Title: Predicting youth athlete sleep quality and the development of a translational tool to inform practitioner decision-making

Submission type: Original research (Cross-sectional study)

Running head: Predicting youth athlete sleep quality

Abstract word count: 303

Word count: 3123

Number of figures: 2

Number of tables: 3

Abstract

Background: Identifying key variables that predict sleep quality in youth athletes allows practitioners to monitor the most parsimonious set of variables that can improve athlete buy-in and compliance for athlete self-report measurement. Translating these findings into a decision-making tool could facilitate practitioner willingness to monitor sleep in athletes.

Hypothesis: Key predictor variables, identified by feature reduction techniques, will lead to higher predictive accuracy in determining youth athletes with poor sleep quality.

Study Design: Cross-sectional study.

Level of Evidence: Level 3.

Methods: A group (N = 115) of elite youth athletes completed questionnaires consisting of the Pittsburgh Sleep Quality Index (PSQI) and questions on sport participation, training, sleep environment and sleep hygiene habits. A least absolute shrinkage and selection operator (LASSO) regression model was used for feature reduction and to select factors to train a feature-reduced sleep quality classification model. These were compared to a classification model utilizing the full feature set.

Results: Sport type, training before 8 am, training hours per week, pre-sleep computer usage, pre-sleep texting or calling, pre-bedtime reading and during-sleep time checks on digital devices were identified as variables of greatest influence on sleep quality and used for the reduced feature set modeling. The reduced feature set model performed better (AUC: 0.80, Sensitivity: 0.57, Specificity: 0.80) than the full feature set models in classifying youth athlete sleep quality.

Conclusions: The findings of our study highlight that sleep quality of elite youth athletes is best predicted by specific sport participation, training and sleep hygiene habits.

Clinical Relevance: Education and interventions around the training and sleep hygiene factors that were identified to most influence the sleep quality of youth athletes could be prioritized to optimize their sleep characteristics. The developed sleep quality nomogram may be useful as a decision-making tool to improve sleep monitoring practice amongst practitioners.

Keywords: Youth athletes, sleep quality, feature reduction, machine learning, nomogram

Introduction

The interest in sleep as a means to optimize athlete physiological and psychological health, well-being and mood, cognition, and performance has increased dramatically in the past decade. [22] This can be attributed to the recognized role of sleep in facilitating these positive outcomes, as well as the documented prevalence of chronic sleep restriction experienced by high-performing athletes [14] and adolescent athletes. [34] Research on the sleep regulatory systems of adolescents have highlighted a potential interaction between a multitude of bioregulatory and psychosocial pressures that contribute to shortened sleep in this population. [8] Amongst these considerations, the biologically-governed phase-delay of circadian rhythms, extrinsic alerting factors, such as evening light from devices with screens, and the societal pressure of early awakenings for early school starts or sports practices culminate towards sleep opportunities that are chronically short and quality deficient. [8]

While the practice to capture details about an athlete's sleep durations and/or quality has been generally adopted by way of athlete monitoring systems to understand the inter-relationships between training demands, wellness and potential maladaptive recovery profiles, [15, 19] there appears to be a lack of data on the factors that contribute to unsatisfactory sleep characteristics amongst athletes, evidenced by recent recommendations for such information to be obtained through sleep questionnaires. [35] From the perspective of a sports practitioner, it would be relevant to also obtain insights into modifiable contributory factors that influence the sleep characteristics of their athletes, and by extension, potentially determine how these factors impact the sleep quality experienced by an athlete throughout the season. It is not always possible to employ objective methods to monitor sleep due to logistical and fiscal considerations. [17] Self-report sleep monitoring instruments provide valuable estimates of sleep characteristics but may be tedious to employ regularly and often require expert guidance or referral to a sleep professional to address the underlying symptoms.

Athlete monitoring systems, or athlete self-report measures (ASRM), would be a seemingly logical alternative to gather insights on factors contributing to sleep disturbance and determine sleep quality. However, because of the regularity and comprehensiveness of data collected in such an approach, practitioners have to consider the trade-off in the number of questions asked and athlete compliance in data collection (i.e., data meaningfulness vs data completeness) when designing the ASRM. [32] In this regard, the process of feature selection, which seeks to reduce dimensionality by 1) identifying a subset of variables critical to the construction of the prediction model and 2) removing irrelevant variables, can be a useful methodology to determine the most parsimonious set of variables from athlete monitoring data, [29] ensuring the most optimal balance of athlete buy-in [31] and potentially improve prediction accuracy.

While developing statistical models to predict sleep quality may be warranted from a research perspective, efforts could be made to translate them to an applied sport science setting. In this regard, a rapidly deployable visual tool to communicate the results of a predictive model assessing the risk of poor sleep quality in youth athletes could facilitate decision-making with respect to recovery. Recent translational medicine approaches in healthcare research have begun to propose nomograms as a tool to screen for individuals at greater risk of certain medical maladies. [20, 36] A novel sleep-specific nomogram that provides a point estimate of a youth athlete's quality of sleep based on predictors of youth athlete sleep quality could allow practitioners to easily identify athletes at risk of poorer sleep and decide on the extent of appropriate management strategies that should be employed when necessary.

The aims of this study were to identify factors that best predict youth athlete sleep quality. Specifically, we aim to use a feature reduction approach to develop a classification model using behavioural, environmental and training factors to predict sleep quality. Using the most accurate predictive model, a secondary aim was to develop a nomogram as a visualization and translational tool to facilitate the interpretation of the predictive analysis for practitioner adoption.

Methods

Participants

A total of 115 male and female elite national youth athletes from 11 National Sports Associations (NSAs) across Singapore were recruited. Convenience sampling was used to select participants. Only youths under 21 years of age were included in this study. Informed consent and assent were given by participants and their parent or guardian. The characteristics of the athletes are presented in Table 1.

*****Insert Table 1 about here*****

Design

A cross-sectional observational design was adopted. The self-administered questionnaire included the Pittsburgh Sleep Quality Index (PSQI), a Sleep Environment and Hygiene Survey and the athletes' training and demographic factors. The paper questionnaire was administered once within a group setting during national training assemblies and took approximately 10 minutes to complete. Ethical approval for the study was obtained from an Institutional Review Board.

Procedure

Pittsburgh Sleep Quality Index Questionnaire.

The PSQI is a valid sleep monitoring tool consisting of 19 self-rated questions that provide a global rating of subjective sleep quality and sub-scale measures of sleep quality, sleep latency, sleep duration, sleep efficiency, sleep disturbances, sleep medication consumption habits and daytime dysfunction. [6] Its administration within initial consultations allows practitioners to determine an athlete's sleep history and identify potential sources of sleep disturbances based on responses and scores in individual components. [17] A global PSQI score of > 5 is reported to have a diagnostic sensitivity of 89.6% and specificity of 86.5% in distinguishing "good" and "poor" sleepers,[6] and was used as the criterion threshold in the present study. The binary classifications of "good" and "poor" sleepers were used as dependent variables.

Sleep environment and hygiene survey.

Sleep environment and hygiene information were collected as predictor variables using a 29-item survey, adapted from the Athlete Sleep Behaviour Questionnaire,[9] on maladaptive behaviours that may influence sleep (Supplementary material 1). The frequency of personal experiences relating to the sleep environment was evaluated on a six-point Likert scale (1 = Never; 6 = Always/daily), including questions pertaining to the immediate sleep environment (noise, light and temperature). Additionally, questions relating to food and fluid intake in the hour preceding sleep, the frequency of electronic device use and the application of blue light filters on electronic devices, the use of stress management and relaxation techniques, irregular sleep-wake times and poor napping habits were investigated. Questions pertaining to lifestyle activities associated with the peri-sleep time, including work, study, media and social media interaction, gaming, reading, music and texting, were also included.

Athlete demographic and training factors.

Athlete-specific data were collected as additional predictor variables. Athletes indicated their age, gender, sport, sport type (individual or team sport), current training phase, training frequency, training time of day and total training hours per week. For training frequency, response options were: Never, 1–2 Trainings/Week, 3-4 Trainings/Week, 5-6 Trainings/Week, and Daily. Training time of day responses were: Earlier than 8 am, 8 am to 5 pm and Later than 5 pm.

Statistical Methods

Statistical analysis was conducted using R statistical software (R. 4.0.2, R Foundation for Statistical Computing). Descriptive data were expressed in means and standard deviations.

Feature reduction -LASSO

In the current study, to uncover the smallest subset of features that best predict sleep quality while maintaining the practical interpretability of the features, we utilized a LASSO regression for feature reduction as it addresses the issues of overfitting and model performance overestimation commonly related to regression methods. [27] The LASSO regression places a penalty on the sum of the coefficients, using that penalty to shrink the coefficient estimates with a minor contribution to the model. The final model may have the coefficients of some variables shrunk all the way to zero, effectively removing them from the model. This results in a smaller set of predictor variables and a more parsimonious and interpretable model. [25]

Forty-two predictor variables from the questionnaire were used as predictor variables. Continuous variables were standardized (z-score) to reduce the influence of magnitude differences in measurement scales on the shrinkage of regression coefficients. Model fit was performed using the *glmnet* package [13] in R with 10-fold cross-validation. Alpha was set as 1, and the optimal value for lambda (0.058) was determined using a hyperparameter grid-search. The most important predictive features determined from the LASSO regression were then used in the model development.

Data pre-processing

Model training and evaluation occurred in R using the *caret* package. [21] A generalized linear model (GLM, logistic regression) classification approach was chosen to predict the categorical outcomes of sleep quality (poor vs good). A 70-30% split was selected for the training and testing sets, respectively, by random sampling. Two models were trained and tested, one using all 42 features (full-feature model) and another using features determined to be significant predictors by the LASSO feature reduction (feature-reduced model).

Model validation and evaluation

This study employed the use of 5-repeated 10-fold cross-validation. Cross-validation performance was used to determine the best of the 2 classification models (full-feature vs feature-reduced model). The predictive accuracy of the trained models was determined by the ability to correctly classify sleep quality within the "held-out" test data set. Area under the curve (AUC) in receiver operating characteristic (ROC), sensitivity and specificity were used to evaluate cross-validation performance between the trained models and predictive accuracies. The sum of sensitivity and specificity (Youden's method) was used for threshold selection.

Nomogram

The *rms* package in R [18] was used to construct the nomogram based on the final chosen model with the best predictive accuracy. The analysis workflow is presented in Figure 1.

*****Insert Figure 1 about here*****

Results

Feature reduction-LASSO

The LASSO regression reduced 42 variables to 7. These included *training hours per week, sport type, training frequency before 8 am, pre-bedtime computer usage frequency, pre-bedtime reading frequency, pre-bedtime text/call frequency and time check frequency*. The LASSO regression coefficients are presented in Table 2. Standard errors are not included in the table as they are not meaningful for biased estimates that arise from penalized estimation methods. [16]

*****Insert Table 2 about here*****

These variables were used to train the feature-reduced GLM model.

Model selection and accuracy

The feature-reduced model provided higher cross-validation performance (mean AUC: 0.75, mean sensitivity: 0.60, mean specificity: 0.80) than the full-feature model (mean AUC: 0.47, mean sensitivity: 0.44, mean specificity: 0.51). When evaluated upon the independent "held-out" test set, the feature-reduced model had a higher AUC (AUC: 0.80, Sensitivity: 0.57, Specificity: 0.80) with a threshold criteria of 0.39 in comparison to the full-feature model that reported an AUC of 0.55 (AUC: 0.55, Sensitivity: 0.57, Specificity: 0.80) with a best threshold criteria of 0.5 (Table 3).

*****Insert Table 3 about here*****

Nomogram

The nomogram was developed based on the feature-reduced model as it had the highest predictive accuracy. To aid readers in its usage, a case study is added to the nomogram (Figure 1a).

*****Insert Figure 2 about here*****

Discussion

Elite athletes may obtain a plethora of physiological and cognitive benefits from achieving adequate sleep. However, despite these health and performance advantages, the prevalence of sleep inadequacy among elite youth athlete populations has been reported as high. [35] Due to the various influences on youth athlete sleep, it would be vital to capture these multiple factors to inform their recovery needs. While ASRMs provide a means to gather these insights, it can be challenging to extract key variables that impact sleep quality from a high number of candidate predictors and maintain practitioner interpretability. To this end, this study aimed to identify key factors that best predict youth athlete sleep quality through machine learning techniques such as feature reduction (LASSO regression) and classification predictive modeling (GLM).

Of the 42 candidate variables analyzed, the LASSO regression identified 7 key variables as predictors of sleep quality amongst elite youth athletes. Of these predictor variables, *training frequency before 8 am*, *training hours per week*, *pre-bedtime computer usage frequency*, *pre-bedtime text/call frequency*, *and time check frequency* concur with previous literature in their influence on sleep. [1, 8, 12, 30, 33] In our findings, participating in a team sport resulted in a higher likelihood of having poor sleep quality, which is inconsistent with previous findings. [23] These equivocal reports may be explained by contextual differences. Given the possibility that athletes self-select into sports and accompanying training schedules that best suit their chronotype,[24] it is plausible that youth athletes with an evening preference may be more prevalent within team sports in the current population. Adolescents with this later chronotype preference have been reported to obtain lesser and poorer sleep due to a misalignment with societal demands.[28] Another notable finding was the association between total training hours per week and sleep quality. More specifically, a shorter duration of weekly training increased the risk of poor sleep quality amongst youth athletes within our study. This is consistent with findings by Brand et al.,[4, 5] who observed better subjective and objective sleep quality in adolescent athletes that participated in frequent exercise. This has been hypothesized to be due to the effect of exercise on brain energy metabolism in young people, resulting in improved sleep efficiency.[11] However, the cross-sectional design of the present study makes the nature of this association unclear as other factors, such as having a limited amount of time to undertake other pre-sleep activities known to disrupt sleep quality in adolescents, may also contribute to this relationship.[2] Our results also highlight an association between pre-bedtime reading frequency and sleep quality, whereby individuals that read more frequently had a lower likelihood of having poor sleep quality. Similar findings were reported in a study among Canadian children that found a positive impact of reading a printed book an hour before bedtime on sleep quality in comparison to the use of electronic devices during the same time of day.[10] The improvement in sleep quality following a reading opportunity among children and adolescents could be attributed to the reduced opportunity for blue light exposure from electronic devices, which have demonstrated an ability to suppress evening melatonin release, resulting in poorer sleep quality in adolescents.[7] However, melatonin and blue light exposure were not measured in the present study and further research to clarify the mechanisms underlying these associations is warranted.

When comparing the predictive accuracy of the full and feature-reduced classification models, the latter elicited greater predictive accuracy. This may be due to errors in the predictions of a model on a held-out data set (often referred to as out-of-sample or generalization error) which tend to increase if the model has been overfit to the data set used to construct it. [25] Adding more predictor variables to a model increases its complexity, and subsequently, its capacity to overfit a sample of data. This can cause increases in the generalization error, as seen in this study. The application of feature reduction could also be considered to refine athlete monitoring practices, whereby the number of questions asked from athletes is gradually reduced, following the identification of variables that do not contribute significantly to the predictive model. This is a salient consideration for practitioners looking to optimize athlete monitoring by improving athlete buy-in. Specifically, employing a more expedient data collection process allows for athletes and coaches to focus on data interpretation and generating critical feedback without overloading practitioner resources.

A secondary aim was to develop a simple-to-use nomogram as a novel visualization and translational tool to facilitate the interpretation of the predictive analysis to practitioners. The proposed nomogram in this study incorporates the selected training and pre-sleep hygiene variables used in the logistic regression model and can be used as a convenient, easy-to-read tool to identify youth athletes that may have a greater propensity for poor sleep (Figure 1b). This tool could be easily employed to perform a rapid assessment of sleep quality in youth athletes with minimal burden to the athlete by collecting parsimonious “risk factors” of poor sleep. This may facilitate athlete compliance, especially if monitoring is done frequently (e.g. daily), by incorporating a condensed set of critical questions. [32] As such, variable selection techniques and nomograms may present an even greater value-proposition in such a time where practitioners are monitoring wellness and performance parameters of multiple youth athletes (e.g., team sport settings, multi-sport support) and need to rapidly identify youth athletes that may require greater sport science support.

Previous instruments, such as the Athlete Sleep Behavior Questionnaire [9] and the clinically validated Athlete Sleep Screening Questionnaire,[3] were specifically developed to determine maladaptive sleep hygiene behaviors and identify sleep disorders in athletes. While there is great value in the use of these instruments, the nomogram has potential application as a quick and practitioner-friendly option within the early stages of the athlete sleep monitoring process to improve the frequency and perceived need for sleep monitoring by practitioners. [26] The ability of a sleep-specific nomogram to generate individualized predictions enables their use in the early identification and stratification of youth athletes' into groups defined by poor sleep quality risk, which could also better prioritize and target management interventions. This study extends on the previous development of sleep questionnaires providing practitioners with an easy-to-use nomogram that can help to guide decision-making on factors related to recovery and periodization of training.

Limitations and future research

The findings should be interpreted with caution. First, self-report tools like the PSQI are prone to recall bias. Objective measures, such as actigraphy, are valuable in their reliability and validity to quantify sleep and wake characteristics and are supplemental in the identification of potential sleep disorders. However, these methods require expertise in usage and are not always fiscally feasible. [17] The PSQI was adopted to determine sleep quality due to its time-cost effectiveness and accuracy. The cross-sectional approach employed in this study implies that the findings are associative and should not be interpreted as causal relationships. Interventions targeting the variables included in the predictive models have not been explicitly tested for their ability to remedy poor sleep quality. While the sleep nomogram in this study provides a useful reference to determine an athlete's sleep quality, it was developed using predictive models that were trained using data from Singaporean youth athletes. Alternate nomograms can be developed for variant athlete populations using their respective data.

Conclusions

Considering the complex considerations that may disturb athlete sleep, there is a need to use a data collection approach that is able to elucidate these multivariate factors. Athlete self-report measures are a cost-effective method to achieve this but can be overwhelming for the athlete when overly comprehensive and frequent. In the present study, we isolated 7 key sport participation, training and sleep hygiene variables that influence and best predict the sleep quality of elite youth athletes from a set of 42 candidate variables. This reduces the need to collect data on potentially irrelevant predictors and helps refine ASRM processes. Practitioners will also be able to better prioritize and focus their education and intervention strategies to optimize their athletes' sleep characteristics.

References

1. Adams SK, Kisler TS. Sleep quality as a mediator between technology-related sleep quality, depression, and anxiety. *Cyberpsychol Behav Soc Netw*. Jan 2013;16(1):25-30. doi:10.1089/cyber.2012.0157
2. Bartel KA, Gradisar M, Williamson P. Protective and risk factors for adolescent sleep: A meta-analytic review. *Sleep Medicine Reviews*. 2015/06/01/ 2015;21:72-85. doi:https://doi.org/10.1016/j.smrv.2014.08.002
3. Bender AM, Lawson D, Werthner P, Samuels CH. The Clinical Validation of the Athlete Sleep Screening Questionnaire: an Instrument to Identify Athletes that Need Further Sleep Assessment. *Sports Med Open*. Jun 4 2018;4(1):23. doi:10.1186/s40798-018-0140-5
4. Brand S, Gerber M, Beck J, Hatzinger M, Pühse U, Holsboer-Trachsler E. Exercising, sleep-EEG patterns, and psychological functioning are related among adolescents. *The World Journal of Biological Psychiatry*. 2010/01/01 2010;11(2):129-140. doi:10.3109/15622970903522501
5. Brand S, Gerber M, Beck J, Hatzinger M, Pühse U, Holsboer-Trachsler E. High Exercise Levels Are Related to Favorable Sleep Patterns and Psychological Functioning in Adolescents: A Comparison of Athletes and Controls. *Journal of Adolescent Health*. 2010/02/01/ 2010;46(2):133-141. doi:https://doi.org/10.1016/j.jadohealth.2009.06.018
6. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res*. May 1989;28(2):193-213. doi:10.1016/0165-1781(89)90047-4
7. Chang AM, Aeschbach D, Duffy JF, Czeisler CA. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *Proc Natl Acad Sci U S A*. Jan 27 2015;112(4):1232-7. doi:10.1073/pnas.1418490112

8. Crowley SJ, Wolfson AR, Tarokh L, Carskadon MA. An update on adolescent sleep: New evidence informing the perfect storm model. *Journal of Adolescence*. 2018/08/01/ 2018;67:55-65. doi:https://doi.org/10.1016/j.adolescence.2018.06.001

9. Driller MW, Mah CD, Halson SL. Development of the athlete sleep behavior questionnaire: A tool for identifying maladaptive sleep practices in elite athletes. *Sleep Sci*. Jan-Feb 2018;11(1):37-44. doi:10.5935/1984-0063.20180009

10. Dube N, Khan K, Loehr S, Chu Y, Veugelers P. The use of entertainment and communication technologies before sleep could affect sleep and weight status: a population-based study among children. *International Journal of Behavioral Nutrition and Physical Activity*. 2017/07/19 2017;14(1):97. doi:10.1186/s12966-017-0547-2

11. Dworak M, Wiater A, Alfer D, Stephan E, Hollmann W, Strüder HK. Increased slow wave sleep and reduced stage 2 sleep in children depending on exercise intensity. *Sleep Medicine*. 2008/03/01/ 2008;9(3):266-272. doi:https://doi.org/10.1016/j.sleep.2007.04.017

12. Fox JL, Scanlan AT, Stanton R, Sargent C. Insufficient Sleep in Young Athletes? Causes, Consequences, and Potential Treatments. *Sports Medicine*. 2020/03/01 2020;50(3):461-470. doi:10.1007/s40279-019-01220-8

13. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*. 2010;33(1):1-22.

14. Fullagar HH, Skorski S, Duffield R, Hammes D, Coutts AJ, Meyer T. Sleep and athletic performance: the effects of sleep loss on exercise performance, and physiological and cognitive responses to exercise. *Sports Med*. Feb 2015;45(2):161-86. doi:10.1007/s40279-014-0260-0

15. Gallo TF, Cormack SJ, Gabbett TJ, Lorenzen CH. Self-Reported Wellness Profiles of Professional Australian Football Players During the Competition Phase of the Season. *The*

Journal of Strength & Conditioning Research. 2017;31(2):495-502. doi:10.1519/jsc.0000000000001515

16. Goeman J, Meijer R, Chatuverdi N. L1 and L2 Penalized Regression Models. 2018:chap Section 6- A note on standard errors and confidence intervals. https://cran.r-project.org/web/packages/penalized/vignettes/penalized.pdf

17. Halson SL. Sleep Monitoring in Athletes: Motivation, Methods, Miscalculations and Why it Matters. *Sports Medicine*. 2019/10/01 2019;49(10):1487-1497. doi:10.1007/s40279-019-01119-4

18. *rms: Regression Modeling Strategies*. Version 6.2-0. 2016. https://CRAN.R-project.org/package=rms

19. Horgan BG, Drew MK, Halson SL, et al. Impaired recovery is associated with increased injury and illness: A retrospective study of 536 female netball athletes. *Scandinavian Journal of Medicine & Science in Sports*. 2021;31(3):691-701. doi:https://doi.org/10.1111/sms.13866

20. Huang J, Cheng A, Lin S, Zhu Y, Chen G. Individualized prediction nomograms for disease progression in mild COVID-19. *Journal of medical virology*. 2020;92(10):2074-2080.

21. *caret: Classification and Regression Training*. 2020. https://CRAN.R-project.org/package=caret

22. Lastella M, Memon AR, Vincent GE. Global Research Output on Sleep Research in Athletes from 1966 to 2019: A Bibliometric Analysis. *Clocks & Sleep*. 2020;2(2):99-119.

23. Lastella M, Roach GD, Halson SL, Sargent C. Sleep/wake behaviours of elite athletes from individual and team sports. *European Journal of Sport Science*. 2015/02/17 2015;15(2):94-100. doi:10.1080/17461391.2014.932016

24. Lastella M, Roach GD, Halson SL, Sargent C. The Chronotype of Elite Athletes. *J Hum Kinet*. Dec 1 2016;54:219-225. doi:10.1515/hukin-2016-0049

25. McNeish DM. Using Lasso for Predictor Selection and to Assuage Overfitting: A Method Long Overlooked in Behavioral Sciences. *Multivariate Behavioral Research*. 2015/09/03 2015;50(5):471-484. doi:10.1080/00273171.2015.1036965

26. Miles KH, Clark B, Fowler PM, Miller J, Pumpa KL. Sleep practices implemented by team sport coaches and sports science support staff: A potential avenue to improve athlete sleep? *J Sci Med Sport*. Jul 2019;22(7):748-752. doi:10.1016/j.jsams.2019.01.008

27. Ranstam J, Cook JA. LASSO regression. *British Journal of Surgery*. 2018;105(10):1348-1348. doi:10.1002/bjs.10895

28. Russo PM, Biasi V, Cipolli C, Mallia L, Caponera E. Sleep habits, circadian preference, and school performance in early adolescents. *Sleep Medicine*. 2017/01/01/ 2017;29:20-22. doi:https://doi.org/10.1016/j.sleep.2016.09.019

29. Ryan S, Kempton T, Coutts AJ. Data Reduction Approaches to Athlete Monitoring in Professional Australian Football. *Int J Sports Physiol Perform*. Nov 4 2020;16(1):59-65. doi:10.1123/ijspp.2020-0083

30. Sargent C, Lastella M, Halson SL, Roach GD. The impact of training schedules on the sleep and fatigue of elite athletes. *Chronobiol Int*. Dec 2014;31(10):1160-8. doi:10.3109/07420528.2014.957306

31. Saw AE, Kellmann M, Main LC, Gastin PB. Athlete Self-Report Measures in Research and Practice: Considerations for the Discerning Reader and Fastidious Practitioner. *Int J Sports Physiol Perform*. Apr 2017;12(Suppl 2):S2127-s2135. doi:10.1123/ijspp.2016-0395

32. Saw AE, Main LC, Gastin PB. Monitoring athletes through self-report: factors influencing implementation. *J Sports Sci Med*. Mar 2015;14(1):137-46.

33. Tang NKY, Anne SD, Harvey AG. Sleeping with the enemy: Clock monitoring in the maintenance of insomnia. *Journal of Behavior Therapy and Experimental Psychiatry*. 2007/03/01/ 2007;38(1):40-55. doi:https://doi.org/10.1016/j.jbtep.2005.07.004

34. Vlahoyiannis AA, G., Bogdanis GC, Sakkas GK, Andreou E, Giannaki CD. Deconstructing athletes' sleep: A systematic review of the influence of age, sex, athletic expertise, sport type, and season on sleep characteristics. *Journal of Sport and Health Science*. 2020/03/19/ 2020;doi:https://doi.org/10.1016/j.jshs.2020.03.006

35. Walsh NP, Halson SL, Sargent C, et al. Sleep and the athlete: narrative review and 2021 expert consensus recommendations. *British Journal of Sports Medicine*. 2021;55(7):356-368. doi:10.1136/bjsports-2020-102025

36. Xue M, Su Y, Feng Z, et al. A nomogram model for screening the risk of diabetes in a large-scale Chinese population: an observational study from 345,718 participants. *Scientific Reports*. 2020/07/14 2020;10(1):11600. doi:10.1038/s41598-020-68383-7

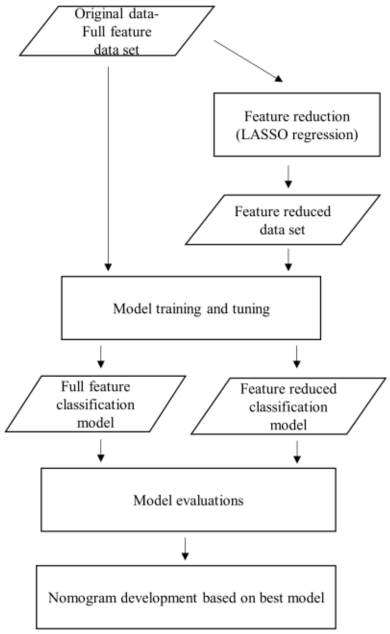

Figure 1. Feature selection and analysis workflow.

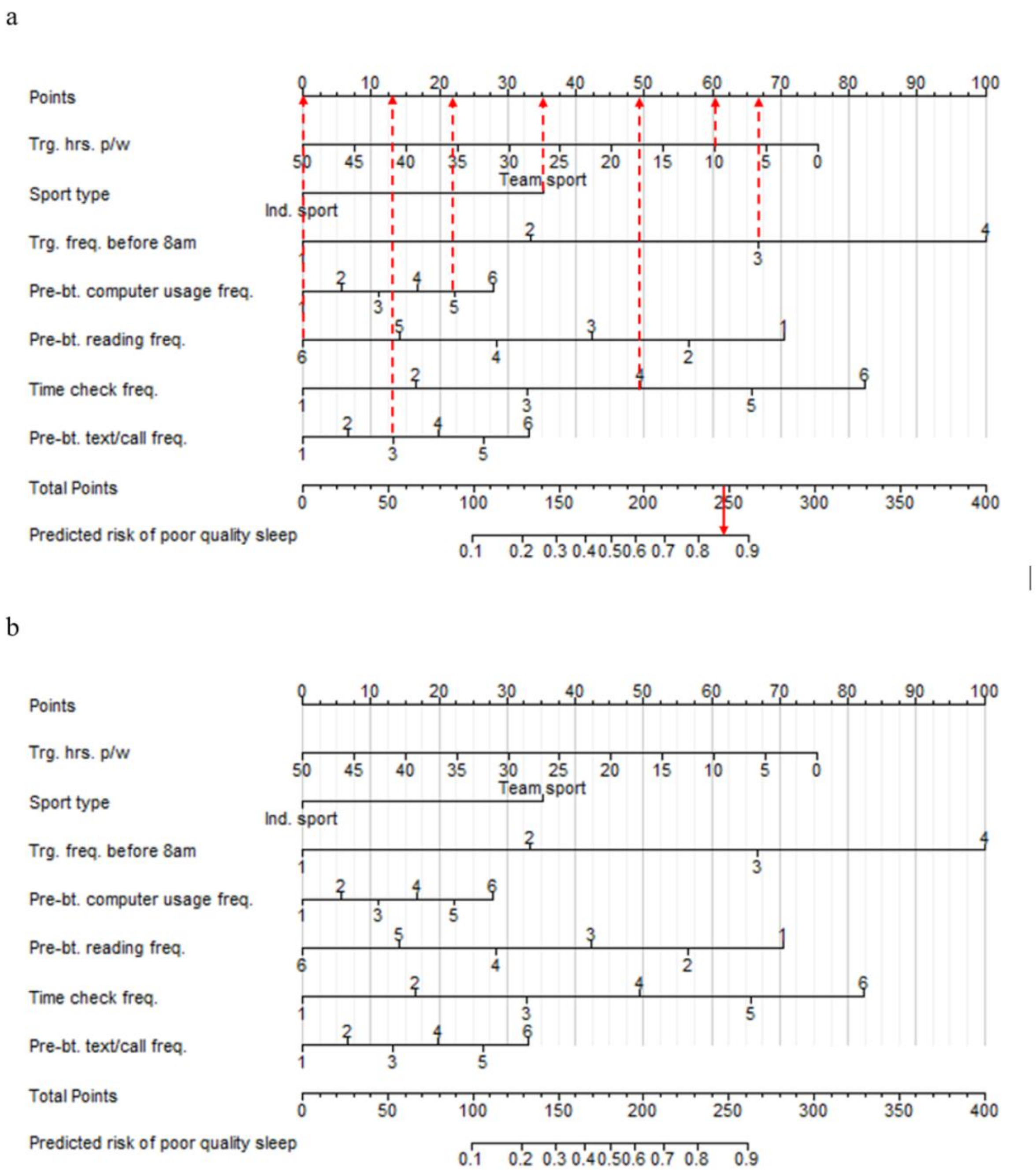

Figure 2. Nomogram for calculating the predicted risk of having poor quality sleep. (a) Case study scenario of an athlete that participated in 10 hours of training per week, from a team sport, that trained before 8 am 3-4 times a week, used a computer during or close to bedtime 5-6 times a week, read during or close to bedtime daily, sometimes checked the time on a digital clock and/or device after lights-out, and texted or called using a communication device 1-2 times a week. Extrapolating this information upwards to the points obtained for each of these

variables, such an individual would obtain a total number of points of approximately 249, indicating approximately an 85% risk of experiencing poor sleep quality as determined by the PSQI. (b) No case study annotations.

Abbreviations: Trg. = Training; Hrs. = Hours; Freq. = Frequency; Pre-bt. = Pre-bedtime

Table 1. Participant demographics by sport, Mean (SD)

Characteristic	**Individual Sport**, N = 57	**Team Sport**, N = 58
Age, yrs	14.58 (1.39)	16.57 (2.08)
Height, cm	167 (9)	169 (7)
Mass, kg	57 (8)	62 (12)
Gender, n (%)		
Female	25 (44)	37 (64)
Male	32 (56)	21 (36)

Table 2. LASSO regression coefficients

Variable	Standardized coefficients
Intercept	-0.59
Trg. hrs. p/w	-0.09
Sport type (individual vs. team)	-0.64
Trg. freq. before 8am	0.06
Pre-bt. Computer usage freq.	0.02
Pre-bt reading freq.	-0.02
Time check freq.	0.25
Pre-bt. Text/call freq.	0.10

Abbreviations: Trg. = Training; Hrs. = Hours; Freq. = Frequency; Pre-bt. = Pre-bedtime

Table 3. Comparison of performance models for training and test set

Model	Training set[a]			Test set		
	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
Full-feature	0.47 (0.19)	0.44 (0.27)	0.51 (0.23)	0.55	0.57	0.65
Feature-reduced	0.75 (0.18)	0.60 (0.27)	0.80 (0.20)	0.80	0.57	0.80

[a] Average results of 5-repeated 10-fold cross-validations

Supplementary material 1. Sleep Environment and Hygiene Survey