Dose effects in behavioural treatment of post-stroke aphasia: a systematic review and meta-analysis

Sam R. Harvey [a,b], Marcella Carragher [a,b], Michael Walsh Dickey[b,c,d], John E. Pierce[a,b], and Miranda L. Rose[a,b]

[a]School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, Australia; [b]Centre of Research Excellence in Aphasia Recovery and Rehabilitation, Australia; [c]Geriatric Research Education and Clinical Center, VA Pittsburgh Healthcare System, USA; [d]University of Pittsburgh, USA

Sam R. Harvey. Health Sciences 1, La Trobe University, Plenty Road, Bundoora 3083, Australia. ORCID identifier: 0000-0002-4839-2117 Twitter: @SRHarvey_ sam.harvey@latrobe.edu.au

Marcella Carragher. Health Sciences 1, La Trobe University, Plenty Road, Bundoora 3083, Australia. ORCID identifier: 0000-0002-7200-6968 Twitter: @MarcellaC_SP marcella.carragher@latrobe.edu.au

Michael Walsh Dickey. 6077 Forbes Tower, University of Pittsburgh, Pittsburgh PA 15260, USA. ORCID identifier: 0000-0002-9068-3313 mdickey@pitt.edu

John E. Pierce. 2-6 Hopetoun Street, Elsternwick 3185, Australia. ORCID identifier: 0000-0001-5164-5106 Twitter: @johnpierce85 pierce.john.e@gmail.com

Miranda L. Rose. Health Sciences 1, La Trobe University, Plenty Road, Bundoora 3083, Australia. ORCID identifier: 0000-0002-8892-0965 Twitter: @rose_mirandaros M.rose@latrobe.edu.au

Corresponding author:

Professor Miranda Rose, PhD
Discipline of Speech Pathology
School of Allied Health, Human Services and Sport
College of Science, Health and Engineering
La Trobe University
Bundoora 3086 AUSTRALIA
T: +61 3 9479 2088
E: m.rose@latrobe.edu.au

Conflicts of interest

The authors declare no potential conflict of interest.

Funding

This work was supported by an Australian Government Research Training Program Scholarship and the NHMRC funded Centre for Research Excellence in Aphasia Recovery and Rehabilitation (#1153236).

ABSTRACT

Purpose: Aphasia is a debilitating chronic acquired language disorder that impacts heavily on a person's life. Behavioural treatments aim to remediate language processing skills or to enhance communication between the person with aphasia and others, and a number of different treatments are efficacious. However, it is unclear how much of a particular treatment a person needs in order to optimise recovery of language and communication skills following stroke.

Materials and methods: Systematic search for and meta-analysis of experimental studies that directly compared different amounts of the same behavioural aphasia treatment, following PRISMA guidelines.

Results: Treatment dose research in aphasia is an emerging area. Just six studies comparing different doses of the same intervention met all criteria for inclusion. Evidence from these studies was synthesised and meta-analysed, where possible. Meta-analyses were inconclusive due to limited data; however, there are indications that suggest increased dose may confer greater improvement on language and communication measures, but with diminishing returns over time. Aphasia severity and chronicity may affect dose-response relationships.

Conclusions: There is currently insufficient evidence to determine the effect of dose on treatment response. A dedicated and coordinated research agenda is required to systematically explore dose-response relationships in post-stroke aphasia interventions.

Word count: 7,442 (including abstract, tables, figures, and citations, ex bibliography and supplementary material)

Keywords: Aphasia, stroke, treatment, dose, meta-analysis

INTRODUCTION

Aphasia is a common and significant acquired communication disability which affects up to 40% of stroke survivors [1] and persists as a chronic condition in up to 50% of cases [2-4]. Aphasia is associated with an increased risk of mortality [5], higher healthcare costs [6], negative consequences for personal relationships, vocational participation, and economic independence [3, 7], and poorer health-related quality of life than many other debilitating health conditions including Alzheimer's disease and cancer [8]. Aphasia treatments have been shown to improve language skills, social participation, and quality of life [9]; however, people with aphasia may not be receiving enough therapy to maximise recovery of language skills and communication following stroke (e.g., [10-15]) despite suggestions that higher doses of treatment may lead to better recovery [9, 10, 16, 17]. Finding the right dose of aphasia treatment is important for treatment prescription, refining research agendas, and will impact service delivery and health policy.

Dose conceptualisation and the dose/intensity confound

Treatment dose can be conceptualised in two ways: as the amount of time spent in therapy and as the number of therapeutic elements provided or received over an intervention period [18, 19]. In the absence of consensus definitions but informed by Baker [18], the following concepts and definitions will be referred to in this review:

Therapeutic element	The basic unit of therapy; either a therapeutic input or a client act.
Session dose	A quantitative measure of the therapeutic content provided in a session, in minutes or therapeutic elements.
Session intensity	The rate at which therapeutic elements are provided in a session, e.g., 300 naming attempts per hour.

Session frequency	The number of therapy sessions per week.
Total dose	Amount of therapy provided or received over an intervention period, in time or therapeutic inputs, e.g., total hours, total number of therapeutic elements.

The *total dose* of treatment is equal to the *session dose* x *session frequency* x *intervention duration* [18, 20]. Hours of therapy is a convenient measure; it is economical to capture, is easy to calculate and compare from one study to the next, has clinical relevance and is easily understood by consumers and policy makers, and is the most commonly reported measure of treatment dose in aphasia intervention studies [19]. Conversely, conceptualising, measuring, and reporting dose as a collection of therapeutic elements may allow more refined inspection of dose-response relationships for a given intervention [18, 21]. There are many potential therapeutic elements for any given intervention. These include inputs such as the presentation of therapy stimuli, clinician-delivered cues, clinician-generated responses, and feedback/reinforcement. Client acts may include accurate, inaccurate and self-corrected responses, and the use of self-cueing strategies. Therapeutic elements may contain the active ingredients of treatment which "teach or enhance new learning and behaviour" [17, p.71]. Closer examination of these active ingredients may ultimately enhance our understanding of the mechanisms of action that transform received treatment into improved health and wellbeing [22]. Once identified, maximising delivery of active ingredients has the potential to increase treatment efficiency and effectiveness. However, measuring dose in terms of therapeutic elements can be more difficult and labour-intensive to capture.

A number of reviews have examined the literature for evidence of dose-related treatment effects. In 2003, Bhogal and colleagues asserted that when it comes to the impact

of aphasia treatment on recovery of language and communication functions, "more is better" [23]. In their synthesis of findings, studies that demonstrated a statistically significant positive treatment effect provided a total of 98.4 hours of therapy or more, whereas ineffective studies provided a total of 43.6 hours of therapy or less [16]. Although based on few studies (n = 8), the assertion that "more is better" has heavily influenced the subsequent examination of dose-response in aphasia research.

The "more is better" finding was not conclusively supported by Brady and colleagues [9] who meta-analysed group-level outcome data from five randomised controlled trials (RCTs) in which participants received either a higher dose (range 27 – 208 hours) or lower dose (range 5 – 78 hours) of treatment. Brady and colleagues found that people with aphasia who received a higher dose of treatment had significantly better functional communication, although this finding was based on data from just one RCT [24]. However, there were no statistically significant findings regarding the effect of dose on measures of receptive or expressive language, or aphasia severity. Overall, the lower dose condition resulted in significantly fewer dropouts and better treatment adherence.

A number of limitations of these reviews necessitate the current review. First, examination of dose-response relationships may be confounded when simultaneously comparing different interventions. For example, two of the five studies included in Brady et al. [9] dose analyses compared different amounts of different treatments: the VERSE I trial [24] compared VERSE therapy to usual care, and Denes et al. [25] compared a conversational approach focused on auditory comprehension to standard speech and language therapy based on a stimulation approach. Differences between treatments may obscure the effect of dose on treatment outcomes and, therefore, the validity of making dose comparisons across interventions is questionable.

Second, the examination of dose effects in post-stroke aphasia is an emerging research area. As such, findings from small-scale Phase I exploratory experiments that have yet to be scaled up to large-scale Phase II group studies may contribute important knowledge to guide future examination of dose-response relationships. Small-scale experimental studies were not included in previous dose effect reviews.

Another limitation of these dose explorations is the conceptualisation of dose as 'the amount of time spent receiving therapy'. As has been previously argued [e.g., [26, 27], measuring the dose of aphasia intervention in hours is inadequate because of the inherent inaccurate assumption that all hours of treatment are equal. Clinically, one hour of treatment to the next may comprise a variety of different tasks targeting different goals requiring the provision of a different number and combination of therapeutic elements. In research, especially in large pragmatic trials, it is difficult to know how often different therapeutic elements are being provided unless treatment details are clearly reported and monitored. Therefore, measuring the dose of complex interventions only in hours makes it impossible to examine responses to specific therapeutic elements.

Despite this, measuring and reporting total hours remains the most common approach. A scoping review found that of 112 aphasia intervention studies reporting dose, 96% (n = 108) reported hours while only 27% (n = 30) reported therapeutic elements in sufficient detail that total dose could be calculated, the latter more frequently reported for naming treatment studies [19]. A recent review of naming treatment studies [28] found that time spent in treatment does not correlate with treatment outcomes and that the number of words treated in therapy correlates with the number of words learned (for a similar finding, [see [29]). A limitation of Thomas and colleagues' [28] high-quality review is that the authors chose to focus their exploration on the relationship between stimulus set size and treatment outcomes without exploring other parameters that might reveal dose-response relationships. This is

165 despite a number of recent naming treatment studies having systematically manipulated or
166 measured therapeutic elements to examine dose-response relationships [e.g., [30-32]; for a
167 review [see [19]). More careful attention to therapeutic elements is needed if the relationship
168 of dose to aphasia treatment outcomes is to be understood.

169 The final major limitation of all dose studies to date is that treatment dose is
170 consistently confounded with treatment *intensity* [33]. In the aphasia literature, intensity has
171 come to be synonymous with *frequency* and means the rate at which a particular dose is
172 provided: it is the quotient of dose over time. Dose and intensity are, therefore,
173 interdependent. For example, in studies comparing massed to distributed practice (dose-
174 controlled studies of intensity [e.g., [34]), the session dose is static while the session
175 frequency and intervention duration are manipulated relative to each other to produce the
176 same total dose (e.g., 1 hour per session, 4-5 sessions per week, 3 weeks = 14 hours
177 compared to 1 hour per session, 1-2 sessions per week for 8 weeks = 14 hours). This raises
178 the question: are observed differences in treatment effects attributable to different session
179 frequency or different intervention duration, or both? Likewise, in dose-effect studies, the
180 challenge with manipulating the total dose is that two of the three schedule parameters (i.e.,
181 session dose, session frequency, intervention duration) must change. This dose/intensity
182 relationship also exists at the session level; high session dose vs low session dose
183 comparisons also compare high session intensity to low session intensity if the session
184 duration is constant. Again, we are faced with the issue of determining which parameter, if
185 any, confers the treatment effect. It is possible, perhaps probable, that the overall impact on
186 outcome is a result of the interaction between a number of these variables [27].

187 In summary, there are signals emerging from the aphasia intervention literature
188 regarding dose-response relationships but the evidence has been scant, at times contradictory,
189 and overall, inconclusive. Traditional methods for measuring treatment dose may lack the

specificity required to adequately investigate these relationships. Furthermore, existing evidence comes from a small number of studies which employed different treatment paradigms with heterogenous samples, most of which did not directly compare different doses of the same intervention. Only a small number of studies have analysed the comparative effects of different doses of the same therapy within each study [19]. To our knowledge, there is no published meta-analysis of data from these dose-effect studies.

AIMS

The primary aim of this review is to examine the current evidence for dose effects in behavioural post-stroke aphasia interventions. We aim to answer the following questions by meta-analysing data from experimental studies that directly compare different amounts of the same intervention:

1) Does a larger dose of intervention result in better language and communication outcomes for people with aphasia following stoke?
2) Does time post stroke impact dose effects?
3) Are there specific person-level characteristics that help explain variability in dose-response relationships?
4) Is there evidence of dose effects in specific language or communication interventions?

METHODS

Search strategy

A comprehensive and systematic search was undertaken in June, 2019 for peer-reviewed randomized controlled studies, quasi-experimental studies, and single-case design studies, which reported the amount of behavioural aphasia therapy provided, and investigated the dose-response relationship of that intervention on language impairment and communication

activity/participation for adults with aphasia following stroke. The search was replicated and the yield updated in September, 2020.

Using the Preferred Reporting Items for Systematic reviews and Meta-Analysis Guidelines (PRISMA, [35]), the following databases were searched, with no language or date limits set: PubMed, Medline, EMBASE, CINAHL, PsycINFO, and Cochrane Library. Table 1 shows search terms for the key domains of post-stroke aphasia, intervention, and dose, identified from relevant literature. These search domains were combined using the AND operator, and the terms within each domain combined using OR. Search terms were modified in line with individual database subject headings. An example of the final search strategy is provided in Appendix A. Reference lists of included studies were examined to identify additional studies not captured during the systematic search.

< Table 1 Search terms relating to treatment dose in aphasia >

Selecting studies

Figure 1 shows a PRISMA flow diagram detailing the results of study identification, screening, eligibility, and inclusion. The search yield was imported into Rayyan online software [36] and duplicates removed using software and manual checking. Titles and abstracts were then screened by the first author as per the inclusion criteria to determine eligibility for full text review. Exclusion criteria are listed in figure 1. Twenty percent of full texts were re-screened by a second reviewer (JEP) for inclusion, achieving 95% agreement between reviewers. Inconsistencies were discussed and resolved, and inclusion criteria refined to improve accurate classification.

Inclusion criteria

- Full text peer-reviewed journal article in English
- Includes adults presenting with post-stroke aphasia, at any time after stroke

- Reports primary data from behavioural treatment targeting language impairment or communication activity/participation
- Measures and reports the amount of treatment provided
- Provides a comparative analysis of the effect of different amounts of the same intervention

Study categorisation and methodological quality appraisal

Studies were categorised and appraised by the first and fourth authors, reaching consensus through discussion where necessary. Included articles were categorised by study type using the Oxford Centre for Evidence Based Medicine levels of evidence [37]. Single-case methodologies are commonly used in aphasia research; however, the OCEBM fails to distinguish experimental from non-experimental single-case designs. Single-case experimental designs (i.e., multiple baseline, withdrawal/reversal, alternating treatments, and changing criterion designs) provide a method for understanding causal relationships in complex behavioural interventions, whereas non-experimental pre/post designs and case studies do not [38]. Therefore, the RoBiNT manual [38] was used to further classify single-case designs. Methodological rigour was assessed using the PEDro-P scale [39] for RCT and quasi-RCT, and the RoBiNT scale [38] for single-case designs. Pre/post case series were excluded from further analysis as these provide a low level of evidence due to a lack of experimental control [38].

Data extraction and analysis

Data were entered into a spreadsheet including participant characteristics, treatment type, outcome measures, therapy schedule, and results. Data extraction was completed by two

reviewers (SRH, JEP) and compared for accuracy. Where data were unavailable, study authors were contacted. Table 2 shows the data items extracted from each paper.

< Table 2 Data items extracted from selected studies >

Effect size calculation

While the initial intention was to meta-analyse effect sizes derived from mean change scores of primary outcomes in group studies, this was not possible due to insufficient comparable studies at the group study level. Where possible, group study effect sizes as reported by each study are reported below (*see Results*). For studies employing a single-case design, Tau-U was calculated based on individual patient naming accuracy data. Tau-U measures the degree of improvement across adjacent treatment phases by measuring the proportion of data points in the treatment phase that are above data in the baseline, adjusting for trend in the baseline phase [40] and is considered superior to other non-overlap measures when handling small data sets [41]. Raw data were manually extracted from single-case design case charts using online software (https://apps.automeris.io/wpd/). Tau-U was adjusted for baseline trend if Tau for the baseline phase exceeded 0.4 and a trend was apparent by visual inspection [40].

RESULTS

The literature search yielded 4,223 unique articles. Of those, 16 articles reporting on 15 studies met the inclusion criteria outlined for this review (figure 1).

< Figure 1 PRISMA flow diagram showing the study selection process >

Levels of evidence

The included studies comprise four RCTs [42-45], one quasi-RCT in which participants were sequentially allocated to cohorts that had been randomly assigned to

different treatment arms [46], and three quasi-experimental "AB with follow up" single-case designs [31, 32, 47]. Eight non-experimental pre/post case series were not appropriate for the meta-analysis given the low level of evidence of these designs [21, 29, 30, 48-52].

Methodological quality

Figure 2 and figure 3 show the quality ratings for controlled trials (PEDro-P) and single-case designs (RoBiNT). Only findings from studies considered moderate to high quality were considered in further analysis. Cut-off scores for moderate to high quality studies are 5 points and above for the PEDro-P scale [53]. Benchmark cut-off scores for single-case designs have yet to be established. However, in a paper examining the reliability of the RoBiNT scale [38] the mean score of included studies was 12 points. This score has been used in lieu of formalised benchmarks in a previous systematic review in aphasia [54] and was adopted for this review.

< Figure 2 PEDro-P scale scores for included group studies with cut-off score ≥5 >

< Figure 3 RoBiNT scale scores for included single case design studies with cut-off score ≥12 >

Study characteristics

Appendix B contains the study characteristics for the six studies that met all criteria for inclusion, level of evidence, and methodological rigour. Studies reported data on 323 participants (153 men, 170 women) with a mean reported age of 62 (SD 7). One study recruited participants across the acute-subacute phase and five in the chronic phase of recovery. All six studies investigated impairment-level lexical retrieval interventions with two RCTs [42, 43] including additional aspects of functional communication rehabilitation and educational counselling. Five studies investigated change in language impairment as the

primary outcome, as measured on standardised aphasia test batteries (i.e., WAB AQ, Aachen Aphasia Test) or by confrontation naming. One study measured changes in functional communication skills using the Amsterdam-Nijmegen Everyday Language Test (ANELT) as the primary outcome [43]. Four studies included secondary outcomes that measured some aspect of functional communication either using a standardised tool (e.g., ANELT) or measures of informativeness (e.g., content information units in a narrative task) in an attempt to quantify generalisation of treatment effects.

Treatment schedule and total dose

Table 3 and table 4 show the treatment schedule of each included study. The total number of hours of treatment provided ranged between six and 60 hours. The prescribed doses and actual amount of treatment received are noted for each group within these studies. Participants in the Bakheit study received less treatment than prescribed and those in the Breitenstein study generally received more. The three RCTs were primarily designed to investigate the effects of treatment *intensity* which was achieved by manipulating session frequency (i.e., weekly intensity) and intervention duration across groups. However, relevant to this systematic review, each also explored the effect of different doses received between groups or at different time points within groups. In addition to reporting treatment duration, two studies [31, 32] also reported total number of therapeutic elements provided (total number of naming attempts) which ranged from 1,200 to 3,200 across participants.

< Table 3 Treatment schedules for studies included in analysis which reported total dose in hours >

< Table 4 Treatment schedules for studies included in analysis which reported total dose in hours and therapeutic elements >

339 *Dose effects*

340 *Evidence from studies reporting dose in hours*

341 Two large, high quality pragmatic RCTs provide conflicting evidence of dose effects
342 in speech and language interventions. In the early phase of recovery, Bakheit and colleagues
343 [42] found no significant difference between participants randomly assigned to receive 60
344 hours (intensive group, n = 51) compared to 24 hours (conventional group, n = 48) of
345 individualised language intervention over 12 weeks. None of the participants in the intensive
346 group received the full dose of treatment. Subgroup analysis of 13 participants from the
347 intensive group who attended over 80% of prescribed sessions (receiving a mean dose of 51.6
348 hours compared to 19.2 hours in the conventional group) also failed to demonstrate between-
349 group difference in language performance (WAB AQ) at any timepoint following
350 intervention (raw data unavailable, correspondence with authors 26/2/2020).

351 In contrast, Breitenstein and colleagues [43] found significant effects following an
352 average of 31 hours of speech and language therapy (intervention group, n = 78) vs. 4.5 hours
353 (control group treatment deferral, n = 78) over three weeks in the chronic phase of recovery
354 (Cohen's $d = 0.58$, $p = 0.0004$). In addition, secondary within-group analysis of a subgroup of
355 participants (n = 39) who received an additional three-week block of therapy showed that the
356 mean change in ANELT A-scale score was roughly one point larger after a median of six
357 weeks of intensive therapy (IQR 5–7) than after the initial three weeks of intensive therapy
358 (mean ANELT A-scale at 3-week timepoint: 3.32 points [SD 5.64], 95% CI 1.35 - 5.29 vs. at
359 6-week timepoint: 4.23 points [4.28], 2.74 - 5.73). These results suggest that a double-dose of
360 intensive patient-specific intervention confers, on average, approximately 30% increased
361 improvement as measured on the ANELT A-scale in the chronic phase.

362

Two studies investigated the effect of a double-dose of constraint-based therapy. Stahl and colleagues [45] conducted an RCT comparing high intensity Intensive Language Action Therapy (ILAT) (4 hours a day for two 2-week therapy periods) versus low intensity (2 hours a day, two 2-week blocks) for people in the chronic phase of recovery post stroke (n = 30). The two groups received different total doses (48 and 24 hours, respectively) at different intensities. Results demonstrated statistically significant improvements in language and communication outcomes for both groups ($0.4 <$ Cohen's $d \leq 1.4$) but no significant interaction of time and group [$F(3, 78) = 0.80$, NS] suggesting that, while both groups improved, there was no added benefit of receiving an additional 24 hours of ILAT within a four-week treatment period.

Mozeiko and colleagues [47] investigated the effect of a double administration (total dose: 60 hours) of modified Constraint-Induced Language Therapy (CILT) in a small (n = 4) quasi-experimental "AB with follow up" design. Naming accuracy and informativeness measures were compared to baseline performance after each of the two treatment phases and effect sizes calculated for each phase. Close inspection of reported effect sizes reveals variable treatment responses across participants (table 5).

< Table 5 Busk & Serlin's d effect size ranges by treatment outcome reported in Mozeiko et al., [47] >

Evidence from studies reporting dose as a count of therapeutic elements

Two studies employing single-case designs investigated the effect of computer-assisted cued picture naming treatment on measures of language impairment in the chronic phase of post-stroke recovery.

Harnish and colleagues [31] reported a case series (n = 8) exploring the effect of a cued picture naming paradigm on picture naming accuracy using a high session dose in a

388 "saturated" practice schedule [27, p.S287]. In this study, each picture (n = 50) was presented
389 within a protocolised cueing hierarchy allowing eight naming attempts per picture, totalling
390 400 naming attempts per session. The total dose was 3,200 naming attempts. Within-subject
391 analysis demonstrated that six participants achieved statistically significant gains in picture
392 naming accuracy after one treatment session (400 naming attempts). The remaining two
393 participants achieved significant gains after three sessions (1,200 attempts). Based on change
394 in confrontation naming accuracy of trained picture items, the overall treatment period
395 yielded small (n = 5/8), medium (n = 1/8), and large (n = 2/8) Busk and Serlin's *d* effect
396 sizes, as per lexical retrieval benchmarks [55]. Six of the seven participants with follow up
397 measures maintained these gains on trained items, and two of seven on untrained items, at
398 approximately 60-days follow up.

399 Building on these preliminary findings, Off and colleagues [32] compared the effects
400 of lower- and higher-dose of therapeutic inputs on confrontation naming for people with
401 chronic aphasia (n = 7). Pictures in the low-dose condition (n = 20) were presented once per
402 session, whereas pictures in high-dose condition (n = 20) were shown four times. Each
403 picture presentation involved two naming attempts, one cued and one uncued, resulting in 40
404 naming attempts per low-dose condition and 160 per high-dose condition per session. The
405 high-dose condition resulted in large effect sizes for two participants (P1, P7) and a small
406 effect size for one (P2) whereas the low-dose condition resulted in a medium effect size for
407 one participant (P7), relative to lexical retrieval benchmarks. All other effect sizes for the
408 remaining participants and dose conditions were negligible (i.e., $d < 4.0$).

409 Tau-U effect sizes were calculated for confrontation naming of treated items
410 immediately post treatment for these two studies (figure 4). The two dose conditions
411 administered by Off and colleagues [32] were analysed separately. Therefore, three effect
412 sizes representing each total dose of naming attempts across the two studies are presented.

The overlap of 95% Confidence Intervals suggest that cued picture naming therapy is effective and no dose condition from these two studies is significantly superior.

< Figure 4 Tau-U effect sizes in cued picture naming therapy studies exploring dose effects >

DISCUSSION

The aim of this systematic review was to examine and compare evidence for dose effects in behavioural treatments for post-stroke aphasia. This review is important to improve the prescription of treatment for people living with aphasia, to optimise the delivery of clinical rehabilitation services, and to inform our theoretical understanding of language processing and recovery following stroke.

The investigation of dose effects in aphasia, and more broadly, stroke rehabilitation, is an emerging research area [e.g., [56]. This systematic review shows that the current state of research is exploratory; there is very limited evidence in the published literature regarding dose effects on impairment-level and activity/participation outcomes, while no evidence from these experimental studies was found for quality of life outcomes related to treatment dose. Preliminary attempts to experimentally control dose parameters have been reported and results from these studies provide a starting point from which to build a focused research agenda. Although based on limited evidence, there are a number of trends in the literature that warrant exploration. The results will now be discussed within the context of the existing literature for each research question addressed in this review.

Does a larger dose of intervention result in better language and communication outcomes for people with aphasia following stoke?

Unlike previous reviews, the current review specifically set out to examine dose effects in studies that provided different amounts of the same intervention. Only three studies

conducted planned comparisons of dose effects and three studies conducted exploratory post-hoc analysis after participants received different doses through deviations to the prescribed treatment schedule.

One study conducted in the acute-subacute phase of post-stroke recovery did not find a dose effect [42]. It has been suggested that the higher dose and more intensive group in this study did not receive enough treatment to elicit statistically significant treatment effects relative to the conventional group [57], which may be true given suspicions that behavioural stroke rehabilitation interventions are under-dosed potentially by several orders of magnitude [58, 59]. However, higher doses of treatment provided over a short duration may not be agreeable or tolerable for people in the early stages of recovery after stroke (*see below*).

Findings are difficult to compare due to the different interventions, outcomes, and treatment schedules used. Participants in Breitenstein et al. [43] who received more therapy did so over a longer intervention duration relative to their lower-dose counterparts, while Stahl et al. [45] purposively increased the number of hours per day for the intensive group while maintaining a fixed intervention duration. In very broad terms, the data suggest that a dose of 60 hours of functional, multicomponent, patient-specific intervention provided at 10 hours per week results in marginally better functional communication outcomes than 30 hours [43], while a dose of 48 hours of constraint treatment (ILAT) confers no additional benefit than 24 hours provided over the same four-week intervention period when treatment effect is measured using impairment-level outcomes [45]. It is possible that functional communication may have a higher threshold to show an effect of treatment due to increased demands on multiple levels of linguistic processing and cognitive skills whereas performance on impairment-level measures, such as confrontation naming tasks, may reach a ceiling relatively sooner due to the discrete and specific nature of isolated linguistic processing skills and tasks [58, 60].

Furthermore, it was not possible to ascertain from the reported group-level data which participants were responders and the potential impact of aphasia severity on treatment response. Like Breitenstein and colleagues, Mozeiko et al. [47] found that additional treatment blocks may add value for some participants and with a diminished return. Participants with mild aphasia benefited from the second treatment phase for impairment-level outcomes but not necessarily for discourse-level measures of informativeness, whereas severe aphasia may be associated with an opposite pattern of improvement, although the evidence for this assertion is based on a very small sample. This finding appears contradictory to Stahl et al. [45], which is curious given both Stahl et al. and Mozeiko et al. utilised constraint induced therapies. However, both the Mozeiko and Breitenstein studies provided 25% more treatment hours than Stahl (60 hours vs 48 hours) and perhaps this demonstrates a dose threshold.

In summary, when effects of longer treatment duration were observed, the additional treatment resulted in improvements that were roughly half the size of improvements associated with the initial dose. This suggests that higher doses (even when tolerated) may be associated with diminishing returns in the chronic phase (*see below*). More research is required to examine correlations between aphasia severity and dose effects.

Two studies demonstrated that manipulation of the delivery of therapeutic elements has the potential to increase the efficiency of treatments whereby gains in language skills can be achieved after relatively brief intervention periods [31, 32]. In traditional dose terms, the preliminary results from Harnish and colleagues [31] suggest that approximately one hour of cued picture naming treatment is sufficient to elicit modest, statistically significant gains in naming accuracy for some people with chronic aphasia. The key learnings from these studies are that treatment dose can be increased independently of treatment duration by increasing

488 the number of therapeutic inputs provided within a session of fixed duration and that people
489 with aphasia across the severity continuum can tolerate these high session doses in the
490 chronic phase of recovery. However, there is currently insufficient evidence to determine if
491 higher session dose is superior to lower session dose for acquisition, generalisation, and
492 maintenance of picture naming skills. Further experimental comparison of low and high dose
493 conditions is required across larger participant cohorts, follow up periods, and using
494 measures more closely aligned to impacts on interaction and quality of life.
495
496 Does time post stroke impact dose effects?
497 The feasibility and appropriateness of delivering high doses of language treatment
498 over a brief intervention duration in the early recovery period following stroke is
499 questionable [61]. High doses may need to be spread out over a longer duration to reduce
500 treatment intensity to tolerable levels and to maximise treatment effect [33]. Furthermore,
501 there are many sequelae of stroke that impact a person's ability to participate in
502 neurorehabilitation during the acute/subacute recovery phase. In the Bakheit study, nearly
503 twice as many people from the intensive group failed to complete the prescribed therapy
504 protocol (n = 20 intensive, n = 11 conventional); many refused treatment or were too ill to
505 participate, particularly in the first four weeks of recovery [42]. Providing more than 50 hours
506 of therapy to this population in the acute-subacute period of recovery therefore appears
507 neither practical nor appropriate. This study provides a preliminary range estimate of the
508 maximum tolerable dose of between 20 to 60 hours of treatment during the acute-subacute
509 period. More evidence is needed to refine this range estimate for the early stages of recovery
510 following stroke, and to determine if such ranges differ between treatments.
511 Intervention tolerance may be less of an issue in the chronic phase of recovery. Many
512 tolerated 48 hours [45] and some up to 60 hours [43, 47] of impairment-based treatment with

513 no reported increase in drop outs, refusal to participate, or increased frustration. Likewise,
514 studies that provided high session doses were also well tolerated [e.g., [31, 32]. Although not
515 common, some degradation of performance was noted, with 10% of participants in
516 Breitenstein et al. [43] showing a deterioration of ≥3 points on the ANELT A-scale after 3
517 weeks of intensive treatment. While intervention tolerance may be greater in the chronic
518 phase, some people with aphasia may stop benefitting from treatment before it becomes
519 intolerable [45]. There is likely to be significant variability in individual tolerance of high
520 dose, high frequency interventions and this requires investigation.
521
522 Are there specific person-level characteristics that help explain variability in dose-response
523 relationships?
524 Treatment responsiveness is mediated by factors intrinsic and extrinsic to the person
525 with aphasia [62]. The mixed findings from studies included in this review may well be
526 explained by complex interactions between person-related variables such as aphasia severity,
527 time post stroke, and motivation as well as treatment-related variables such as intervention
528 type and treatment schedule. Furthermore, domain-general cognitive processes such as ability
529 to attend, maintain focus, self-monitor, and self-motivate are likely to play a significant role
530 in intervention tolerance and, therefore, treatment response [63-65]. How and to what extent
531 these person- and treatment-level variables mediate treatment response is not yet well
532 understood. Large trials that recruit heterogenous samples have the advantage of producing
533 conclusions that may be broadly applicable to diverse clinical populations at the expense of
534 detailed person-level recommendations, [e.g., [42, 43, 45]. In small studies, sample
535 heterogeneity often complicates the interpretation and synthesis of results while allowing
536 deeper exploration of factors likely to explain variations in dose-response relationships, [e.g.,
537 [31, 32, 47].

538 Inadequate description and control of person-level and treatment-level factors makes
539 interpretation of these findings difficult. More experimental research is needed to explore
540 factors mediating the dose-response relationship.
541
542 Is there evidence of dose effects in specific language or communication interventions?
543 *Constraint approaches*
544 Constraint treatments have demonstrated efficacy with a dose of 30 hours provided
545 over two weeks [e.g., [66, 67]. It remains unclear whether similar treatment effects could be
546 achieved with lower dose therapy. Results from the two studies comparing different doses of
547 constraint treatments are inconsistent: Stahl et al. [45] found that 48 hours of ILAT was not
548 superior to 24 hours over the same treatment period whereas Mozeiko et al. [47] found
549 positive treatment effects after both treatment phases of modified CILT. Further direct
550 comparison is required to determine the optimal dose of constraint treatments.
551 *Cued picture naming therapy*
552 Cued picture naming therapy also has demonstrated efficacy and the two studies
553 employing cued picture naming in this review demonstrated positive treatment effects. While
554 all participants' picture naming improved in these studies, with some evidence of
555 maintenance at follow up, the magnitude of improvement varied across participants and there
556 is insufficient evidence to determine the optimal dose of cued picture naming therapy for any
557 particular individual.
558
559 Additional emerging factor: Arbitrary dose prescription
560 To date, all dose prescription has been arbitrary. As Doogan and colleagues note: "It
561 makes intuitive sense that our trial designs should not be constrained by a set dose when we
562 have no clear guidance as to what this should be" [29, p.90]. To further our understanding of

563 the dose-response relationship, it is critical to identify appropriate individualised doses of
564 treatment and failure to do so may contribute to inconsistent dose-effect findings. One
565 possible step toward achieving this long-term goal is to calibrate treatment dose relative to
566 one or more observable person-level baseline characteristics. In lexical retrieval paradigms,
567 one such variable is naming response time, that is, the amount of time taken to correctly
568 name a picture stimulus [68]. In picture naming treatments, there is a theoretical maximum
569 number of pictures that can be named in a given amount of time for a given individual. It may
570 be possible to use an individual PWA's baseline picture naming response times to
571 theoretically determine a maximum session dose for that individual against which alternative,
572 lower doses could be calibrated. Within-subject comparison of individualised dose conditions
573 may elucidate a 'sweet spot' at which optimal acquisition and maintenance of picture naming
574 skills is achieved. Future research should explore individualised calibration of dose relative to
575 baseline person-level characteristics.
576
577 ***Future directions for research on treatment dose in post-stroke aphasia***
578 Consensus definitions for dose parameters in aphasia interventions are required.
579 Inconsistent measurement and reporting of dose parameters across the aphasia literature
580 stems from a lack of standard definitions [19]. Multidisciplinary collaboration across stroke
581 recovery is required to establish core dose constructs. Consistent use of terminology will
582 have important implications for the development, implementation, and evaluation of dose
583 effect studies, for synthesis of data across these studies, for the theoretical exploration of
584 what drives treatment response in these interventions, for clinical decision-making regarding
585 service delivery, and for health policy makers. Once consensus definitions are in place,
586 reporting guidelines (e.g., TIDieR) need to be extended to encourage systematic routine
587 measurement and reporting of dose parameters.

More research is required comparing the effects of higher and lower doses of the same intervention on the acquisition, maintenance, and generalisation of language skills and communication. As yet, no group-level studies have attempted to comparatively manipulate therapeutic elements to examine dose effects. Large-scale studies should be informed by evidence from high-quality single-case experimental design studies exploring which therapeutic elements confer treatment effects and how these effects are mediated by person- and treatment-level factors. Furthermore, the use of alternative study designs (e.g., dose escalation or dose ranging methodologies) should be explored for applicability to aphasia treatment research.

Finally, sophisticated modelling techniques (e.g., linear mixed effects, Bayesian approaches) are required to estimate the relative effects of session dose, session frequency, intervention duration, and treatment intensity on treatment outcomes.

Limitations

There are a number of limitations pertaining to the studies included in the review and the methods employed in conducting the review.

Limitations of included studies

The small-scale studies included in this review are quasi-experimental. Experimental small-scale designs are required to explore causal dose-response relationships before scaling up to Phase II dose feasibility studies and Phase III dose effectiveness studies. Intensity and dose are confounded in the studies included in this review, meaning that more careful work is needed to determine what dose effects are independent of intensity. Furthermore, there is an absence of discussion regarding *task difficulty* as a parameter of dose and intensity in the aphasia literature. Ultimately, the small number of disparate studies comparing the effect of

providing different amounts of the same post-stroke aphasia intervention precluded conclusive answers to our research questions.

Limitations of review methods

In conducting a recent scoping review [19], we found a number of dose comparison studies and determined it would be appropriate to attempt meta-analysis of data from these studies. The decision to conduct meta-analysis was made after considerable work had been conducted. Therefore, this review was not protocolised or registered with PROSPERO.

Our meta-analysis of single-case data was limited to an exploration of skill acquisition and did not address maintenance or generalisation. For these studies, we chose to calculate Tau-U, which is gaining popularity as an adjunct to traditional visual analysis [69]. However, it is only a valid comparison of adjacent phases [40]. In impairment-based aphasia treatments, the treatment effect is not predicted to resolve in the post-treatment phase, therefore we could not compare baseline to maintenance phases (e.g., Harnish data). Further, it was not possible to calculate Tau-U for the Mozeiko study which provided double dose across different treatment levels with an intervening no-treatment period. Alternative statistical methods for evaluating, modelling, and synthesising treatment response would assist future analysis.

CONCLUSION

Treatment dose research in aphasia is an emerging area with few studies comparing different doses of the same intervention. There are indications in the literature that increased dose may confer greater improvement on language and communication measures, but with diminishing returns over time. Large-scale group studies comparing dose effects have used total hours of treatment as the measure of dose which lacks the specificity to examine dose-response relationships. Conversely, small-scale studies experimentally exploring therapeutic

637 elements provide a test bed for closer examination of person- and treatment-level factors
638 mediating treatment response. A dedicated and coordinated research agenda is required to
639 systematically explore dose-response relationships in post-stroke aphasia research.

Bibliography

1. Mitchell, C., et al., *Prevalence of aphasia and dysarthria among inpatient stroke survivors: describing the population, therapy provision and outcomes on discharge.* Aphasiology., 2020.
2. El Hachioui, H., et al., *Long-term prognosis of aphasia after stroke.* J Neurol Neurosurg Psychiatry, 2013. **84**(3): p. 310-315.
3. Hilari, K., J.J. Needle, and K.L. Harrison, *What are the important factors in health-related quality of life for people with aphasia? A systematic review.* Archives of physical medicine and rehabilitation, 2012. **93**(1): p. S86-S95.
4. Pedersen, P.M., K. Vinter, and T.S. Olsen, *Aphasia after stroke: type, severity and prognosis.* Cerebrovascular Diseases, 2004. **17**(1): p. 35-43.
5. Laska, A.C., et al., *Aphasia in acute stroke and relation to outcome.* Journal of internal medicine, 2001. **249**(5): p. 413-422.
6. Ellis, C., et al., *The one-year attributable cost of poststroke aphasia.* Stroke, 2012. **43**(5): p. 1429-1431.
7. Dickey, L., et al., *Incidence and profile of inpatient stroke-induced aphasia in Ontario, Canada.* Archives of physical medicine and rehabilitation, 2010. **91**(2): p. 196-202.
8. Lam, J.M.C. and W.P. Wodchis, *The relationship of 60 disease diagnoses and 15 conditions to preference-based health-related quality of life in Ontario hospital-based long-term care residents.* Medical care, 2010: p. 380-387.
9. Brady, M., et al., *Speech and Language Therapy (SLT) for aphasia after stroke: cochrane systematic review evidence of therapy regimens, delivery models and theoretical approaches.* European stroke journal, 2016. **1**(1): p. 693-.

10. Godecke, E., et al., *Amount of therapy matters in very early aphasia rehabilitation after stroke: a clinical prognostic model.* Semin Speech Lang, 2013. **34**(3): p. 129-41.
11. Palmer, R., et al., *Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (Big CACTUS): a multicentre, single-blinded, randomised controlled trial.* The Lancet Neurology, 2019. **18**(9): p. 821-833.
12. Wenke, R., et al., *Communication and well-being outcomes of a hybrid service delivery model of intensive impairment-based treatment for aphasia in the hospital setting: a pilot study.* Disabil Rehabil, 2018. **40**(13): p. 1532-1541.
13. Kong, A.P.-H. and C.W.-K. Tse, *Clinician survey on speech pathology services for people with aphasia in Hong Kong.* Clinical Archives of Communication Disorders, 2018. **3**(3): p. 201-212.
14. Manning, M., et al., *Supporting people with post-stroke aphasia to live well: A cross-sectional survey of Speech & Language Therapists in Ireland.* Health & Social Care in the Community, 2020. **27**: p. 27.
15. Rose, M., et al., *Aphasia rehabilitation in Australia: Current practices, challenges and future directions.* International Journal of Speech-Language Pathology, 2014. **16**(2): p. 169-180.
16. Bhogal, S.K., R. Teasell, and M. Speechley, *Intensity of aphasia therapy, impact on recovery.* Stroke, 2003. **34**(4): p. 987-93.
17. Robey, R.R., *A meta-analysis of clinical outcomes in the treatment of aphasia.* J Speech Lang Hear Res, 1998. **41**(1): p. 172-87.
18. Baker, E., *Optimal intervention intensity.* International Journal of Speech-Language Pathology, 2012. **14**(5): p. 401-409.

19. Harvey, S.R., et al., *Dose effects in behavioural treatment of post-stroke aphasia: a systematic review.* under review.
20. Warren, S.F., M.E. Fey, and P.J. Yoder, *Differential treatment intensity research: A missing link to creating optimally effective communication interventions.* Mental retardation and developmental disabilities research reviews, 2007. **13**(1): p. 70-77.
21. Cherney, L.R., *Aphasia treatment: intensity, dose parameters, and script training.* Int J Speech Lang Pathol, 2012. **14**(5): p. 424-31.
22. Turkstra, L., Rocio, N., Whyte, J., Dijkers, M. P., Hart, T., *Knowing What We're Doing: Why Specification of Treatment Methods Is Critical for Evidence- Based Practice in Speech-Language Pathology.* American Journal of Speech-Language Pathology, 2016. **25**: p. 164-171.
23. Bhogal, S.K., et al., *Rehabilitation of Aphasia: More is better.* Topics in Stroke Rehabilitation, 2003. **10**(2): p. 66-76.
24. Godecke, E., et al., *Very early poststroke aphasia therapy: a pilot randomized controlled efficacy trial.* Int J Stroke, 2012. **7**(8): p. 635-44.
25. Denes, G., et al., *Intensive versus regular speech therapy in global aphasia: A controlled study.* Aphasiology, 1996. **10**(4): p. 385-394.
26. Carpenter, J. and L.R. Cherney, *Increasing aphasia treatment intensity in an acute inpatient rehabilitation programme: a feasibility study.* Aphasiology, 2016. **30**(5): p. 542-565.
27. Togher, L., *Challenges inherent in optimizing speech-language pathology outcomes: It's not just about counting the hours.* International Journal of Speech-Language Pathology, 2012. **14**(5): p. 438-442.

28. Thomas, L., et al., *Speech and language therapy for aphasia: parameters and outcomes.* Aphasiology, 2020. **34**(5): p. 603-642.
29. Snell, C., K. Sage, and M.A.L. Ralph, *How many words should we provide in anomia therapy? A meta-analysis and a case series study.* Aphasiology, 2010. **24**(9): p. 1064-1094.
30. Gravier, M.L., et al., *What matters in semantic feature analysis: Practice-related predictors of treatment response in aphasia.* American Journal of Speech-Language Pathology, 2018. **27**(1S): p. 438-453.
31. Harnish, S.M., et al., *Dosing of a Cued Picture-Naming Treatment for Anomia.* American Journal of Speech-Language Pathology, 2013. **23**(2): p. S285-99.
32. Off, C.A., et al., *The impact of dose on naming accuracy with persons with aphasia.* Aphasiology, 2016. **30**(9): p. 983-1011.
33. Doogan, C., et al., *Aphasia Recovery: When, How and Who to Treat?* Curr Neurol Neurosci Rep, 2018. **18**(12): p. 90.
34. Dignam, J., et al., *The relationship between novel word learning and anomia treatment success in adults with chronic aphasia.* Neuropsychologia, 2016. **81**: p. 186-197.
35. Liberati, A., et al., *The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration.* PLoS medicine, 2009. **6**(7): p. e1000100.
36. Ouzzani, M., et al., *Rayyan - a web and mobile app for systematic reviews.* Systematic Reviews, 2016. **5**(210).
37. *The Oxford 2011 Levels of Evidence*. 2011; Available from: http://www.cebm.net/index.aspx?o=5653.

38. Tate, R.L., et al., *The Risk-of-Bias in N-of-1 Trials (RoBiNT) scale: An expanded manual for the critical appraisal of single-case reports*. 2015: Robyn Tate.

39. Murray, E., et al., *The reliability of methodological ratings for speechBITE using the PEDro-P scale.* International journal of language & communication disorders, 2013. **48**(3): p. 297-306.

40. Parker, R.I., et al., *Combining nonoverlap and trend for single-case research: Tau-U.* Behavior Therapy, 2011. **42**(2): p. 284-299.

41. Parker, R.I., K.J. Vannest, and J.L. Davis, *Effect size in single-case research: A review of nine nonoverlap techniques.* Behavior Modification, 2011. **35**(4): p. 303-322.

42. Bakheit, A.M.O., et al., *A prospective, randomized, parallel group, controlled study of the effect of intensity of speech and language therapy on early recovery from poststroke aphasia.* Clinical Rehabilitation, 2007. **21**(10): p. 885-894.

43. Breitenstein, C., et al., *Intensive speech and language therapy in patients with chronic aphasia after stroke: A randomised, open-label, blinded-endpoint, controlled trial in a health-care setting.* The Lancet, 2017. **389**(10078): p. 1528-1538.

44. DeDe, G., E. Hoover, and E. Maas, *Two to Tango or the More the Merrier? A Randomized Controlled Trial of the Effects of Group Size in Aphasia Conversation Treatment on Standardized Tests.* J Speech Lang Hear Res, 2019. **62**(5): p. 1437-1451.

45. Stahl, B., et al., *Efficacy of intensive aphasia therapy in patients with chronic stroke: a randomised controlled trial.* J Neurol Neurosurg Psychiatry, 2018. **89**(6): p. 586-592.

46. Marshall, J., et al., *Evaluating the Benefits of Aphasia Intervention Delivered in Virtual Reality: Results of a Quasi-Randomised Study.* PLoS One, 2016. **11**(8): p. e0160381.

47. Mozeiko, J., E.B. Myers, and C.A. Coelho, *Treatment Response to a Double Administration of Constraint-Induced Language Therapy in Chronic Aphasia.* J Speech Lang Hear Res, 2018. **61**(7): p. 1664-1690.

48. Duncan, E.S. and S.L. Small, *Imitation-based aphasia therapy increases narrative content: a case series.* Clinical rehabilitation, 2017. **31**(11): p. 1500-1507.

49. Lee, J.B., R.C. Kaye, and L.R. Cherney, *Conversational script performance in adults with non-fluent aphasia: treatment intensity and aphasia severity.* Aphasiology, 2009. **23**(7): p. 885-897.

50. Herbert, R., D. Webster, and L. Dyson, *Effects of syntactic cueing therapy on picture naming and connected speech in acquired aphasia.* Neuropsychol Rehabil, 2012. **22**(4): p. 609-33.

51. Marshall, J., et al., *Computer delivery of gesture therapy for people with severe aphasia.* Aphasiology, 2013. **27**(9): p. 1128-1146.

52. Schuchard, J., K.A. Rawson, and E.L. Middleton, *Effects of distributed practice and criterion level on word retrieval in aphasia.* Cognition, 2020. **198**: p. 104216.

53. *PEDro statistics*. 2016 [cited 2020 February 12]; Available from: http://www.pedro.org.au/english/downloads/pedro-statistics/.

54. Pierce, J.E., et al., *Constraint and multimodal approaches to therapy for chronic aphasia: A systematic review and meta-analysis.* Neuropsychological Rehabilitation, 2019. **29**(7): p. 1005-1041.

55. Beeson, P.M. and R.R. Robey, *Evaluating single-subject treatment research: Lessons learned from the aphasia literature.* Neuropsychology review, 2006. **16**(4): p. 161-169.

56. Dalton, E.J., et al., *Dose Articulation in Preclinical and Clinical Stroke Recovery: Refining a Discovery Research Pipeline and Presenting a Scoping Review Protocol.* Frontiers in Neurology, 2019. **10**: p. 1148.

57. Aerts, A., et al., *Aphasia therapy early after stroke: behavioural and neurophysiological changes in the acute and post-acute phases.* Aphasiology, 2015. **29**(7): p. 845-871.

58. Nardo, D., et al., *Less is more: neural mechanisms underlying anomia treatment in chronic aphasic patients.* Brain, 2017. **140**(11): p. 3039-3054.

59. Varley, R., *Rethinking aphasia therapy: A neuroscience perspective.* International Journal of Speech-Language Pathology, 2011. **13**(1): p. 11-20.

60. Dignam, J.K., A.D. Rodriguez, and D.A. Copland, *Evidence for Intensive Aphasia Therapy: Consideration of Theories From Neuroscience and Cognitive Psychology.* Pm r, 2016. **8**(3): p. 254-67.

61. Nouwens, F., et al., *Early cognitive-linguistic treatment for aphasia due to stroke; A randomised controlled trial (rats- 3).* European Stroke Journal, 2017. **2**(1): p. 4.

62. Kiran, S. and C.K. Thompson, *Neuroplasticity of Language Networks in Aphasia: Advances, Updates and Future Challenges.* Frontiers in neurology, 2019. **10**: p. 295.

63. Brownsett, S.L.E., et al., *Cognitive control and its impact on recovery from aphasic stroke.* Brain, 2014. **137**(1): p. 242-254.

64. Harnish, S.M. and J.P. Lundine, *Nonverbal working memory as a predictor of anomia treatment success.* American journal of speech-language pathology, 2015. **24**(4): p. S880-S894.

65. Simic, T., et al., *Baseline executive control ability and its relationship to language therapy improvements in post-stroke aphasia: a systematic review.* Neuropsychological rehabilitation, 2019. **29**(3): p. 395-439.

66. Pulvermuller, F., et al., *Constraint-induced therapy of chronic aphasia after stroke.* Stroke, 2001. **32**(7): p. 1621-6.

67. Meinzer, M., et al., *Long-term stability of improved language functions in chronic aphasia after constraint-induced aphasia therapy.* Stroke, 2005. **36**(7): p. 1462-6.

68. Evans, W.S., et al., *How Much Time Do People With Aphasia Need to Respond During Picture Naming? Estimating Optimal Response Time Cutoffs Using a Multinomial Ex-Gaussian Approach.* Journal of Speech, Language, and Hearing Research, 2020. **63**(2): p. 599-614.

69. Lee, J.B. and L.R. Cherney, *Tau-U: A quantitative approach for analysis of single-case experimental data in aphasia.* American Journal of Speech-Language Pathology, 2018. **27**(1S): p. 495-503.

Figure captions

Figure 1 PRISMA flow diagram showing the study selection process

Figure 2 PEDro-P scale scores for included group studies with cut-off score ≥5

Figure 3 RoBiNT scale scores for included single case design studies with cut-off score ≥12

Figure 4 Tau-U effect sizes in cued picture naming therapy studies exploring dose effects

820 Table 1 Search terms relating to treatment dose in aphasia

Population	Intervention	Qualifier
Aphasia/ Dysphasia/ (aphasia OR dysphasia).ti.ab	Therapy/ Intervention/ Treatment/ Rehabilitation/ (therap* OR intervention OR treatment OR rehab*).ti.ab	Dose/ Dosage/ Amount/ Intensity/ Frequency/ (dose* OR dosage* OR amount* OR intensi* OR frequenc*).ti.ab

821

822 Table 2 Data items extracted from selected studies

Domain	Data items
Study characteristics	Author name, year, title, study design, ICF domain (i.e., impairment, activity/participation), treatment description, aphasia chronicity, and key findings relevant to dose
Participant characteristics	Sample size, age, education, gender, handedness, time post-onset, aetiology, aphasia type and severity, and aphasia severity rating measure
Dose characteristics	Session dose (duration and/or elements), session frequency, total intervention duration, total sessions, total dose (hours), total dose (elements)
Results	Statistical analyses utilised, acquisition, generalisation, maintenance

823

824 Table 3 Treatment schedules for studies included in analysis which reported total dose in hours

Study, design, and sample size	**SESSION DOSE**	**SESSION FREQUENCY**	**INTERVENTION DURATION**	**TOTAL DOSE**
	In minutes	Per week	In weeks	In time
Bakheit et al. [42] RCT n = 51 intensive group (IG) n = 48 conventional group (CG) Subgroup n = 13 from IG	60 minutes	*Prescribed, actual mean (SD)* IG: 5x/week, 3.1 (1.4) CG: 2x/week, 1.6 (0.5) Subgroup: 4.3 (1.0)	12 weeks	*Prescribed, actual mean (SD)* IG: 60 hours, 35.6 (16.4) CG: 24 hours, 19.3 (6.4) Subgroup: 51.6 (12.0)
Breitenstein et al. [43] RCT n = 78 intervention (IG) n = 78 control/treatment deferred (CG) Subgroup n = 34 n = 19 from IG n = 15 from CG	At least 60 minutes of individual/group treatment At least 60 minutes of self-directed treatment	At least 10 hours per week with therapist At least 5 hours per week of self-directed treatment	*Prescribed, actual median (IQR)* IG: At least 3 weeks, 4.8 (IQR 3.0-5.6) CG: 4.0 (3.0-5.0) Subgroup: 6 weeks (5-7)	*Prescribed, actual median (IQR)* IG: 30 hours, 31 (30-34.5) CG: 0 hours, 4.5 (3.0-6.8) Subgroup IG: 51.8 hours (47.2-58.0) CG: 48.0 hours (44.0-56.8)
Stahl et al. [45] RCT n = 30	60 minutes	G1: 4 hours/day, 3x/week G2: 2 hours/day, 3x/week	2 weeks x 2 = 4 weeks	Group 1: 48 hours Group 2: 24 hours
Mozeiko et al. [47] Case series AB+ design n = 4	180 minutes	5x/week	2 weeks x 2 = 4 weeks	60 hours

825

826 Table 4 Treatment schedules for studies included in analysis which reported total dose in hours and therapeutic elements

Study, design, and sample size	**SESSION DOSE**		**SESSION FREQUENCY**	**INTERVENTION DURATION**	**TOTAL DOSE**	
	In minutes	In therapeutic elements	Per week	In weeks	In time	In therapeutic elements
Harnish et al. [31] Case series AB+ design n = 8	60 minutes	50 picture presentations x 8 naming attempts per picture = 400 naming attempts per session	4x/week	2 weeks	8 hours	400 picture presentations 3200 naming attempts
Off et al. [32] Case series AB+ design n = 7	~60 minutes	20 pictures per dose condition *Low dose condition* 1 presentation per picture with 2 naming attempts = 40 naming attempts per session *High dose condition* 4 presentations per picture with 2 naming attempts = 160 naming attempts per session Total per session: 100 picture presentations, 200 naming attempts	2-3x/week	Up to 5 weeks	6-15 hours	<u>Picture presentations</u> Low dose: 120-300 High dose: 480-1200 Total: 600-1500 <u>Naming attempts</u> Low dose: 240-600 High dose: 960-2400 Total: 1200-3000

827

828 Table 5 Busk & Serlin's *d* effect size ranges by treatment outcome reported in Mozeiko et al.,

829 [43]

	Outcome	Treatment phase 1	Treatment phase 2
Impairment	Naming trained items	4.33 – 27.58 Participants with mild aphasia (n = 2) had larger ESs than participants with severe aphasia (n = 2)	-1.06 – 40.31 Only participants with mild aphasia demonstrated a response to second treatment phase.
	Naming untrained items	-0.86 – 21.92 Negligible treatment effect for participants with severe aphasia across both phases.	0.71 – 47.06 Participants with mild aphasia had larger treatment effects after the second treatment phase than after the first treatment phase.
Discourse	Average CIUs in narrative task	0.57 – 9.55	-0.63 – 12.22
	CIUs/min	-32.09 – 10.58 One participant with severe aphasia had large ES. Negligible ESs for other participants. One participant with mild aphasia had a marked decrease in CIUs/min although visual inspection suggests this result is due to a single outlier.	-1.94 – 12.17 A different participant with severe aphasia had large ES but no effect following first phase. Negligible ESs for other participants.
	% CIUs of total word count	-0.19 – 2.38 Negligible ESs for all participants.	-2.90 – 4.66 One participant with mild aphasia had small ES after second treatment phase. Negligible ESs for other participants.

830