Sleep Medicine 114 (2024) 178–181

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Number of days required to reliably measure weekday and weekend total sleeping time with accelerometer: A secondary data analysis with National Health and Nutritional Survey (NHANES) 2011–2014 data

Paul H. Lee *

Southampton Clinical Trials Unit, University of Southampton, United Kingdom

ARTICLE INFO

Keywords:
Actigraphy
Measurement
Reliability
Sleep

ABSTRACT

The current standard practice for measuring sleeping time with accelerometer is to ask the participants to wear it for 7 consecutive days and analysing data from participants who have provided at least 4 days of valid data. However, this standard lacks supporting evidence. This study aims to evaluate this standard of practice by examining the reliability of measuring total sleeping time in a representative sample of US adults using accelerometer data from the National Health and Nutritional Survey (NHANES) waves 2011–2012 and 2013–2014. The sample included a total of 14,676 participants, out of which only those who provided data for seven days (n = 9510) were included in the analysis. The results revealed that the intra-class correlation coefficient (ICC) for a single day of measurement was 0.38 for weekdays and 0.27 for weekends. To achieve a reliability of 0.7, measurements for 4 and 7 nights were necessary for weekdays and weekends, respectively. Our simulation study found that the randomly-selected 3-day average of weekday sleeping time strongly correlated with the actual mean (ρ = 0.92), capturing at least 80 % of the variance. However, the randomly-selected 1-day average of weekend sleeping time only captured about 60 % of the variance. In conclusion, we recommend that future accelerometer research adopts a 9-day continuous measurement period, covering four weekend days, to reliably estimate both weekday and weekend sleeping time.

1. Introduction

The accelerometer, a device that measures sleep and physical activity by tracking wrist movement, is increasingly used in medical research due to its proven validity [1] and decreasing costs. While accelerometers have become the standard of measuring sleep [2] and physical activity level [3] at population-level research, certain standard of practices in their use lack evidence-based support. For example, many accelerometer studies rely on a one-week measurement period and analyse data from participants who provide at least 4 days of valid data, without considering the specific population or variables of interest [4]. The underlying assumption is that 4 days of accelerometer data can reliably represent the entire week, but this assumption is rarely tested.

Multiple studies have investigated the reliability, or internal consistency, of accelerometer-measured data to determine the minimum number of days required to achieve a reasonably high reliability (e.g., >0.7 or >0.8) for weekly data. A pioneering study conducted in 1999 focused on children and adolescents [5], and found that the single-day intra-class correlation coefficient (ICC) for most sleep parameters measured by the Mini Motionlogger was approximately 0.3. This study concluded that 5–6 days of data were necessary to achieve a reliability of 0.7 [5]. Subsequent studies conducted with different accelerometer brands and age groups consistently arrived at similar conclusions, recommending a measurement period of 5–6 days [6–8] for sleep variables.

One significant limitation of the studies mentioned above that examined the reliability of accelerometer-measured sleep parameters is their reliance on calculating single-day intra-class correlation coefficient (ICC) values and using the Spearman-Brown prophecy formula to determine the number of days needed to achieve specific reliability thresholds (usually 0.7, 0.8, and 0.9) [5–8]. This approach overlooks the differences in sleep patterns between weekdays and weekends. A population-based study found that the Japanese working population slept an average of 7.5 h on weekends and 6.5 h on weekdays [9]. Therefore, when estimating total sleep time and other sleep parameters

* Southampton Clinical Trials Unit, University of Southampton, Southampton General Hospital, Tremona Road, Southampton, Hampshire, SO16 6YD, United Kingdom.

E-mail address: paul.h.lee@southampton.ac.uk.

https://doi.org/10.1016/j.sleep.2024.01.006
Received 18 September 2023; Received in revised form 22 December 2023; Accepted 5 January 2024
Available online 6 January 2024
1389-9457/© 2024 The Author. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

across the entire week, it is essential to consider both weekday and weekend sleep duration separately. Additionally, the assumption of equal correlations between sleep parameters on weekdays and weekends when applying the Spearman-Brown prophecy formula was not adequately justified. Moreover, while most existing studies focus on children, there is a lack of reliability data for sleep parameters in adults.

In this study, we utilised the publicly available accelerometer data from the National Health and Nutritional Survey (NHANES) waves 2011–2012 and 2013–2014 to assess the reliability of total sleeping time in a representative sample of adults in the United States. To estimate the minimum number of days required to achieve acceptable reliabilities of 0.7 and 0.8, we applied the Spearman-Brown prophecy formula separately on weekday and weekend data. Additionally, we employed a simulation approach among participants who provided complete data for a full 7-day week. This allowed us to generate different scenarios by randomly removing available sleep data, specifically 1–4 days for weekdays and 1 day for the weekend. We then correlated the means calculated from total sleeping time with and without the randomly removed data to examine the impact of data availability on reliability.

2. Methods

2.1. Participants

The complete details of the NHANES recruitment procedure can be accessed on the NHANES official website at https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/overview.aspx?BeginYear=2011. For this study, we examined the accelerometer data from a total of 14,676 participants who were recruited in NHANES 2011–2012 and 2013–2014, and only included participants who provided seven valid days of data on accelerometer-measured total sleeping time (n = 9510) where a valid day was defined as a day with accelerometer-measured sleep duration between 3 and 12 h/day. Note that no data were removed due to valid hours as 96.6 % of the records had at least 20 valid hours. To maintain participant privacy and mitigate identification issues, NHANES assigned an age of 80 to all individuals aged 80 or older (n = 394).

2.2. Accelerometer-measured total sleeping time

The complete details of the accelerometer data processing procedure can be accessed on the NHANES official website at https://wwwn.cdc.gov/nchs/data/nhanes/2011-2012/manuals/2012-Physicial-Activity-Monitor-Procedures-Manual-508.pdf In summary, participants were instructed to wear an ActiGraph GT3X + accelerometer (details available at https://actigraphcorp.com/) on their non-dominant wrist during the day of the examination. They were then asked to continue wearing the accelerometer 24 h a day for 7 consecutive days, removing it on the morning of the 9th day. The accelerometer recorded acceleration data at a frequency of 80Hz, and the epoch length was set to 1 min.

To determine the total sleeping time, a machine learning algorithm was employed. This algorithm classified each recorded minute as wake, sleep, non-wear, or unknown, based on factors such as signal power, variance of orientation, and changes in orientation [10]. The total sleeping time for each day was calculated by summing the sleep minutes recorded during that particular day. Note that this study used 24-h sleeping time instead of nighttime sleeping time. The total sleeping time represented the sleep in the 24-h period, but not only the nighttime sleep of the participants.

2.3. Statistical analysis

Spearman correlation was used to examine the correlation between the total sleeping time across the seven days of a week (S/M/T/W/T/F/S). Given the observed weekday-weekend difference in total sleeping time within the dataset (as presented in Table 2), separate estimates were made for the total sleeping time on weekdays and weekends. The two-way mixed effect single measurement intra-class correlation coefficient (ICC) was calculated for the entire week's total sleeping time, as well as separately for weekdays and weekends. The single measurement ICC was used to determine the number of days required to achieve reliabilities of 0.7, 0.8, and 0.9 by applying the Spearman-Brown prophecy formula (number of days required $= \frac{\rho_r(1-\rho)}{\rho(1-\rho_r)}$, where ρ and ρ_r are the single measurement ICC and the target reliability, respectively) [11,12].

To assess the impact of the number of wearing days on the accuracy of total sleeping time estimation, we conducted simulations using different numbers of days (1/2/3/4 for weekday mean and 1 for weekend mean). For each participant, their sleeping time was calculated by taking the mean of randomly-sampled days from all available days. The randomisation process was repeated 1000 times to generate multiple samples. The accuracy of the four weekday means and weekend means was evaluated by comparing them against the actual weekday and weekend means, respectively. The actual means were calculated by taking the mean of five weekdays and two weekends for all participants. To evaluate the biasedness and precision of the average sleeping time calculated with sleep measurements of different number of days, we considered several metrics, including 1) mean of the means, 2) standard deviation (SD) of the means, 3) root-mean-square error, 4) mean of the correlations between the actual mean and estimated mean, 5) SD of the correlations, and 6) agreement of observations with the actual mean and the mean calculated from randomly-selected days belonged to the same category (<420 min, 420–479 min, 480–539 min, 540+ min). A sensitivity analysis was conducted to test the definition of weekend as Friday and Saturday sleep. These metrics were also reported across different subgroups across sex (male vs female), age (below or above 35 years old), and ethnicity (Non-Hispanic White/not Non-Hispanic White). The statistical analysis was performed using R version 4.2.0.

Table 1
Demographic characteristics of the participants (n = 9510).

Variable	Frequency	Percentage
Sex		
Male	4617	48.5
Female	4893	51.5
Race		
Mexican American	1500	15.8
Other Hispanic	932	9.8
Non-Hispanic White	3332	35.0
Non-Hispanic Black	2300	24.2
Others	1446	15.2
Education level		
Less than 9th grade	545	5.9
9–11th grade	812	8.9
High school graduate/GED or equivalent	1352	14.7
Some college or AA degree	1807	19.7
College graduate or above	1611	17.6
Participants below 20 years old	3030	31.9
Marital status[a]		
Married	3250	53.0
Widowed	512	8.3
Divorced	665	10.8
Separated	181	3.0
Never married	1096	17.9
Living with partners	425	6.9
Annual household income		
<$20,000	2013	22.0
$20,000-$44,999	3125	34.2
$45,000-$74,999	1555	17.0
$75,000-$99,999	835	9.1
>$100,000	1616	17.7
Variable	**Mean**	**SD**
Age	36.6	24.0

[a] Only participants above 20 years old were required to answer.

3. Results

Table 1 displays the demographic characteristics of the NHANES cycles 2011–2012 and 2013–2014 participants who provided 7 valid nights of accelerometer data (n = 9510). The mean (SD) age of the participants was 36.6 (24.0), and 48.5 % (n = 4617) were male. Approximately one-third of the participants were Non-Hispanic White (35 %, n = 3332).

Table 2 presents the descriptive statistics of the total sleeping time across the week. As mentioned in the Methods, we observed a weekday-weekend difference of total sleeping time. Specifically, the mean total sleeping time for weekdays ranged from 444.4 min to 461.8 min, while for weekends, it ranged from 473.0 min to 496.6 min. The correlations between total sleeping time across the different days were generally consistent, ranging from 0.24 to 0.43, with weekdays correlations (0.35–0.43) consistently higher than the other correlations. Therefore, in the subsequent simulation studies, we would use the mean total sleeping time of weekdays (454.6 min) and weekends (484.7 min) as the actual means.

Table 3 presents the single-item intra-class correlation coefficient (ICC) of total sleeping time and the number of measurement days required to achieve reliabilities of 0.7, 0.8, and 0.9. The results indicate that one week of accelerometer data collection is insufficient to achieve a reliability of 0.8. However, 7 days of data collection, which is equivalent to 3.5 weekends, are required to achieve a reliability of 0.7 for estimating the weekend sleeping time. In contrast, for estimating the weekday sleeping time, 4 days of accelerometer data are sufficient to achieve a reliability of 0.7.

Table 4 shows the simulation results for estimating total sleeping time based on the number of randomly-selected days. For the weekday means, regardless of the number of days selected (1/2/3/4), the estimated means were found to be unbiased. As anticipated, the accuracy of the estimation improved with an increased number of randomly-selected days. This improvement was observed in terms of the standard deviation (SD) of means, root-mean-square error, and correlation. The 3-day weekday mean correlated strongly with the actual mean (ρ = 0.92), in which we can expect at least 80 % (=0.92^2) of the variance of the actual weekday mean was captured. On the other hand, the 1-day weekend mean could only capture about 60 % (=0.78^2) of the variance actual weekend mean. The agreement of the 1-day weekday and weekend means were around 50 % (51.5 % and 54.5 %, respectively), and the agreement increased with the number of days used.

Results of the sensitivity analysis (Supplementary Table S1) show that the conclusions are similar regardless of the definition of weekend. The stratified simulation results are shown in Supplementary Tables S2–S7. The simulation results were consistent across all subgroups examined.

4. Discussion

To the best of our knowledge, this is the first study examining the reliability of accelerometer-measured sleeping time among a population-representative sample (NHANES 2011–2012 and 2013–2014), specifically separating weekday and weekend data. The ICCs of total sleeping time revealed that sleeping time was more consistent across the weekdays then weekends. We found that collecting accelerometer data for 4 weekdays achieved a reliability of 0.7 for weekday sleeping time. However, due to the lower ICC for weekend sleeping time (0.27), it was necessary to collect data for 7 weekend days to achieve the same reliability. This implies that the total accelerometer data collection period would need to span at least 22 days (3 weeks plus one weekend) to capture sufficient data. In large-scale studies where a large sample size is required, such an extended measurement period poses too much burden on participants. Therefore, we suggest using a 9-day measurement period that covers 4 weekend days and 5 weekdays. This duration yielded a reliability of 0.6 for estimating weekend sleeping time, which can provide a more feasible option. It is important to note that the reliabilities of sleep parameters were generally lower than those of physical activity parameters, which have been reported in the range of 0.30–0.45 in previous studies [6,7,13,14]. Therefore, our recommendation of a 9-day measurement period should allow for estimating habitual physical activity levels with a reliability of 0.7 or better.

The results of our simulation studies support a similar conclusion with Spearman-Brown prophecy formula. We found a strong correlation between the mean sleeping time calculated from randomly-selected 3 weekdays and the original 5 weekday mean, suggesting that as few as 3 valid sleep records may be sufficient for weekday sleeping time estimation. It is worth noting that if the goal of collecting accelerometer-measured total sleeping time is to estimate the mean total sleeping time of the population, measuring one weekday and one weekend day would be adequate. However, we did observe a 20-min difference in total sleeping time between Saturdays and Sundays. This raises the

Table 2
Spearman correlations and means of total sleeping time (in min) across the seven days of a week (n = 9510).

Correlation	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Sun	1	0.24	0.27	0.28	0.29	0.24	0.29
Mon		1	0.41	0.41	0.39	0.35	0.25
Tue			1	0.42	0.43	0.37	0.26
Wed				1	0.41	0.38	0.28
Thu					1	0.38	0.28
Fri						1	0.27
Sat							1
Mean	496.6	461.8	456.3	456.9	453.7	444.4	473.0
SD	103.4	99.2	96.6	96.4	95.2	99.1	103.7

Table 3
Intra-class correlation coefficients (ICCs) of total sleeping time throughout the week, the weekday, and the weekend (n = 9510).

		Number of days to achieve a reliability of		
Measure	ICC	0.7	0.8	0.9
Week	0.31	6	9	21
Weekday	0.38	4	7	15
Weekend	0.27	7	11	25

Table 4
Simulation results (number of simulations: 1000).

	Actual mean	Mean calculated from randomly-selected days			
		1	2	3	4
Weekday					
Mean of means	454.6	454.7	454.6	454.7	454.6
SD of means		0.73	0.43	0.28	0.18
Root-mean-square error		68.2	41.8	27.8	17.0
Mean of correlations between actual mean		0.71	0.85	0.92	0.97
SD of correlations		4.99×10^{-3}	2.68×10^{-3}	1.37×10^{-3}	5.76×10^{-4}
Agreement[a]		51.5 %	62.6 %	72.7 %	79.9 %
Weekend					
Means	484.7	484.8			
SD of means		0.65			
Root-mean-square error		62.9			
Mean of correlations between actual mean		0.78			
SD of correlations		2.89×10^{-3}			
Agreement[a]		54.5 %			

[a] Percentage of observations with the actual mean and the mean calculated from randomly-selected days belonged to the same category (<420 min, 420–479 min, 480–539 min, 540+ min).

question of whether there are habitual sleeping pattern differences between these two days. Unfortunately, we only collected data for one Saturday and one Sunday in the current sample, which prevented us from examining the Saturday-Sunday differences. Further research with a longer measurement period would be needed to explore this aspect.

This study is not without limitations. Firstly, only participants who provided 7 valid accelerometer sleeping records were included, which may introduce selection bias and limit the generalizability of the findings to the entire sample. However, a previous study demonstrated that altering the inclusion criteria had minimal impact on the sample intraclass correlation coefficient (ICC) [6], suggesting that this limitation may have had a minor effect on the results. Additionally, several factors that can influence the reliability of total sleeping time, such as seasonality and working hours, were not taken into account due to the lack of available data. Seasonal variations and job-related factors can impact sleep patterns, and their absence in the study could affect the generalizability of the findings. To address seasonality, longer-term accelerometer data collection (e.g., over the course of a year) would be necessary to capture the potential fluctuations in sleep patterns.

In conclusion, this study aimed to determine the optimal number of days required to accurately estimate total sleeping time in a week using accelerometers. Based on the findings, we recommend that future accelerometer research adopts a 9-day continuous measurement period that includes four weekend days. This extended duration will provide more reliable estimates of weekend sleeping time. By addressing the limitations of previous studies and considering the weekday-weekend differences in sleep patterns, we believe this 9-day data collection approach can enhance the accuracy and representativeness of accelerometer-measured sleep data. Further research in this area is warranted to validate and refine these recommendations in other populations.

Funding

None.

Ethics approval

This study was approved by the NCHS Research Ethics Review Board (ERB) (Protocol #2011–17).

Consent to participate

All participants consented to participate.

Data availability

The dataset(s) supporting the conclusions of this article is(are) available in the NHANES website (https://wwwn.cdc.gov/nchs/nhanes/).

CRediT authorship contribution statement

Paul H. Lee: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sleep.2024.01.006.

References

[1] van Hees VT, Sabia S, Jones SE, Wood AR, Anderson KN, Kivimäki M, et al. Estimating sleep parameters using an accelerometer without sleep diary. Sci Rep 2018;8:12975.
[2] Full KM, Kerr J, Grandner MA, Malhotra A, Moran K, Godoble S, et al. Validation of a physical activity accelerometer device worn on the hip and wrist against polysomnography. Sleep Health 2018;4:209–16.
[3] Lee I-M, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. Br J Sports Med 2014;48: 197–201.
[4] Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nystrom C, Mora-Gonzalez J, Lof M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic Review and practical considerations. Sports Med 2017;47:1821–45.
[5] Acebo C, Sadeh A, Seifer R, Tzischinsky O, Wolfson AR, Hafer A, et al. Estimating sleep patterns with activity monitoring in children and adolescents: how many nights are necessary for reliable measures? Sleep 1999;22:95–103.
[6] Antczak D, Lonsdale C, Del Pozo Cruz B, Parker P, Sanders T. Reliability of GENEActiv accelerometers to estimate sleep, physical activity, and sedentary time in children. Int J Behav Nutr Phys Activ 2021;18:73.
[7] Ridgers ND, Hnatiuk JA, Vincent GE, Timperio A, Barnett LM, Salmon J. How many days of monitoring are needed to reliably assess SenseWear Armband outcomes in primary school-aged children? J Sci Med Sport 2016;19:999–1003.
[8] Taylor RW, Williams SM, Farmer VL, Taylor BJ. The stability of sleep patterns in children 3 to 7 years of age. J Pediatr 2015;166:697–702 e1.
[9] Adachi H, Yamamoto R, Fujino R, Kanayama D, Sakagami Y, Akamine S, et al. Association of weekday-to-weekend sleep differences and stress response among a Japanese working population: a cross-sectional study. Sleep Med 2021;82:159–64.
[10] John D, Tang Q, Albinali F, Intille S. An open-source monitor-independent movement summary for accelerometer data processing. J Measure Phys Behav 2019;2:268–81.
[11] Brown W. Some experimental results in the correlation of mental abilities. British Journal of Psychology, 1904-1920 1910;3:296–322.
[12] Spearman C. Correlation calculated from faulty data. British Journal of Psychology, 1904-1920 1910;3:271–95.
[13] Aadland E, Ylvisaker E. Reliability of objectively measured sedentary time and physical activity in adults. PLoS One 2015;10:e0133296.
[14] Dillon CB, Fitzgerald AP, Kearney PM, Perry IJ, Rennie KL, Kozarski R, et al. Number of days required to estimate habitual activity using wrist-worn GENEActiv accelerometer: a cross-sectional study. PLoS One 2016;11:e0109913.