement error Content validity Structural validity ✔ Hypothesis testing Cross-cultural validity ✔ Responsiveness Interpretability
Blennerhassett *et al*. (2010)	Diagnosis = Stroke Time since diagnosis *(months)* = less than 6 *n* = 22 Age (*year*), median (IQR), [range] = 63 (50–69), [23–80] Sex, number male (%) = 17 (77) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✔ Hypothesis testing Cross-cultural validity ✔ Responsiveness Interpretability
Burridge, Turk, Notley, Pickering and Simpson (2009)	Diagnosis = Stroke Time since diagnosis *(months)* = greater than 6 *n* = 17 Age *(year)*, mean (SD) = 57 (13.4) Sex, number male (%) = 11 (65) Sample included people with spasticity = yes	Internal consistency Reliability Measurement error Content validity Structural validity ✔ Hypothesis testing Cross-cultural validity Responsiveness Interpretability
Celik *et al*. (2010)	Diagnosis = Stroke Time since diagnosis *(months)* = greater than 6 *n* = 9 Age *(year)*, range = 48–67 Sex, number male (%) = 7 (78) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✔ Hypothesis testing Cross-cultural validity Responsiveness ✔ Interpretability
De Weerdt and Harrison (1985)	Diagnosis = Stroke Time since diagnosis *(months)* = less than 6 *n* = 53 Age *(year)*, mean (SD) = 68.6 (9.3) Sex, number male (%) = 25 (47) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✔ Hypothesis testing Cross-cultural validity ✔ Responsiveness Interpretability

TABLE 1: (*Continued*)

Studies included	Summary of study participants	Psychometric property tested
Dromerick *et al.* (2006)	Diagnosis = Stroke Time since diagnosis *(months)* = less than 6 *n* = 39 Age *(year)*, mean (SD) = 64.54 (14.13) Sex, number male (%) = 17 (44) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity Responsiveness ✓ Interpretability
Edwards *et al.* (2012)	Diagnosis = Stroke Time since diagnosis *(months)* = less than 6 *n* = 40 Age *(year)*, mean (SD) = 63.7 (13.6) Sex, number male (%) = 21 (42) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity ✓ Responsiveness ✓ Interpretability
Fleming *et al.* (2014)	Diagnosis = Stroke Time since diagnosis *(months)* = greater than 6 *n* = 33 Age *(year)*, mean (SD) = 61.5 (14.2) Sex, number male (%) = 20 (61) Sample included people with spasticity = yes	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity Responsiveness ✓ Interpretability
Lang *et al.* (2008)	Diagnosis = Stroke Time since diagnosis *(months)* = less than 6 *n* = 12 Age *(year)*, mean (SD) = 64 (14) Sex, number male (%) = 21 (40) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity Hypothesis testing Cross-cultural validity Responsiveness ✓ Interpretability
Lang *et al.* (2006)	Diagnosis = Stroke Time since diagnosis *(months)* = less than 6 *n* = 50 Age *(year)*, mean (SD) = 63.7 (13.6) Sex, number male (%) = 21 (42) Sample included people with spasticity = yes	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity ✓ Responsiveness Interpretability

TABLE 1: (*Continued*)

Studies included	Summary of study participants	Psychometric property tested
Lyle (1981)	Diagnosis = mixed (Stroke *n* = unknown, TBI *n* = unknown) Time since diagnosis *(months)* = Mixed (mean greater than 6) *n* = 20 Age *(year)*, mean (range) = 53.2 (26–72) Sex, number male (%) = 13 (65) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity ✓ Structural validity Hypothesis testing Cross-cultural validity Responsiveness Interpretability
McDonnell *et al.* (2006)	Diagnosis = Stroke Time since diagnosis *(months)* = less than 6 *n* = 17 Age *(year)*, range = 45–94 Sex, number male (%) = 9 (53) Sample included people with spasticity = not reported Rater characteristics Nil provided	Internal consistency ✓ Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity Responsiveness Interpretability
Morris, van Wijck, Joice and Donaghy (2013)	Diagnosis = Stroke Time since diagnosis *(months)* = greater than 6 *n* = 85 Age *(year)*, median (range) = 69 (36–88) Sex, number male (%) = 49 (58) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity Responsiveness Interpretability
Notley, Turk, Pickering, Simpson and Burridge (2007)	Diagnosis = Stroke Time since diagnosis *(months)* = greater than 6 *n* = 10 Age *(year)*, mean (SD) = 63 (13.8) Sex, number male (%) = 6 (60) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity Responsiveness Interpretability
O'Dell *et al.* (2013)	Diagnosis = Stroke Time since diagnosis *(months)* = greater than 6 *n* = 32 Age *(year)*, mean (SD) = 56 (12.4), Sex, number male (%) = 23 (72) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity ✓ Responsiveness Interpretability

TABLE 1: (*Continued*)

Studies included	Summary of study participants	Psychometric property tested
Page *et al.* (2015)	Diagnosis = Stroke	Internal consistency
	Time since diagnosis *(months)* = greater than 6	✓ Reliability
	n = 32	Measurement error
	Age *(year)*, mean (SD) = 56.6 (10.1)	Content validity
	Sex, number male (%) = 15 (47)	Structural validity
	Sample included people with spasticity = not reported	✓ Hypothesis testing
	Rater characteristics	Cross-cultural validity
	Rater n = 1 Clinical experience *(year)* = 8	Responsiveness
	Observations *n* = 64	Interpretability
Page *et al.* (2012)	Diagnosis = Stroke	Internal consistency
	Time since diagnosis *(months)* = greater than 6	✓ Reliability
	n = 29	✓ Measurement error
	Age *(year)*, mean (SD) = 60.8 (12.3)	Content validity
	Sex, number male (%) = 23 (79)	Structural validity
	Sample included people with spasticity = not reported	✓ Hypothesis testing
	Rater characteristics	Cross-cultural validity
	Rater *n* = 1 Clinical experience *(year)* = 8	Responsiveness
	Observations *n* = 58	Interpretability
Rabadi and Rabadi (2006)	Diagnosis = Stroke	Internal consistency
	Time since diagnosis *(months)* = less than 6	Reliability
	n = 104	Measurement error
	Age *(year)*, mean (SD) = 72.0 (13)	Content validity
	Sex, number male (%) = 43 (41)	Structural validity
	Sample included people with spasticity = not reported	✓ Hypothesis testing
		Cross cultural validity
		✓ Responsiveness
		Interpretability
Rand and Eng (2015)	Diagnosis = Stroke	Internal consistency
	Time since diagnosis *(months)* = less than 6	Reliability
	n = 32	Measurement error
	Age *(year)*, mean (SD) = 58.1 (12.4)	Content validity
	Sex, number male (%) = 25 (78)	Structural validity
	Sample included people with spasticity = not reported	✓ Hypothesis testing
		Cross-cultural validity
		Responsiveness
		Interpretability
Stinear, Barber, Petoe, Anwar and Byblow (2012)	Diagnosis = Stroke	Internal consistency
	Time since diagnosis *(months)* = less than 6	Reliability
	n = 40	Measurement error
	Age *(year)*, median (range) = 70 (31–91)	Content validity
	Sex, number male (%) = 16 (40)	Structural validity
	Sample included people with spasticity = not reported	✓ Hypothesis testing
		Cross-cultural validity
		Responsiveness
		Interpretability

TABLE 1: (*Continued*)

Studies included	Summary of study participants	Psychometric property tested
Urbin, Waddell and Lang (2015)	Diagnosis = Stroke Time since diagnosis *(months)* = Mixed (77% greater than 6) *n* = 35 Age *(year)*, mean (SD) = 56 (10.4), 62 (9.4) Sex number male (%) = 6 (75), 20 (74) Sample included people with spasticity = not reported	Internal consistency Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity Responsiveness Interpretability
Yozbatiran *et al.* (2008)	Diagnosis = Stroke Time since diagnosis *(months)* = greater than 6 *n* = 12 (validity) *n* = 9 (inter-rater reliability) *n* = 8 (intra-rater) Age *(year)*, mean (SD) = 61.0 (15.0) Sex, number male (%) = 6 (50) Sample included people with spasticity = not reported Rater characteristics Rater *n* = 2 Clinical experience *(year)* = 8 Observations *n* = 58	Internal consistency ✓ Reliability Measurement error Content validity Structural validity ✓ Hypothesis testing Cross-cultural validity Responsiveness Interpretability

a lack of clearly specified hypotheses, application of less than optimal statistical approaches, including effect sizes, and uncertainty as to what occurred in the interim period between measurements. Three of the nine studies included participants with upper limb spasticity (Barden, Baguley, Nott & Chapparo, 2014; Beebe & Lang, 2009; Lang *et al*., 2006). Synthesis of best evidence found that, for all nine studies, there was a *positive moderate level of evidence for responsiveness*.

Studies including participants with no identified limb spasticity used a range of statistical approaches to evaluate responsiveness and found the original ARAT to be responsive to change over time in acute through to chronic stroke and in chronic TBI. In comparison to 'like' measures of upper limb activity performance, the original ARAT performed well but was less responsive than the Chedoke Arm and Hand Activity Inventory (Barreca, Stratford, Lambert, Masters & Streiner, 2005) when used with people within the first 6 months post-stroke. When used within a sample of people with upper limb spasticity, which included participants post-stroke and TBI, the synthesised evidence rating was reduced to 'limited' as a result of only three studies contributing to the evidence. All studies were of fair methodological quality and only one (Lang *et al*., 2006) had a positive Terwee criteria due to a responsiveness ratio (RR) >1.96. The original ARAT was found to have moderate to high effect sizes ranging from 0.55 to 0.78 across studies (Barden *et al*., 2014; Beebe & Lang, 2009).

Measurement error

A single study which did not include participants with upper limb spasticity was located (Page *et al*., 2012). This reported a smallest detectable change (SDC) of 22.54 – the smallest amount of change attributed to true change and not random measurement error. COSMIN notes the close relationship and influence of measurement error and reliability, between and within raters and over time, on estimated SDCs. This study's methodological quality rating was 'fair' due to a lack of clear description of test conditions and independent administration; the assumption was that this was similar and independent. The criteria for good measurement properties thus received an 'indeterminant' rating because the MIC was not defined and no convincing explanation was given that agreement was acceptable, particularly given the SDC was 40% of the total score of 57. This evidence synthesis concluded that the original ARAT has a *conflicting level of evidence for measurement error*, further reduced due to the small sample size.

Structural validity

Lyle's, 1981, study detailing the development of the ARAT was the sole paper which evaluated structural validity. Lyle reported that the ARAT met Guttman Scaling criteria within each subscale achieving coefficients of scalability greater than 0.6 and coefficients of reproducibility greater than 0.9. This study had an

TABLE 2: *Summary of study results*

Study author, year	Internal consistency	Inter-rater reliability	Intra-rater reliability	Retest reliability	Measurement error	Content Validity	Structural validity	Hypothesis Testing (Construct validity)	Responsiveness	Interpretability	Floor and ceiling effect
Barden *et al.* (2014)								r = 0.50–0.63 DCD	ES = 0.78		100% floor
Barden *et al.* (2012)								Predictive of grasp/release P = −0.43 to 0.73	ROC curve = 0.88 (95% CI 0.76–1.00)		Floor 26% Ceiling 5%
Barreca *et al.* (2005)								CMSA r = 0.81–0.93	ROC curve 0.88 (95% CI 0.76–1.00)		
Barreca, Stratford, Masters, Lambert & Griffiths (2006)								CAHAI −9,13 r = 0.93–0.95	ROC curve 0.72		
Barreca, Stratford, Masters, Lambert, Griffiths & McBay (2006)								CMSA, CAHAI-7, 8, 9 week 0 r = 0.87–0.95 (one sided 95% CI 0.68–0.91) Week 2–6 r = 0.92–0.94 (one sided 95% CI 0.81–0.90)			
Beebe and Lang (2009)								1, 3, 6 months r_s = 0.61–0.95, grip and dexterity	ES = 0.55,0.63		

TABLE 2: (*Continued*)

Study author, year	Internal consistency	Inter-rater reliability	Intra-rater reliability	Retest reliability	Measurement error	Content Validity	Structural validity	Hypothesis Testing (Construct validity)	Responsiveness	Interpretability	Floor and ceiling effect
Blennerhassett *et al.* (2010)								HFS baseline $r = 0.96$ (95% CI 0.90–0.98) 4–6 weeks $r = 0.95$ (95% CI 0.87–0.98)	Rho_c = 062, 95% CI 0.35–0.90, Kw = 0.65		
Burridge *et al.* (2009)								Negative UMN $r = 0.025$–0.710, Positive UMN $r = -0.008$ to 0.231			
Celik *et al.* (2010)								Robotics $r = -0.83$ to 0.51			Floor 0% Ceiling 22%
De Weerdt and Harrison (1985)								B-FM 2 weeks $r = 0.91$, 8 weeks $r = 0.94$	$T = 1$, $n = 28$, $z = 4.60$		
Dromerick *et al.* (2006)								FIM $r = 0.47$, MAL QOM $r = 0.61$, WMFT time, functional ability $r = 0.65$, 0.95.			Ceiling 41%
Edwards *et al.* (2012)									ES = 1.018, 1.390		Floor 2% to 5.9% Ceiling 3.9% to 33%

TABLE 2: (*Continued*)

Study author, year	Internal consistency	Inter-rater reliability	Intra-rater reliability	Retest reliability	Measurement error	Content Validity	Structural validity	Hypothesis Testing (Construct validity)	Responsiveness	Interpretability	Floor and ceiling effect
Fleming *et al.* (2014)								54 = MAL 2.5 Predictive validity $R_2 = 0.6$; F1, 17 = 25.518; $P < 0.001$.			Floor 0% Ceiling 0%
Lang *et al.* (2008)										MCID 12, 17 (dominant, non-dominant)	
Lang *et al.* (2006)								Spasticity $r = -0.28$ to -0.49, disability $r = 0.2$–0.6	RR = 5.200, 7.067 ES = 1.018		
Lyle (1981)							Guttman Scaling criteria met				
McDonnell *et al.* (2006)				ICC $_{(3,1)}$ 0.93 ± 0.05 range 0.83–0.90				FMA $r = 0.75$, grip strength $r = 0.73$, tapping speed $r = 0.61$			

TABLE 2: (*Continued*)

Study author, year	Internal consistency	Inter-rater reliability	Intra-rater reliability	Retest reliability	Measurement error	Content Validity	Structural validity	Hypothesis Testing (Construct validity)	Responsiveness	Interpretability	Floor and ceiling effect
Morris *et al.* (2013)								RMA $r = -0.80$, NHP $r = -0.25$. Greater UL dysfunction associated with poorer HRQOL. Not predictive of overall HRQoL.			
Notley *et al.* (2007)								Tracking accuracy $r = -0.441$ to 0.829			
O'Dell *et al.* (2013)								AMAT-9 $r = 0.79$, SIS hand $r = 0.40$	SRM 0.89		
Page *et al.* (2015)			ICC 0.99 (95% CI 0.98, 0.99					W/h UE FM $r = 0.67$–0.74 ($P < 0.001$)			
Page *et al.* (2012)			ICC 0.71 (0.53–0.89, 95% CI)		SDC 22.54			W/h UE FM $r_s = 0.72$			
Rabadi and Rabadi (2006)								FMA $r = 0.77$ $P < 0.001$–0.87, $P < 0.001$	SRM 0.68		

TABLE 2: (*Continued*)

Study author, year	Internal consistency	Inter-rater reliability	Intra-rater reliability	Retest reliability	Measurement error	Content Validity	Structural validity	Hypothesis Testing (Construct validity)	Responsiveness	Interpretability	Floor and ceiling effect
Rand and Eng (2015)								Discharge ARAT and 12 months MAL $r = 0.78$, $P < 0.001$, accelerometer $r = 0.58$, $P < 0.001$, predicting 12 months UL function $R_2 = 0.776$, $P < 0.001$ (MAL), 0.470, $P = 0.001$ (accelerometer)			
Stinear *et al.* (2012)								PREP algorithm predicted 12 weeks ARAT			
Urbin *et al.* (2015)								Sensor acceleration $r = 0.73$–0.85, $P < 0.001$			
Yozbatiran *et al.* (2008)		$P = 0.96$, ICC 0.9986	$P = 0.99$ ICC 0.99					Arm Fugyl-Meyer $r = 0.94$ ($P < 0.01$)			

TABLE 2: (*Continued*)

Study author, year	Internal consistency	Inter-rater reliability	Intra-rater reliability	Retest reliability	Measurement error	Content Validity	Structural validity	Hypothesis Testing (Construct validity)	Responsiveness	Interpretability	Floor and ceiling effect
Summary of results	N/A	ICC 0.9986	ICC 0.71 0–0.99	$ICC_{(3,1)}$ 0.93 ± 0.05	MDC 22.54	N/A	Guttman Scaling criteria met	Measures UL function as intended. r = 0.25–0.95, not predictive HRQoL	ROC curve = 0.72–0.88 ES = 0.52–1.390 RR = 5.20–7.067 SRM = 0.68–0.89 Rho_c = 0.62 (95% CI 0.35–0.90) Kw = 0.65 Z = 4.60–4.85	MCID 12, 17	Floor effect = 0–100% Ceiling effect 0–41%

excellent level of methodological quality and met requirements for a positive rating for good measurement properties. The small sample size ($n < 25$), however, resulted in an *'unknown' level for structural validity*.

Hypothesis Testing (Construct validity)

A total of 26 studies evaluated the psychometric property of construct validity. Only four of these had a primary study purpose of evaluating psychometric properties; 16 included the original ARAT to validate new or modified outcome measurement tools and the final six studies evaluated predictors of the use and recovery of the upper limb with the original ARAT as the measure of activity limitation. Methodological quality ranged from poor to excellent within individual studies. A lack of clearly stated hypotheses reduced ratings for the majority of studies. Less commonly, ratings were reduced due to limited details regarding blinding of assessors and handling of missing data.

The original ARAT was found to have a high correlation with other like-tests, including the Brunnstrom-Fugl Meyer test (Fugl-Meyer, Jaakso, Leyman, Olsson & Steglind, 1975), the WMFT (Wolf, Lecraw, Barton & Jann, 1989), all versions of the Chedoke Arm & Hand Activity Inventory (Gowland *et al.*, 1993) and the Arm Motor Ability Test-9 (McCulloch, Cook, Fleming, Novack & Taub, 1988). Activity limitation as measured by the original ARAT was a predictor of grasp and release but not of overall quality of life.

The synthesis of evidence found the original ARAT to have s*trong positive evidence to support construct validity*. The synthesis of evidence considering participants with upper limb spasticity only, reduced in strength to a level of limited evidence with only six studies contributing. There is a *moderate level of positive evidence to support construct validity* in a sample of participants without upper limb spasticity with a total of 18 papers contributing to the final synthesis.

Interpretability

The original ARAT had highly variable results regarding floor and ceiling effects across the six studies that discussed this property. The variation overall ranged from 0 to 100% of participants for floor effects and 0–41% ceiling effects. COSMIN does not assign a rating for floor and ceiling effects – however, Terwee's criteria for good measurement properties assigns a rating. Positive ratings are assigned to studies if ≤15% of respondents achieved the highest or lowest possible score and a negative for ≥15% despite adequate design and methods. This is now applied in the evidence synthesis.

Dromerick *et al.* (2006) reported up to 41% of participants with a moderate degree of upper limb motor dysfunction with no identified spasticity scored the maximum 90 days post-stroke in comparison to 36% for the WMFT (Wolf *et al.*, 1989) functional ability scale whilst still recording limitations on the FIM (Granger, Hamilton, Linacre, Heinemann & Wright, 1993) and MAL (Uswatte & Taub, 1999). None of the participants in the study by Fleming, Newham, Roberts-Lewis and Sorinola (2014) who were greater than 6 months post-stroke and had identified upper limb spasticity were scored maximum or minimum in the original ARAT. Fleming's study found, however, that a score of 54 (out of a possible 57) was required to score a 2.5 on the MAL. Conversely, a floor effect was noted in Barden, Nott, Heard, Chapparo and Baguley's (2012) study including patients with TBI and upper limb spasticity; and O'Dell *et al.* (2013) study sample which included people with severe functional limitations and no upper limb spasticity. Edwards *et al.* (2012) found no floor effects within the first 90 days post-stroke. These findings indicate that it is more the level of upper limb activity limitation which impacts on the 1981-ARAT's ability to detect change rather than the time post-diagnosis. The synthesis revealed *conflicting evidence regarding interpretability* for all studies and for studies when stratifying by the presence of upper limb spasticity,

Interpretability can also be affected by the MIC. MIC was reported in a single study in an acute setting by Lang, Edwards, Birkenmeier and Dromerick (2008). MIC was 12 for the dominant affected limb and 17 for the non-dominant limb. Further studies are required, but evidence for *MIC was synthesised to be positive*.

Discussion

This systematic review considered methodological quality of the original ARAT when used within neurorehabilitation. The original ARAT is a tool commonly used in clinical practice and trials to measure the ability to perform activities with the upper limb. The original ARAT is one of the earliest measures of upper limb performance (Page *et al.*, 2012), yet despite such history, this review located very few studies that specifically investigated performance of this tool. This is surprising because it has been used as a presumably psychometrically acceptable comparator to validate or evaluate other existing, new or modified tools.

Study results demonstrate clinicians and researchers alike can be confident that when using the original ARAT with people post-stroke and with TBI without the presence of upper limb spasticity that a *moderate to strong level of evidence supports the intra-rater reliability, construct validity and responsiveness* of this tool. This review identified *several areas where only limited or inconclusive evidence exists*, specifically: inter and test–retest reliability (do different raters administer and score differently? Is there an influence on performance of the test items on repeat session over time?); measurement error (what is true change in performance and not the

TABLE 3: *Synthesis of best evidence and criteria*

	Internal consistency	Reliability (Inter)	Reliability (intra)	Reliability (retest)	Measurement Error	Content Validity	Structural validity	Construct validity	Cross cultural validity	Responsiveness	Interpretability	Floor or ceiling effect
Whole sample	0	?	++	?	+/−	0	?	+++	0	++	+	+/−
UL spasticity present	0	0	0	0	0	0	0	+	0	+	0	+/−
UL spasticity not present	0	?	++	?	+/−	0	?	++	0	++	+	+/−

Level	Rating	Criteria
Strong	+++ or – total sample size ≥100)	Consistent findings in multiple studies of good methodological quality or one study of excellent methodological quality
Moderate	++ or – (total sample size 50–99)	Consistent findings in multiple studies of fair methodological quality or in one study of good methodological quality
Limited	+ or – (total sample size 25–49)	One study of fair quality
Conflicting	±	Conflicting findings
Unknown	? (total sample <25)	Only studies of poor methodological quality

result of systematic or random error?); content validity; structural validity; and floor and ceiling effects (does the type of patient or the timing of use of the ARAT matter?).

This systematic review demonstrates differences within the psychometric properties of the original ARAT when the presence of upper limb spasticity is apparent in study samples. The level of evidence is significantly reduced and/or missing across various properties with only *limited positive evidence identified for construct validity and responsiveness*. An *inconclusive level of evidence was identified for floor and ceiling effects* and nil other evidence for remaining properties located.

This review highlights the importance of study or practice purpose and the clinical or research context in determining whether or not selection of the original ARAT is appropriate. This is because ratings of best evidence synthesis vary across participants and contexts. Practically, conflicting evidence drawn from a single study for the SDC (22.54) casts doubt on assumptions about utility of this measurement tool.

The paucity of studies contributing to best evidence synthesis indicates that more research is needed to provide the evidence needed for sensible interpretation of results. Interpretation decisions should also reflect change deemed important by patients. The MIC has been reported to be 10% of the total score or approximately a change of 6 in the subacute and chronic phase (van der Lee *et al*. 1999). In contrast, this review identified one acute setting study where the MIC was found to be 12 for the dominant affected limb and 17 for the non-dominant limb. A higher MIC in the acute setting has been proposed by Lang *et al*. (2008) to be due to the large portion of recovery that can occur during this period and strong expectations for continued recovery. Recovery expectations exemplify differences in MIC and SDC, the influence of both on the utility of the ARAT and need for clinicians to be aware of this measurement property in interpretation of results.

Studies frequently did not report detailed information regarding the methodology, the manner in which missing items were handled, a lack of clearly defined hypotheses and less than optimal statistical analyses particularly for construct validity and responsiveness. This reduced quality ratings. It is possible that if these had been reported, stronger ratings could have been made. As highlighted by Kennedy *et al*. (2013, p. 2523), COSMIN does not differentiate between poor methodological quality and poor reporting; thus, lower ratings may be a reflection of either poor methodological quality or underreporting of study characteristics.

Study limitations

Studies were excluded from this review because they did not meet inclusion criteria. The exclusion of studies where the original ARAT was not administered in English or the report was not in English is one example. This exclusion was made because of a lack of cross-cultural or translation validation studies. Thus studies completed in the Netherlands, Taiwan and China were excluded which would otherwise have been eligible. Another exclusion that led to a narrowing of the evidence base was the decision to use only those studies where the original ARAT had been administered in full with no modification. This exclusion was made because of the variability in potential modifications and the lack of transparency and consistency that modifications introduce. These exclusions together meant a large body of evidence were sacrificed to attain study results from relatively homogenous studies.

The review method itself holds limitations. Despite COSMIN criteria being quite explicit, implementation of the appraisal tool requires a level of individualisation for each application. COSMIN replicability in systematic reviews is an inherent and unavoidable limitation.

Conclusions

The ARAT is a frequently used outcome measure in clinical neurorehabilitation practice and research. In terms of suitable clinical populations, this review of the original 1981 version provides evidence that it is appropriate to use with people post-stroke and there may be potential for use within TBI populations (although the small sample size means further work is required in a TBI population). In terms of psychometric properties, the original ARAT has been shown to measure what it seeks to measure, that is, upper limb activity limitation. It is able to detect change in upper limb activity performance. Limitations of the instrument relate to SDC and MDC impacts on interpretation of results, particularly since psychometric properties differ across different populations and time points. The presence of upper limb spasticity significantly reduces recommendations regarding routine use in clinical practice due to limited evidence found only for construct validity and responsiveness. There is a need for further work to apply the ARAT to neurorehabilitation post-stroke populations and other populations with upper limb hemiparesis and spasticity with more rigorous research methodology and meticulous reporting to build evidence about use of the original version of the ARAT in neurorehabilitation.

Key points for occupational therapy

- In neurorehabilitation, the 1981 original ARAT is a valid measure of activity limitation and is responsive to change in activity performance.
- More evidence is needed to understand the smallest detectable change (SDC) and minimal important change (MIC) at different recovery time points and

in different neurorehabilitation populations including those people with upper limb spasticity.
- Gaps in reporting and methodological weaknesses limited the scope of evidence available for this review; future research using the ARAT in neurorehabilitation must use rigorous designs and meticulous reporting.

Author contributions

Shannon Pike co-led the design of the review and led the search and appraisal of data and writing of the manuscript, Natasha Lannin co-led the design of the review, assisted with data appraisal and provided critical revision of the manuscript, Anne Cusick contributed to the design of the review, assisted with data appraisal and writing of the manuscript. Kylie Wales provided assistance with the data appraisal and critical revision of the manuscript. All authors read and approved the final manuscript.

Funding

This review is not funded.

References

Barden, H. L., Nott, M. T., Heard, R., Chapparo, C. & Baguley, I. J. (2012). Clinical assessment of hand motor performance after acquired brain injury with dynamic computerised hand dynamometry: Construct, concurrent and predictive validity. *Archives of Physical Medicine and Rehabilitation*, *93*, 2257–2263. https://doi.org/10.1016/j.apmr.2012.06.028.

Barden, H. L., Baguley, I. J., Nott, M. T. & Chapparo, C. (2014). Dynamic computerised hand dynamometry: Measuring outcomes following upper limb botulinum toxin-A injections in adults with acquired brain injury. *Journal of Rehabilitation Medicine*, *46*, 314–320. https://doi.org/10.2340/16501977-1781.

Barreca, S. R., Stratford, P. W., Lambert, C. L., Masters, L. M. & Streiner, D. L. (2005). Test-retest reliability, validity and sensitivity of the Chedoke Arm and Hand Activity Inventory: A new measure of upper-limb function for survivors of stroke. *Archives of Physical Medicine and Rehabilitation*, *86*, 1616–1622. https://doi.org/10.1016/j.apmr.2005.03.017.

Barreca, S. R., Stratford, P. W., Masters, L. M., Lambert, C. L. & Griffiths, J. (2006). Comparing 2 versions of the Chedoke Arm and Hand Activity Inventory with the Action Research Arm Test. *Physical Therapy*, *86*, 245–253.

Barreca, S. R., Stratford, P. W., Masters, L. M., Lambert, C. L., Griffiths, J. & McBay, C. (2006). Validation of three shortened versions of the Chedoke Arm and Hand Activity Inventory. *Physiotherapy Canada*, *58*, 148–156.

Beebe, J. A. & Lang, C. E. (2009). Relationship and responsiveness of six upper extremity function tests during the first six months of recovery after stroke. *Journal of Neurological Physical Therapy*, *33*, 96–103. https://doi.org/10.1097/NPT.0b013e3181a33638.

Blennerhassett, J. M., Avery, R. M. & Carey, L. M. (2010). The test-retest and responsiveness to change of the Hand Function Survey during stroke rehabilitation. *Australian Occupational Therapy Journal*, *57*, 431–438. https://doi.org/10.1111/j.1440-1630.2010.00884.x.

Burridge, J. H., Turk, R., Notley, S. V., Pickering, R. M. & Simpson, D. M. (2009). The relationship between upper limb activity and impairment in post-stroke hemiplegia. *Disability and Rehabilitation*, *31*, 109–117.

Carroll, D. (1965). A quantitative test of upper extremity function. *Journal of Chronic Disability*, *18*, 479–491.

Celik, O., O'Malley, M. K., Boake, C., Levin, H. S., Yozbatiran, N. & Reistetter, T. A. (2010). Normalized movement quality measures for therapeutic robots strongly correlate with clinical motor impairment measures. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *18*, 433–444.

Chen, H. F., Lin, K. C., Wu, C. Y. & Chen, C. L. (2012). Rasch validation and predictive validity of the action research arm test in patients receiving stroke rehabilitation. *Archives of Physical Medicine and Rehabilitation*, *93*, 1039–1045. https://doi.org/10.1016/j.apmr.2011.11.033.

De Weerdt, W. J. G. & Harrison, M. A. (1985). Measuring recovery of arm-hand function in stroke patients: A comparison of the Brunnstrom-Fugl-Meyer test and the Action Research Arm test. *Physiotherapy Canada*, *37*, 65–70.

Dobson, F., Hinman, R. S., Hall, M., Terwee, C. B., Roos, E. M. & Bennell, K. L. (2012). Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: A systematic review. *Osteoarthritis Cartilage*, *20*, 1548–1562. https://doi.org/10.1016/j.joca.2012.08.015.

Dromerick, A. W., Lang, C. E., Birkenmeier, R., Hahn, M. G., Sahrmann, S. A. & Edwards, D. F. (2006). Relationships between upper-limb functional limitation and self-reported disability 3 months after stroke. *Journal of Rehabilitation and Development*, *43*, 401–408. https://doi.org/10.1682/JRRD.2005.04.0075.

Edwards, D. F., Lang, C. E., Wagner, J. M., Birkenmeier, R. & Dromerick, A. W. (2012). An evaluation of the Wolf Motor Function Test in motor trials early after stroke. *Archives of Physical Medicine and Rehabilitation*, *93*, 660–668. https://doi.org/10.1016/j.apmr.2011.10.005.

Fleming, M. K., Newham, D. J., Roberts-Lewis, S. F. & Sorinola, I. O. (2014). Self-perceived utilization of the paretic arm in chronic stroke requires high upper limb functional mobility. *Archives of Physical Medicine and Rehabilitation*, *95*, 918–924. https://doi.org/10.1016/j.apmr.2014.01.009.

Fugl-Meyer, A. R., Jaakso, L., Leyman, I., Olsson, S. & Steglind, S. (1975). The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. *Scandinavian Journal of Rehabilitation Medicine*, *7*, 13–31.

Gowland, C., Stratford, P. W., Ward, M., Moreland, J., Torresin, W., Van Hullenaar, S. *et al.* (1993). Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment. *Stroke*, *24*, 58–63.

Granger, C. V., Hamilton, B. B., Linacre, J. M., Heinemann, A. W. & Wright, B. D. (1993). Performance profiles for the functional independence measure. *American Journal of Physical Rehabilitation*, *72*, 84–89.

Hsueh, I. P. & Hsieh, C. L. (2002). Responsiveness of two upper extremity function instruments for stroke in patients receiving rehabilitation. *Clinical Rehabilitation*, *16*, 617–624. https://doi.org/10.1191/0269215502cr530oa.

Hsueh, I. P., Lee, M. M. & Hsieh, C. L. (2002). The Action Research Arm Test: Is it necessary for patients being tested to sit at a standardised table? *Clinical Rehabilitation*, *16*, 382–388. https://doi.org/10.1191/0269215502cr509oa.

Kennedy, C. A., Beaton, D. E., Smith, P., Van Eerd, D., Tang, K., Inrig, T. *et al.* (2013). Measurement properties of the QuickDASH (Disabilities of the Arm, Shoulder and Hand) outcome measure

and cross-cultural adaptations of the QuickDASH: A systematic review. *Quality of Life Research*, *22*, 2509–2547. https://doi.org/10.1007/s11136-013-0362-4.

Koh, C.-L., Hsueh, I.-P., Wang, W.-C., Sheu, C.-F., Yu, T.-Y., Wang, C.-H. *et al.* (2006). Validation of the Action Research Arm Test using item response theory in patients after stroke. *Journal of Rehabilitation Medicine*, *38*, 375–380. https://doi.org/10.1080/16501970600803252.

Kwakkel, G., Lannin, N. A., Borschmann, K., English, C., Ali, M., Churilov, L. *et al.* (2017). Standardised measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the stroke recovery and rehabilitation roundtable. *International Journal of Stroke*, *12*, 451–461. https://doi.org/10.1177/1747493017711813.

Lang, C. E., Wagner, J. M., Dromerick, A. W. & Edwards, D. F. (2006). Measurement of upper-extremity function early after stroke: Properties of the Action Research Arm Test. *Archives of Physical Medicine and Rehabilitation*, *87*, 1605–1610. https://doi.org/10.1016/j.apmr.2006.09.003.

Lang, C. E., Edwards, D. F., Birkenmeier, R. L. & Dromerick, A. W. (2008). Estimating minimal clinically important differences of upper-extremity measures early after stroke. *Archives of Physical Medicine and Rehabilitation*, *89*, 1693–1700. https://doi.org/10.1016/j.apmr.2008.02.022.

Lin, J.-H., Hsu, M.-J., Sheu, C.-F., Wu, T.-S., Lin, R.-T., Chen, C.-H. *et al.* (2009). Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. *Physical Therapy*, *89*, 840–850. https://doi.org/10.2522/ptj.20080285.

Lin, K.-C., Chuang, L.-L., Wu, C.-Y., Hsieh, Y.-W. & Chang, W.-Y. (2010). Responsiveness and validity of three dexterous function measures in stroke rehabilitation. *Journal of Rehabilitation Research & Development*, *47*, 563–572. https://doi.org/10.1682/JRRD.2009.09.0155.

Lyle, R. C. (1981). A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *International Journal of Rehabilitation Research*, *4*, 483–492.

McCulloch, K. L., Cook, E. W. I., Fleming, W. C., Novack, T. A. & Taub, E. (1988). A reliable test of upper extremity ADL function. *Archives of Physical Medicine and Rehabilitation*, *69*, 755.

McDonnell, M. N., Hillier, S. L., Ridding, M. C. & Miles, T. S. (2006). Impairments in precision grip correlate with functional measures in adult hemiplegia. *Clinical Neurophysiology*, *117*, 1474–1480. https://doi.org/10.1016/j.clinph.2006.02.027.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L. *et al.* (2010). International consensus on taxonomy, terminology, and definitions of measurement properties: Results of the COSMIN study. *Journal of Clinical Epidemiology*, *63*, 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006.

Morris, J. H., van Wijck, F., Joice, S. & Donaghy, M. (2013). Predicting health related quality of life 6 months after stroke: The role of anxiety and upper limb dysfunction. *Disability and Rehabilitation*, *35*, 291–299. https://doi.org/10.3109/09638288.2012.691942.

Notley, S. V., Turk, R., Pickering, R. M., Simpson, D. M. & Burridge, J. H. (2007). Analysis of the quality of wrist movement during a simple tracking task. *Physiological Measurement*, *28*, 881–895. https://doi.org/10.1088/0967-3334/28/8/011.

O'Dell, M. W., Kim, G., Rivera, L., Fieo, R., Christos, P., Polistena, C. *et al.* (2013). A psychometric evaluation of the Arm Motor Mobility Test. *Journal of Rehabilitation Medicine*, *45*, 519–527. https://doi.org/10.2340/16501977-1138.

Page, S. J., Levine, P. & Hade, E. (2012). Psychometric properties and administration of the Wrist/Hand Subscales of the Fugl-Meyer assessment in minimally impaired upper extremity hemiparesis in stroke. *Archives of Physical Medicine and Rehabilitation*, *93*, 2373–2376. https://doi.org/10.1016/j.apmr.2012.06.017.

Page, S. J., Hade, E. & Persch, A. (2015). Psychometrics of the wrist stability and hand mobility subscales of the Fugl-Meyer assessment in moderately impaired stroke. *Physical Therapy*, *95*, 103–108. https://doi.org/10.2522/ptj.20130235.

Pike, S., Lannin, N. A., Cusick, A., Wales, K., Turner-Stokes, L. & Ashford, S. (2015). A systematic review protocol to evaluate the psychometric properties of measures of function within adult neuro-rehabilition. *Systematic Reviews*, *4*, 86. https://doi.org/10.1186/s13643-015-0076-5.

Platz, T., Pinkowski, C., van Wijck, F. & Johnson, G. (2005). *Arm – Arm Rehabilitation Measurement: Manual for Performance and Scoring of Fugl-Meyer test, Action Research Arm Test and Box and Block test*. Baden-Baden: Deutscher Wissenschafts-Verlag.

Platz, T., Pinkowski, C., van Wijck, F., Kim, I. H., di Bella, P. & Johnson, G. (2005). Reliability and validity of arm function assessment with standardised guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: A multicentre study. *Clinical Rehabilitation*, *19*, 404–411. https://doi.org/10.1191/0269215505cr832oa.

Rabadi, M. H. & Rabadi, F. M. (2006). Comparison of the Action Research Arm Test and the Fugl-Meyer Assessment as measures of upper-extremity motor weakness after stroke. *Archives of Physical Medicine and Rehabilitation*, *87*, 962–966. https://doi.org/10.1016/j.apmr.2006.02.036.

Rand, D. & Eng, J. J. (2015). Predicting daily use of the affected upper extremity 1 year after stroke. *Journal of Stroke and Cerebrovascular Diseases*, *24*, 274–283.

Santisteban, L., Teremetz, M., Bleton, J. P., Baron, J. C., Maier, M. A. & Lindberg, P. G. (2016). Upper limb outcome measures used in stroke rehabilitation studies: A systematic literature review. *PLoS ONE*, *11*, e0154792. https://doi.org/10.1371/journal.pone.0154792.

Sheean, G., Lannin, N. A., Turner-Stokes, L., Rawicki, B. & Snow, B. J. (2010). Botulinum toxin assessment, intervention and aftercare for upper limb hypertonicity in adults: International consensus statement. *European Journal of Neurology*, *17* (Suppl 2), 74–93. https://doi.org/10.1111/j.1468-1331.2010.03129.x.

Stinear, C. M., Barber, P. A., Petoe, M., Anwar, S. & Byblow, W. D. (2012). The PREP algorithm predicts potential for upper limb recovery after stroke. *Brain*, *135*, 2527–2535. https://doi.org/10.1093/brain/aws146.

Sullivan, J. E., Crowner, B. E., Kluding, P. M., Nichols, D., Rose, D. K., Yoshida, R. *et al.* (2013). Outcome measures for individuals with stroke: Process and recommendations from the American physical therapy association neurology section task force. *Physical Therapy*, *93*, 1383–1396. https://doi.org/10.2522/ptj.20120492.

Terwee, C. B., Bot, S. D. M., de Boer, M., van der Windt, D. A. W. M., Knol, D. L., Dekker, J. *et al.* (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*, 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W. J. G., Bouter, L. M. & de Vet, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, *21*, 651–657. https://doi.org/10.1007/s11136-011-9960-1.

Terwee, C. B., Bot, S. D. M., de Boer, M., Van der Windt, D. A. W. M., Knol, D. L., Dekker, J. *et al.* (2017). Criteria for good measurement properties. Retrieved ?????, from http://www.cosmin.nl/Systematic%20reviews%20of%20measurement%20properties.html

Urbin, M. A., Waddell, K. J. & Lang, C. E. (2015). Acceleration metrics are responsive to change in upper extremity function of

stroke survivors. *Archives of Physical Medicine and Rehabilitation*, *96*, 854–61. https://doi.org/10.1016/j.apmr.2014.11.018.

Uswatte, G. & Taub, E. (1999). *Constraint-Induced movement therapy: New approaches to measurement in rehabilitation*. Cambridge: Cambridge University Press.

Van der Lee, J. H., De Groot, V., Beckerman, H., Wagenaar, R. C., Lankhorst, G. J. & Bouter, L. M. (2001). The intra- and interrater reliability of the Action Research Arm Test: A practical test of upper extremity function in patients with stroke. *Archives of Physical Medicine and Rehabilitation*, *82*, 14–19. https://doi.org/10.1053/apmr.2001.18668.

Van der Lee, J. H., Roorda, L. D., Beckerman, H., Lankhorst, G. J. & Bouter, L. M. (2002). Improving the Action Research Arm test: A unidimensional hierarchical scale. *Clinical Rehabilitation*, *16*, 646–653. https://doi.org/10.1191/0269215502cr534oa.

Van der Lee, J. H., Wagenaar, R. C., Lankhorst, G. J., Vogelaar, T. W., Deville, W. L. & Bouter, L. M. (1999). Forced use of the upper extremity in chronic stroke patients: Results from a single blind randomized clinical trial. *Stroke*, *30*, 2369–2375. https://doi.org/10.1161/str.30.11.2369.

Wagenaar, R. C., Meijer, O. G., van Wieringen, P. C., Kuik, D. J., Hazenberg, G. J., Lindeboom, J. *et al*. (1990). The functional recovery of stroke: A comparison between neuro-developmental treatment and the Brunnstrom method. *Scandinavian Journal of Rehabilitation Medicine*, *22*, 1–8.

Wales, K., Clemson, L., Lannin, N. A. & Cameron, I. (2016). Functional assessments used by occupational therapists with older adults at risk of activity and participation limitations: A systematic review. *PLoS ONE*, *11*, e0147980. https://doi.org/10.1371/journal.pone.0147980.

Wolf, S. L., Lecraw, D. E., Barton, L. A. & Jann, B. B. (1989). Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head injured patients. *Experimental Neurology*, *104*, 125–132.

Yozbatiran, N., Der-Yeghiaian, L. & Cramer, S. C. (2008). A standardized approach to performing the action research arm test. *Neurorehabilitation and Neural Repair*, *22*, 78–90. https://doi.org/10.1177/1545968307305353.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. COSMIN ratings.
Table S2. Quality of Measurement properties (Terwee analysis).