Sleep Medicine 63 (2019) 127–136

Contents lists available at ScienceDirect

Sleep Medicine

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

Original Article

Validity of two retrospective questionnaire versions of the Consensus Sleep Diary: the whole week and split week Self-Assessment of Sleep Surveys

Jessica R. Dietch, Kevin Sethi [1], Danica C. Slavish, Daniel J. Taylor*

Department of Psychology, University of North Texas, 1155 Union Circle #311280, Denton, TX, 76203, USA

A R T I C L E I N F O

Article history:
Received 11 January 2019
Received in revised form
20 May 2019
Accepted 28 May 2019
Available online 12 June 2019

Keywords:
Sleep diary
Survey
College students
Validity

A B S T R A C T

Objective/Background: Prospective, daily sleep diaries are the gold standard for assessing subjective sleep but are not always feasible for cross-sectional or epidemiological studies. The current study examined psychometric properties of two retrospective questionnaire versions of the Consensus Sleep Diary.
Participants/Methods: College students ($N = 131$, mean age $= 19.39 \pm 1.65$; 73% female) completed seven days of prospective sleep diaries then were randomly assigned to complete either the Self-Assessment of Sleep Survey (SASS), which assessed past week sleep ($n = 70$), or the SASS-Split (SASS-Y), which assessed weekday/weekend sleep separately ($n = 61$). Participants also completed psychosocial/sleep questionnaires including the Pittsburgh Sleep Quality Index (PSQI). Sleep parameters derived from SASS, SASS-Y, PSQI, and sleep diaries were assessed via Bland Altman plots, limits of agreement, mean differences, and correlations.
Results: SASS-Y demonstrated stronger correlations with prospective sleep diaries and slightly less biased estimates ($r = 0.51$ to 0.85, $\alpha = -0.43$ to 1.70) compared to SASS ($r = 0.29$ to 0.84, $\alpha = -1.63$ to 2.33) for terminal wakefulness (TWAK), sleep onset latency (SOL), sleep efficiency (SE), and quality (QUAL). SASS resulted in slightly less bias for total sleep time (TST) and wake after sleep onset (WASO) ($\alpha = -0.65$ and 0.93, respectively) compared to SASS-Y ($\alpha = 14.90$ and 1.05, respectively). SASS and SASS-Y demonstrated greater convergence with sleep diary than PSQI.
Conclusions: Results demonstrated good psychometric properties for the SASS and SASS-Y. When prospective sleep diaries are not feasible, the SASS and SASS-Y are acceptable substitutes to retrospectively estimate sleep parameters. Retrospective estimation of sleep parameters separately for weekdays/weekends may offer advantages compared to whole week estimation.

1. Introduction

Prospective, daily sleep diaries are the gold standard for assessing subjective sleep, but sleep diaries are not always feasible for cross-sectional or large epidemiological studies. Instead, brief retrospective self-report sleep questionnaires are the most commonly used measures in health research. However, sleep questionnaires suffer some substantial limitations [1]. Existing questionnaires do not yield all of the sleep parameters typically produced by sleep diaries and objective measures of sleep, and literature comparing the validity of sleep questionnaires to sleep diaries is limited [2]. Further, existing questionnaires typically fail to separate weekday and weekend sleep despite known discrepancies in sleep between these schedules [2,3]. Given these measurement barriers, we sought to develop and validate a brief retrospective self-report questionnaire that addresses these limitations and provides an accurate measure of subjective sleep parameters.

Sleep diaries are typically used in addition to or in place of objective measures (ie, polysomnography [PSG] or actigraphy) [4]. Sleep diaries are typically prospective (ie, completed over multiple time-points) self-report measures in which patients and participants record their sleep patterns and answer other questions related to their sleep on a daily basis (eg, sleep quality, daytime

* Corresponding author. Department of Psychology, University of Arizona, 1503 E University Blvd., Tucson, AZ, 85721, USA. Fax: +1 (520) 621-9306.
E-mail addresses: Jessica.Dietch@va.gov (J.R. Dietch), Kevin.Sethi@va.gov (K. Sethi), Danica.Slavish@unt.edu (D.C. Slavish), danieljtaylor@email.arizona.edu (D.J. Taylor).
[1] Present address: Psychology Service (116A), South Texas Veterans Health Care System, 7400 Merton Minter, San Antonio, TX, 78229, USA.

https://doi.org/10.1016/j.sleep.2019.05.015
1389-9457/

sleepiness, stimulant use). Sleep diaries capture night-to-night variability in sleep, which a growing literature suggests is associated with poor health outcomes [5]. Sleep diaries completed on a single night have been shown to correspond well with both PSG and actigraphy, although studies have found people overestimate sleep onset latency (SOL) and wake time after sleep onset (WASO) on diaries [6–10]. Recently, a group of insomnia experts developed a standardized sleep diary called the Consensus Sleep Diary and recommended it for widespread use [4].

Sleep diaries provide some challenges, particularly for epidemiologic or other large-scale research. First, it is not always possible to ensure diaries are completed daily, particularly when administered via paper-and-pencil. If participants fail to complete the diary daily and attempt to fill in the diary retrospectively, the advantage of the diary over single-time point retrospective measures is diminished [11]. Second, diaries necessitate a longitudinal study design in which participants are first instructed on use of the sleep diary and then must return the diary at a later date. This format may not be feasible for some study designs (eg, large-scale epidemiological research). Finally, sleep diaries represent substantial participant burden in comparison to brief questionnaires (eg, seven times as many questions or items). Although electronic data capture can mitigate some of these concerns, it does not entirely reduce the burden or challenges posed by typical sleep diary data collection.

Retrospective self-report questionnaires are inexpensive and can be quickly administered, which makes them ideal for screening patients and conducting epidemiological research studies [12] but existing measures demonstrate considerable limitations. For example, for the most widely used measure of sleep quality, the Pittsburgh Sleep Quality Index (PSQI) [13], the method for calculation of quantitative sleep variables (eg, total sleep time [TST], sleep efficiency [SE]) is not consistent with current expert opinion [1,4]. Similarly, the Sleep Timing Questionnaire (STQ; [14]) and the Sleep Habits Survey [15,16] do not yield habitual estimates of important sleep parameters that are essential for researchers of insomnia and other similar sleep disorders (eg, SOL, WASO). In order to facilitate comparison with existing sleep research conducted using sleep diaries and objective measures of sleep, it is important that retrospective self-report questionnaires are able to obtain accurate estimates of multiple facets of sleep (eg, timing, duration, efficiency) that are easy to compare to other measures of sleep (eg, actigraphy).

Furthermore, the validity of retrospective self-report questionnaires may be poor, particularly in populations with greater night-to-night variability in sleep [17]. When people are asked to recall behaviors that are rare or important, they may be able to provide relatively accurate retrospective reports [18]. However, for frequent behaviors (eg, sleep) people are unlikely to be able to recall specific details of the many sleep episodes. Rather, they must rely upon estimation and inference strategies to provide reports, which is subject to error and reporting biases particularly when the behaviors are more irregular (eg, in people with insomnia; [18]). People may also round estimates to the nearest hour, or report desired times rather than actual times when reporting morning wake-times, which may further decrease the accuracy of self-reports [14].

Research on autobiographical memory suggests retrospective reports can be improved by encouraging a cued retrieval process that takes advantage of the hierarchical structure of memory [19,20]. Researchers may be able to garner more accurate responses by prompting participants to consider separate periods of time or themes for which the behavior may have varied [19,20]. For example, a researcher examining sleep in college students may choose to ask participants to recall sleep behaviors separately for the summer and school year, and further delineate between sleep on exam weeks versus sleep on non-exam weeks, allowing multiple cues to aid in the retrieval of memories. This might result in more accurate responses than asking them to recall sleep behaviors throughout the entire year. For epidemiological sleep research, measures may be improved by simply incorporating thematic variations in sleep behaviors that typically occur in the general population such as weekdays/work days and weekends/off days. Indeed, in a nationally representative study of 1040 adults by Lauderdale [2], assessing sleep duration using a single question versus two questions (weekday and weekend, weighted) yielded a significant (15-min) difference in sleep duration estimates. Taken together, this evidence indicates querying weekday and weekend sleep separately may enhance accuracy of retrospective estimates of sleep.

The goals of the current study were to: (1) validate a retrospective questionnaire version of the Consensus Sleep Diary [4] compared to typically-administered prospective sleep diaries, and (2) experimentally examine whether retrospectively measuring weekday and weekend sleep separately using a version of the Consensus Sleep Diary offers improvements over retrospectively measuring sleep over the entire week (ie, the current standard in epidemiological research).

2. Method

2.1. Participants

In sum, 168 participants were recruited from undergraduate psychology courses at the University of North Texas during the Fall semester of 2013. Inclusion criteria were that students needed to be (a) between ages 18–28, and (b) enrolled in an undergraduate psychology course at the University of North Texas. Participants were recruited and offered extra credit through SONA Systems, a web-based human subject pool management software. Thirty-seven participants were ultimately excluded for the following reasons: poor adherence to the sleep diary/sleep questionnaire (ie, <5 days of sleep diary completed; $n = 24$) or age out of range/unreported ($n = 13$). Therefore, a total of 131 participants were included in the analyses. See Table 1 for participant characteristics.

2.2. Procedures

Participants reported to the sleep lab to collect sleep diary and instructions. They were instructed to complete the diary each morning as soon as possible after waking for seven days. Later, they were to return to the lab on the same day of the week and approximate time as their original visit. Participants were then randomly assigned to either the whole week survey or split week survey condition. Simple randomization was performed via a random number generator and assigned based on the order in which participants returned their sleep diary. Participants in both conditions then completed the corresponding sleep questionnaire in addition to the demographics and additional psychosocial measures including the PSQI. All procedures were approved by the University of North Texas Institutional Review Board.

2.3. Measures

2.3.1. Demographics

In order to gain descriptive information about the sample, participants were administered a psychosocial questionnaire including demographic and work information (eg, age, class rank, gender, race, ethnicity, marital status, employment status, work schedule, and whether they worked or had class on a weekend day).

Table 1
Participant characteristics.

	Entire Sample Mean or n (N = 131)	Entire Sample SD or % (N = 131)	SASS Sample Mean or n (n = 71)	SASS Sample SD or % (n = 71)	SASS-Y Sample Mean or n (n = 60)	SASS-Y Sample SD or % (n = 60)	t-value or χ^2 value	p-value
Age	19.39	1.65	19.38	1.60	19.40	1.72	−0.07	0.95
Class Rank							2.7	0.62
Freshman	62	47.3%	33	46.5%	29	48.3%		
Sophomore	28	21.4%	14	19.7%	14	23.3%		
Junior	16	12.2%	11	15.5%	5	8.3%		
Senior	21	16.0%	10	14.1%	11	18.3%		
Not reported	4	3.1%	3	4.2%	1	1.7%		
Gender							0.15	0.93
Male	33	25.2%	17	23.9%	16	26.7%		
Female	96	73.3%	53	74.6%	43	71.7%		
Race							3.66	0.60
White	66	50.4%	35	49.3%	31	51.7%		
Black or African American	27	20.6%	14	19.7%	13	21.7%		
Asian	5	3.8%	2	2.8%	3	5.0%		
American Indian/ Alaska Native	2	1.5%	2	2.8%	0	0%		
More than one race	13	9.9%	6	8.5%	7	11.7%		
Unknown or not reported	18	13.7%	12	16.9%	6	10.0%		
Ethnicity							3.88	0.14
Hispanic/Latino	42	32.1%	28	39.4%	14	23.3%		
Not Hispanic/Latino	79	60.3%	38	53.5%	41	68.3%		
Not reported	10	7.6%	5	7.0%	5	8.3%		
Marital Status							0.05	0.98
Married	2	1.5%	1	1.4%	1	1.7%		
Unmarried	125	95.4%	68	95.8%	57	95.0%		
Not reported	4	3.1%	2	2.8%	2	3.3%		
Employment Status							3.28	0.35
Full-time	5	3.8%	1	1.4%	4	6.7%		
Part-time	57	43.5%	31	43.7%	26	43.3%		
Unemployed	68	51.9%	38	53.5%	30	50.0%		
Not reported	1	0.8%	1	1.4%	0	0%		
Work schedule							4.22	0.38
Daytime	29	22.1%	11	15.5%	18	30.0%		
Nighttime	17	13.0%	10	14.1%	7	11.7%		
Rotating shift	11	8.4%	7	5.6%	4	6.7%		
Worked/had class on a weekend day	46	35.1%	23	32.4%	23	38.3%	1.88	0.06

SASS = whole week measure, SASS-Y = split-week measure.

2.3.2. Sleep diary

Each morning, participants were instructed to record information about their previous night's sleep using the Core Consensus Sleep Diary [4]. These questions allowed for the calculation of the following quantitative variables related to sleep: Total time in bed (TIB), SOL, number of awakenings (NWAK), WASO, terminal wakefulness (TWAK), TST and SE (ie, TST divided by TIB ×100). Participants also rated the "Quality" (QUAL) of their sleep on a 5-point Likert-type scale (1 = poor to 5 = very good). Participants were instructed to leave a sleep diary blank for the day if they forgot to fill it out, rather than try to fill it in the next day. Sleep variables were calculated for the entire week by creating a 5:2 weighted average using available weekdays and weekends. Lastly, participants had the opportunity to provide additional information that may be relevant to their sleep in a free response comment box (eg, "I have a cold"). An additional question was provided on the sleep diary which asks participants to indicate each night whether the next day is a day off, or a workday, as well as when they are scheduled to start class or work that day. Out of seven possible sleep diaries, participants completed and average of 6.73 ± 0.51 for a compliance rate of 96%. Of the total sample, 76% completed all seven days of sleep diaries, 21% completed six days, and 3% completed five days.

2.3.3. Self-Assessment of Sleep Surveys

One of two different retrospective versions of the Consensus Sleep Diary questionnaire was administered to participants, depending on the condition to which they were randomly assigned. The whole week version of the Self-Assessment of Sleep Survey SASS (Supplementary Material 1) asked participants to complete questions recalling their "average" sleep habits "during the previous week." The split week version of the SASS-Y (Supplementary Material 2) asked participants to complete questions twice, once recalling how they slept "on the weekend during the previous week (Friday night through Sunday morning)" and once recalling how they slept "on weekdays during the previous week (Sunday night through Friday morning)." For the SASS-Y, variables were averaged separately for weekend sleep and weekday sleep, and then a weighted (ie, 5:2 weekday to weekend) average for the entire week was calculated. These questionnaires were developed by adding the phrase "on average" to the question stems from the Consensus Sleep Diary and including survey instructions that direct participants to consider a specific period of time (eg, for the SASS, "…during the previous week"). For the SASS-Y, the questions were administered twice with separate instructions referring to the weekdays and weekends.

2.3.4. Pittsburgh Sleep Quality Index

The Pittsburgh Sleep Quality Index (PSQI) is a commonly used retrospective questionnaire that assesses subjective sleep over the past month. The PSQI is a reliable and valid instrument [13] and has been validated in the college student population [21]. It demonstrated high test-retest reliability, high internal consistency (coefficient $\alpha = 0.83$), and high diagnostic sensitivity for distinguishing between good and poor sleepers ($\kappa = 0.75$, $p < 0.001$) [13]. The PSQI includes items that produce a global score and seven component scores, some of which overlap with sleep diary-generated parameters: QUAL, SOL, TST, and SE. Coefficient α was 0.56 in the current study.

2.4. Statistical analyses

The following analyses were conducted in order to establish the convergent validity of the SASS and SASS-Y via comparison with sleep diaries. Prospective sleep diaries (5:2 weighted) served as the gold standard against which the SASS, SASS-Y and PSQI were compared on six sleep parameters (ie, TWAK, SOL, WASO, TST, SE, and QUAL).

2.4.1. Bland-Altman plots and limits of agreement

Sleep parameters produced from the SASS, SASS-Y, and PSQI were compared with sleep diaries using the Bland and Altman technique to examine systematic bias and agreement [22]. One deficit in the sleep measure validation literature is the erroneous use of product-moment correlation coefficients (r) and other global indices to demonstrate agreement between two measures. Correlation coefficients are inadequate for assessing agreement for four primary reasons: (a) high correlation between two methods indicates the strength of a relationship instead of the agreement between the methods; (b) a difference in scale of measurement affects agreement but not correlation (eg, a parameter measured consistently lower by one method [eg, 400 min vs 500 min] could be perfectly correlated with a gold standard even though the two measures never agree), and therefore data that has low agreement can still demonstrate high correlations; (c) correlation fluctuates depending on the range in the sample such that a wide-ranging parameter will inherently have a greater correlation than a narrow-ranging parameter; and (d) significance testing is irrelevant in testing the strength of agreement, as it is expected that two measures would have a certain degree of relationship [22]. Bland and Altman [23] stated, although frequently done, it is inappropriate to assess agreement between measures using correlation, regression, comparison of means, structural equations, or intraclass correlation methods. Instead, plots of means against mean differences (Bland-Altman plots) and estimates of where 95% of differences between measures are expected to fall (limits of agreement) are suggested, as these items give information as to potential systematic bias and variability of estimates in addition to mean differences. We defined satisfactory agreement if differences were smaller than 30 min for SOL, WASO, and TWAK, 60 min for TST, 15% for SE, and 1 point for QUAL. We selected these limits of agreement based on a combination of examples set by previous literature [24,25] and on the basis of our clinical experience. For TST in particular, we selected 60 min due to (1) the common practice of categorizing TST by hour increments on retrospective sleep surveys [26] and (2) the common use of 60 min as a unit for altering sleep schedules to produce noticeable change in performance and health outcomes [27,28,29]. For SE, we selected more generous limits of agreement compared to previous literature (eg, 5%) due to: (1) the number of subjectively estimated variables that are included in the SE calculation, making it more susceptible to additive bias and (2) our clinical experience working with people with insomnia, the most likely disorder to display discrepancies in SE. Notably, few studies exist that have developed precedent for setting acceptable limits of agreement for sleep parameters in adults.

Bland-Altman plots were created to graph the mean of either SASS, SASS-Y, and PSQI measures and the weighted sleep diary measures (x-axis) versus the difference between the two measures (y-axis; [30]). In the plots, the 95% limits of agreement are displayed as the two outer lines and signify that 95% of differences between the two measures (ie, Two standard deviations above and below the mean) are expected to fall between these lines. Mean difference, or bias, is represented as the middle line on these plots, with positive values indicating overestimation by either SASS, SASS-Y, or PSQI measures compared to sleep diaries, and negative values indicating underestimation of either SASS, SASS-Y, or PSQI measures compared to sleep diaries. All analyses were conducted in R [version 1.1.383; [30]]. Bland-Altman analyses were conducted using the BlandAltmanLeh package [31].

2.4.2. Mean differences and correlations

Means on all sleep parameters were compared using paired samples t-tests and bivariate correlations in order to examine the degree and strength of the relationship between SASS, SASS-Y, PSQI, and sleep diary. Spearman correlations were used for TWAK, SOL, and WASO, as these variables were not normally distributed, and Pearson correlations were used for TST, TWAK, QUAL, as these measures were normally distributed.

3. Results

Participant characteristics are displayed in Table 1. There were no significant differences in demographic characteristics between groups.

3.1. Bland-Altman analyses

Mean difference (ie, bias or α), standard deviation of differences (ie, precision or σ), and 95% limits of agreement [23] for all sleep parameters are presented in Table 2. Inspection of these values and Bland-Altman plots (see Figs. 1–6) indicated the SASS-Y displayed slightly less bias than the SASS for estimating TWAK, SOL, SE, and QUAL, whereas the SASS displayed less bias than the SASS-Y for estimating WASO and TST. The SASS-Y questionnaire demonstrated the best (ie, lowest) precision across all sleep parameters resulting in the narrowest limits of agreement (ie, 95% of differences are expected to fall in a narrower range). This effect was most pronounced for TST, in which 95% limits of agreement for the SASS-Y spanned a range of 2.9 h compared to SASS (5.3 h) and PSQI (3.5 h).

There appeared to be a significant positive slope for SASS, SASS-Y, and PSQI measures of TWAK, SOL, TST, SE, and WASO, indicating non-constant bias (ie, greater mean values of each of these parameters tended to be related to greater overestimation by SASS, SASS-Y, and PSQI questionnaires compared to the weighted sleep diaries). There also appeared to be substantial heteroscedasticity for the SASS, SASS-Y, and PSQI questionnaires for all sleep parameters (ie, very high or low mean values of these parameters tended to be related to greater variability in bias between the SASS, SASS-Y, or PSQI questionnaires and the weighted sleep diaries).

Regarding satisfactory agreement with sleep diary, see the "Range" column in Table 2. For the SASS, satisfactory agreement was met for SOL, SE, and QUAL but not for TWAK, TST, or WASO. For the SASS-Y, satisfactory agreement was met for SOL, WASO, SE, and QUAL (and TWAK was <1 min outside of range), but not for TST. For the PSQI, satisfactory agreement was met for SOL and QUAL, but not for TST or SE.

Table 2
Bias, precision, and 95% limits of agreement for SASS, SASS-Y, and Pittsburgh sleep quality index sleep parameters compared to 5:2 weighted consensus sleep diary sleep parameters.

Sleep parameter	α	Σ	LoA lower	LoA upper	Range
SASS					
TWAK (min)	−1.63	18.65	−38.92	35.67	±37.30
SOL (min)	2.33	12.72	−22.61	27.26	±24.94
WASO (min)	0.93	16.27	−31.61	33.47	±32.54
TST (min)	−0.65	79.44	−159.54	158.23	±158.89
SE (%)	−0.62	6.83	−14.28	13.03	±13.66
QUAL (1 = poor, 5 = excellent)	−0.09	0.48	−1.03	0.84	±0.94
SASS-Y					
TWAK (min)	−0.43	15.33	−31.08	30.22	±30.65
SOL (min)	1.70	9.73	−17.76	21.17	±19.47
WASO (min)	1.05	9.81	−18.57	20.67	±19.62
TST (min)	14.90	43.27	−71.65	101.44	±86.55
SE (%)	−0.35	4.73	−9.81	9.11	±9.46
QUAL (1 = poor, 5 = excellent)	0.01	0.34	−0.67	0.69	±0.68
PSQI					
SOL (min)	−0.07	11.87	−23.34	23.20	±23.27
TST (min)	−16.19	52.26	−120.71	88.33	±104.17
SE (%)	−4.17	13.37	−30.38	22.03	±26.21
QUAL (1 = very bad, 4 = very good)	−0.65	0.45	−1.43	0.13	±0.78

Note. LoA = limits of agreement, PSQI = Pittsburgh Sleep Quality Index, SASS = whole week measure, SASS-Y = split-week measure, SD = 5:2 weighted average sleep diary, TWAK = terminal wakefulness, SOL = sleep onset latency, WASO = wake after sleep onset, TST = total sleep time, SE = sleep efficiency, QUAL = sleep quality, α = bias (mean difference), σ = precision (standard deviation of mean difference). $N = 131$ for sleep diary and PSQI measures; $n = 71$ for SASS; $n = 60$ for SASS-Y. Negative α′s indicate underestimation by SASS/SASS-Y/PSQI measures compared to sleep diaries, and positive α′s indicate overestimation by SASS/SASS-Y/PSQI measures compared to sleep diaries.

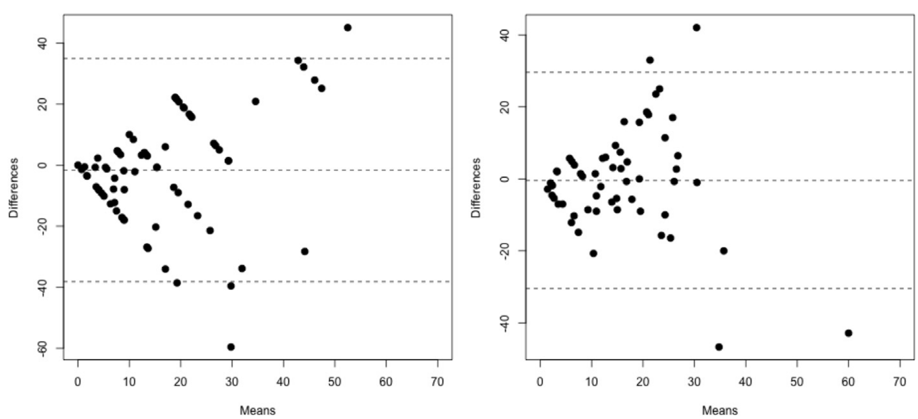

Fig. 1. Bland-Altman plots for time spent in bed after the final awakening (TWAK); SASS to sleep diary (left) and SASS-Y to sleep diary (right). Note. Bland-Altman plot comparing agreement of either whole week (SASS; left) or split week (SASS-Y; right) measurement of time spent in bed after the final awakening (TWAK; in minutes) to sleep diary measurement of TWAK. The x-axis represents the mean between either SASS or SASS-Y TWAK and sleep diary TWAK, and the y-axis represents the difference between SASS or SASS-Y TWAK and sleep diary TWAK. Each point on the graph represents the data from one participant. The center line is the mean difference between the two measures, and the outer dashed lines are 95% CIs. On the y-axis, positive values indicate overestimation by the SASS or SASS-Y, and negative values represent underestimation by the SASS or SASS-Y. (The PSQI does not include equivalent measures of TWAK).

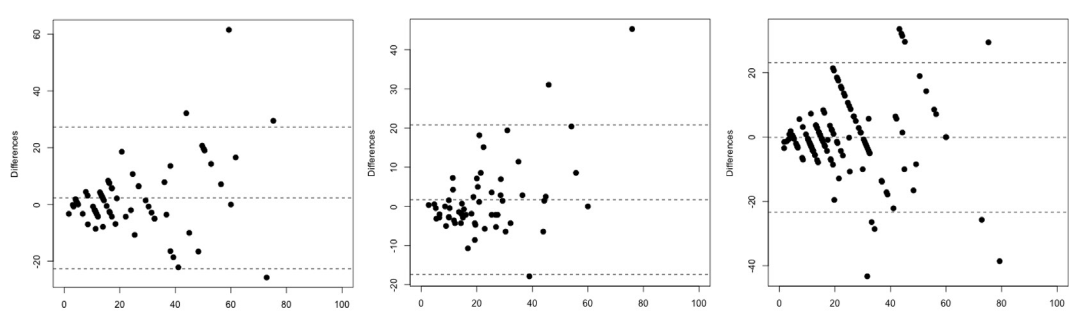

Fig. 2. Bland-Altman plots for sleep onset latency (SOL); SASS to sleep diary (left), SASS-Y to sleep diary (middle), PSQI to sleep diary (right). Note. Bland-Altman plot comparing agreement of either whole week (SASS; left), split week (SASS-Y; middle) or Pittsburgh Sleep Quality Index (PSQI; right) measurement of sleep onset latency (SOL; in minutes) to sleep diary measurement of SOL. The x-axis represents the mean between SASS, SASS-Y, or PSQI SOL and sleep diary SOL, and the y-axis represents the difference between SASS, SASS-Y, or PSQI SOL and sleep diary SOL. Each point on the graph represents the data from one participant. The center line is the mean difference between the two measures, and the outer dashed lines are 95% CIs. On the y-axis, positive values indicate overestimation by the SASS, SASS-Y, or PSQI measures, and negative values represent underestimation by the SASS, SASS-Y, or PSQI measures.

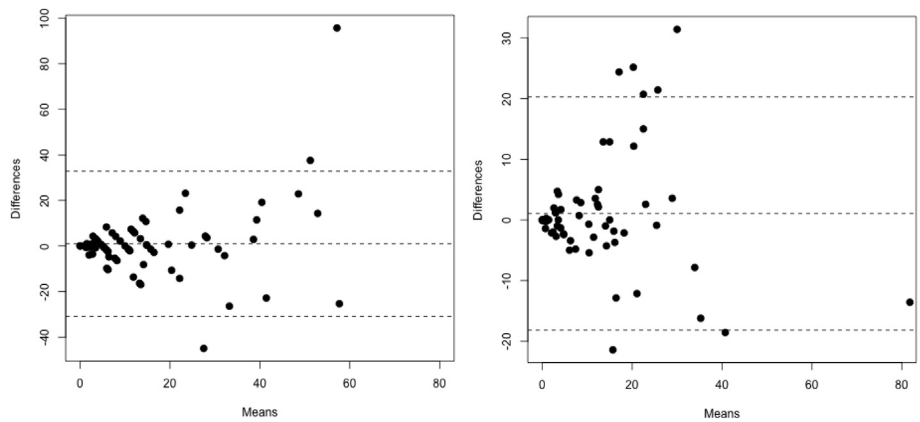

Fig. 3. Bland-Altman plots for wake after sleep onset (WASO); SASS to sleep diary (left) and SASS-Y to sleep diary (right). Note. Bland-Altman plot comparing agreement of either whole week (SASS; left) or split week (SASS-Y; right) measurement of wake after sleep onset (WASO in minutes) to sleep diary measurement of WASO. The x-axis represents the mean between either SASS or SASS-Y WASO and sleep diary WASO, and the y-axis represents the difference between SASS or SASS-Y WASO and sleep diary WASO. Each point on the graph represents the data from one participant. The center line is the mean difference between the two measures, and the outer dashed lines are 95% CIs. On the y-axis, positive values indicate overestimation by the SASS or SASS-Y, and negative values represent underestimation by the SASS or SASS-Y. (The PSQI does not include equivalent measures of WASO).

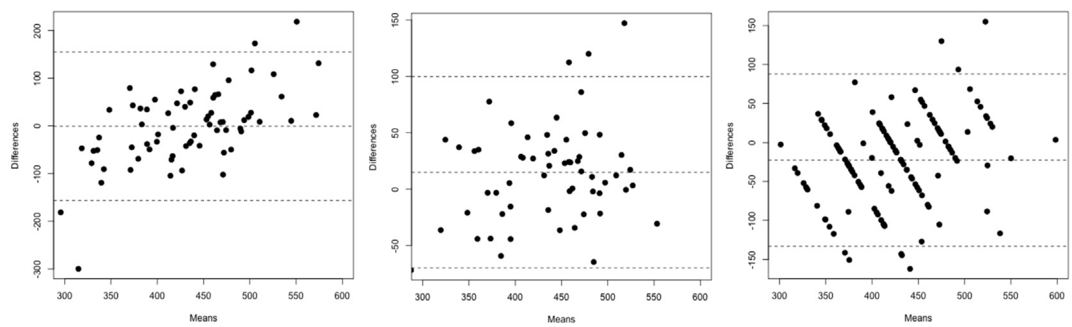

Fig. 4. Bland-Altman plots for total sleep time (TST); SASS to sleep diary (left), SASS-Y to sleep diary (middle), PSQI to sleep diary (right). Note. Bland-Altman plot comparing agreement of either whole week (SASS; left), split week (SASS-Y; middle) or Pittsburgh Sleep Quality Index (PSQI; right) measurement of total sleep time (TST; in minutes) to sleep diary measurement of TST. The x-axis represents the mean between SASS, SASS-Y, or PSQI TST and sleep diary TST, and the y-axis represents the difference between SASS, SASS-Y, or PSQI TST and sleep diary TST. Each point on the graph represents the data from one participant. The center line is the mean difference between the two measures, and the outer dashed lines are 95% CIs. On the y-axis, positive values indicate overestimation by the SASS, SASS-Y, or PSQI measures, and negative values represent underestimation by the SASS, SASS-Y, or PSQI measures.

3.2. Mean differences and correlations

As shown in Table 3, there were no significant mean differences between the SASS and the weighted sleep diaries on any of the sleep parameters (ie, TWAK, SOL, WASO, TST, SE, or QUAL). There were significant mean differences between the SASS-Y and the weighted sleep diaries on TST, but not on TWAK, SOL, WASO, SE or QUAL. There were significant mean differences between the PSQI

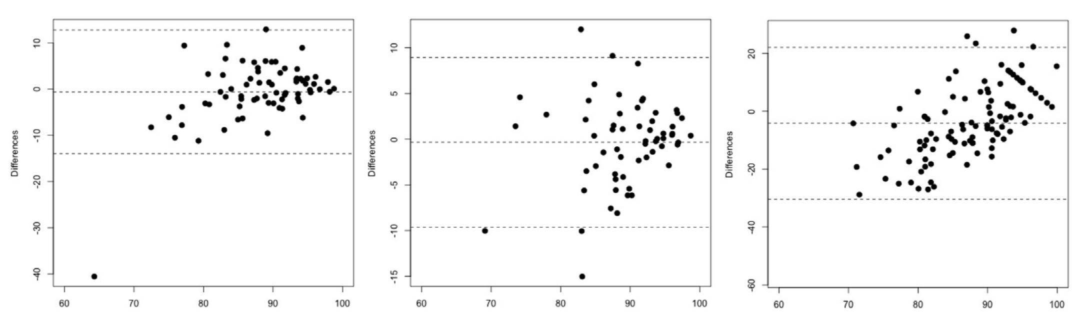

Fig. 5. Bland-Altman plots for sleep efficiency (SE); SASS to sleep diary (left), SASS-Y to sleep diary (middle), PSQI to sleep diary (right). Note. Bland-Altman plot comparing agreement of either whole week (SASS; left), split week (SASS-Y; middle) or Pittsburgh Sleep Quality Index (PSQI; right) measurement of sleep efficiency (SE; in minutes) to sleep diary measurement of SE. The x-axis represents the mean between SASS, SASS-Y, or PSQI SE and sleep diary SE, and the y-axis represents the difference between SASS, SASS-Y, or PSQI SE and sleep diary SE. Each point on the graph represents the data from one participant. The center line is the mean difference between the two measures, and the outer dashed lines are 95% CIs. On the y-axis, positive values indicate overestimation by the SASS, SASS-Y, or PSQI measures, and negative values represent underestimation by the SASS, SASS-Y, or PSQI measures.

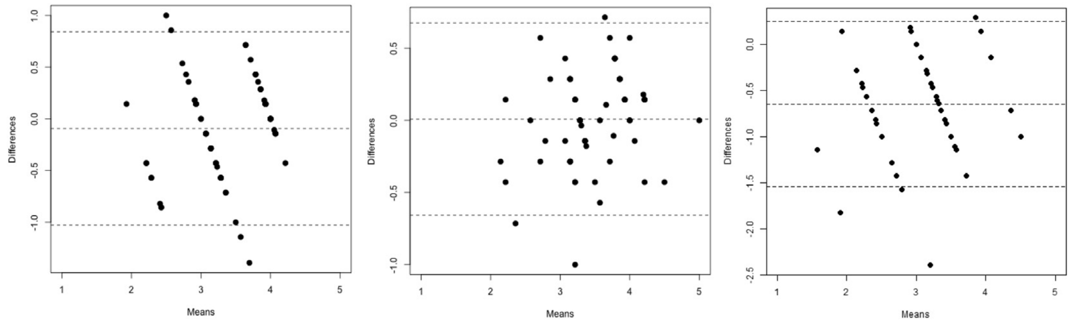

Fig. 6. Bland-Altman plots for sleep quality (QUAL); SASS to sleep diary (left), SASS-Y to sleep diary (middle), PSQI to sleep diary (right). Note. Bland-Altman plot comparing agreement of either whole week (SASS; left), split week (SASS-Y; middle) or Pittsburgh Sleep Quality Index (PSQI; right) measurement of sleep quality to sleep diary measurement of QUAL. The x-axis represents the mean between SASS, SASS-Y, or PSQI QUAL and sleep diary QUAL, and the y-axis represents the difference between SASS, SASS-Y, or PSQI QUAL and sleep diary QUAL. Each point on the graph represents the data from one participant. The center line is the mean difference between the two measures, and the outer dashed lines are 95% CIs. On the y-axis, positive values indicate overestimation by the SASS, SASS-Y, or PSQI measures, and negative values represent underestimation by the SASS, SASS-Y, or PSQI measures.

and the weighted sleep diaries on measures of TST, SE, and QUAL, but not SOL (TWAK and WASO are not assessed by the PSQI).

As shown in Table 4, the SASS and the weighted sleep diaries were significantly correlated on all sleep parameters ($r = 0.24$ to 0.84). The SASS-Y and the weighted sleep diaries also were significantly correlated on all sleep measures, to a somewhat stronger degree than the SASS ($r = 0.51$ to 0.85). The PSQI was correlated with the weighted sleep diaries on SOL, TST, and QUAL ($r = 0.67$ to 0.75), but not on SE ($r = 0.13$).

4. Discussion

The need for valid measurement of specific sleep domains is one of the most fundamental and pressing concerns in sleep research. However, no validated questionnaires exist that directly correspond to the gold standard of subjective sleep measurement, sleep diary. Further, most previous research has not differentiated weekday versus weekend sleep in retrospective assessments, despite known differences in average sleep patterns by different days of the week/schedule obligations [2]. This was the first study to (1) explore validity of a retrospective questionnaire version of the Consensus Sleep Diary [4] and (2) experimentally examine whether retrospectively measuring weekday and weekend sleep separately (ie, SASS-Y) offered improvements over retrospectively measuring sleep over the entire week (ie, SASS) in comparison to prospectively measuring sleep daily for a week (ie, sleep diaries). Our results suggested separate recall of weekday versus weekend sleep may help enhance validity for some sleep parameters (TWAK, SOL, SE, QUAL) when administering prospective sleep diaries is not feasible.

The SASS-Y version of the retrospective sleep diary had stronger correlations with the prospective sleep diaries on all six sleep parameters compared to the SASS, and the SASS-Y appeared slightly less biased than the SASS in estimating TWAK, SOL, SE, and QUAL. However, the differences between the amount of bias in the SASS and SASS-Y were not particularly large (ie, <2 min difference in bias for TWAK, SOL, and SE and 0.08 unit difference in bias for QUAL). Notably, the SASS-Y overestimated TST by approximately 15 min (compared to the SASS, which only underestimated TST by 0.5 min). These results are identical to findings from Lauderdale [2]: in their study, TST reported using a weighted average from two separate questions about weekday and weekend sleep was significantly longer by 15 min compared to a single question about prior week's sleep. However, our findings build on this study by: 1) assessing other sleep parameters beyond just TST, and 2) by comparing the SASS and SASS-Y to a prospective sleep diary, which allowed us to determine convergent validity to the current self-report gold standard. These results are in contrast with two previous studies which found, on average, compared to sleep diary a retrospective questionnaire of sleep underestimated TST by 11 min [32] in a sample of community-dwelling older adults and by 15 min in a sample of 20–40 year-old adults [17]. These discrepancies may be due to differences in population or in the formulation of the sleep questionnaire and sleep diary used in the respective studies.

It is possible that by asking participants to recall weekday and weekend sleep separately, we primed them to rely on heuristics or stereotypes about weekend sleep to estimate their TST (eg, "Everyone sleeps in longer on the weekends.") It is noteworthy that, while the SASS-Y measure of TST was slightly more biased, it also was more strongly correlated with sleep diary TST compared to the SASS ($r = 0.80$ versus $r = 0.61$, respectively). This suggests people may be more likely to overestimate their TST on the SASS-Y, but they do so in a manner consistent with their actual sleep (eg, if TST was longer on diary, their questionnaire estimate would also be longer). In contrast to the SASS and SASS-Y, the PSQI sleep parameters appeared to be more biased and less strongly correlated with the prospective sleep diaries. Because retrospective questionnaires are not highly accurate in comparison to sleep diary, these findings suggest retrospective measures should be avoided when prospective measures are possible particularly if the sleep parameter of interest is TST.

Although bias (ie, aggregate over/underestimation) of sleep parameters for retrospective questionnaires compared to sleep diary was generally within acceptable ranges, the precision (ie, standard deviation around mean differences) was generally poor (ie, large) resulting in broad limits of agreement. Large precision values indicated a great deal of variability in the estimates on all of the retrospective questionnaires compared to sleep diary, meaning that for a given individual the accuracy of questionnaires was highly variable. However, the SASS-Y demonstrated the best precision compared to the SASS and PSQI across all sleep parameters, and this outperformance was particularly prominent for TST. Ultimately, our results suggest retrospective questionnaires do not yield accurate estimates of TST in comparison to sleep diary, as none of the three questionnaires had mean differences that fell within the predetermined acceptable limits of agreement (ie, ±60 min difference).

In sum, the relative benefit of including the additional questions about weekday and weekend sleep must be weighed against potential burden on respondents. The results from the t-tests suggest that the SASS and SASS-Y are similarly able to validly estimate

Table 3
Means, standard deviations, and T-Tests between SASS, SASS-Y, sleep diary, and Pittsburgh sleep quality index.

	Mean	Std. Dev.	t-value (SASS to Sleep Diary)	p-value (SASS to Sleep Diary)	Cohen's d (SASS to Sleep Diary)	t-value (SASS-Y to Sleep Diary)	p-value (SASS-Y to Sleep Diary)	Cohen's d (SASS-Y to Sleep Diary)	t-value (PSQI to Sleep Diary)	p-value (PSQI to Sleep Diary)	Cohen's d (PSQI to Sleep Diary)
TWAK											
SASS	15.99	17.59	−0.74	0.46	−0.10	−0.22	0.83	0.01	n/a	n/a	n/a
SASS-Y	16.45	13.54									
Sleep Diary	17.28	14.90									
SOL											
SASS	27.20	23.47	1.54	0.13	0.11	1.36	0.18	0.11	−0.07	0.94	−0.01
SASS-Y	23.64	18.09									
Sleep Diary	23.53	16.90									
PSQI	23.32	16.89									
WASO											
SASS	16.08	19.04	0.48	0.63	0.05	0.83	0.41	0.07	n/a	n/a	n/a
SASS-Y	13.53	13.95									
Sleep Diary	13.93	14.66									
TST											
SASS	432.96	100.07	−0.07	0.94	−0.01	**2.67**	**0.01**	**0.22**	**−4.58**	**<0.0001**	**−0.34**
SASS-Y	447.09	71.05									
Sleep Diary	432.96	59.05									
PSQI	410.38	73.32									
SE											
SASS	87.76	8.89	−0.77	0.45	−0.08	−0.57	0.57	−0.05	**−3.18**	**<0.01**	**−0.40**
SASS-Y	89.17	6.85									
Sleep Diary	88.89	5.96									
PSQI	85.01	12.88									
QUAL											
SASS	3.30	0.67	−1.62	0.11	−0.13	0.20	0.84	0.01	**−16.05**	**<0.0001**	**−1.11**
SASS-Y	3.47	0.64									
Sleep Diary	3.42	0.57									
PSQI	2.79	0.56									

Note. **Bold** values = significant differences ($p < 0.05$). PSQI = Pittsburgh Sleep Quality Index, SASS = whole week measure, SASS-Y = split-week measure, SD = 5:2 weighted average sleep diary, TWAK = terminal wakefulness, SOL = sleep onset latency, WASO = wake after sleep onset, TST = total sleep time, SE = sleep efficiency, QUAL = sleep quality. $N = 131$ for sleep diary and PSQI measures; $n = 71$ for SASS; $n = 60$ for SASS-Y.

Table 4
Correlations between the SASS, SASS-Y, and Pittsburgh sleep quality index sleep parameters with the 5:2 weighted consensus sleep diary sleep parameters.

	Sleep Diary Variables					
	TWAK	SOL	WASO	TST	SE	QUAL
TWAK						
SASS	**0.29***	0.12	0.07	0.21	−0.21	−0.20
SASS-Y	**0.51****	0.13	0.08	−0.06	−0.32*	−0.20
SOL						
SASS	0.27*	**0.84****	0.19	−0.01	−0.69**	−0.15
SASS-Y	0.16	**0.85****	0.36**	−0.30*	−0.62**	−0.15
PSQI	0.23**	**0.75****	0.20*	−0.10	−0.59**	
WASO						
SASS	0.09	0.14	**0.56****	−0.04	−0.39**	−0.49**
SASS-Y	0.28*	0.39**	**0.76****	−0.14	−0.63**	−0.10
TST						
SASS	−0.07	−0.02	−0.26*	**0.61****	0.29**	−0.04
SASS-Y	0.02	−0.41**	−0.25	**0.80****	0.43**	0.11
PSQI	−0.09	−0.00	−0.12	**0.66****	0.24**	0.18*
SE						
SASS	−0.31**	−0.48**	−0.40**	0.14	**0.64****	0.29*
SASS-Y	0.32*	−0.69**	−0.54**	0.42**	**0.74****	0.21
PSQI	−0.13	−0.14	−0.10	0.01	**0.13**	0.15
QUAL						
SASS	−0.11	−0.11	−0.31**	0.21	0.32**	**0.72****
SASS-Y	−0.21	−0.25	−0.26*	0.30*	0.36**	**0.85****
PSQI	−0.07	−0.13	−0.26**	0.35**	0.29**	**0.67****

Note. *p < 0.05. **p < 0.01. PSQI = Pittsburgh Sleep Quality Index, SASS = whole week measure, SASS-Y = split-week measure, SD = 5:2 weighted average sleep diary, TWAK = terminal wakefulness, SOL = sleep onset latency, WASO = wake after sleep onset, TST = total sleep time, SE = sleep efficiency, QUAL = sleep quality. $N = 131$ for sleep diary and PSQI measures; $n = 71$ for SASS; $n = 60$ for SASS-Y. Any correlations with TWAK, SOL, and WASO measures are non-parametric Spearman correlations (as these variables were not normally distributed), and all TST, TWAK, QUAL measures are parametric Pearson correlations (as these variables were normally distributed).
Bold values indicate comparison of analogous parameters.

TWAK, SOL, WASO, SE, and QUAL and the SASS may actually be slightly less biased than the SASS-Y for estimating TST. However, the SASS-Y demonstrated better precision across all sleep parameters and thus may be preferable in studies with smaller sample sizes or in a clinical setting when accuracy of reporting for individuals is prioritized over the aggregate accuracy. In contrast, for studies where participant burden is a concern, use of a briefer whole week measure like the SASS may be sufficient to obtain summary information about sleep parameters for the sample.

4.1. Limitations and future directions

While this study has a number of strengths, some limitations warrant further research. First, all participants completed the SASS or SASS-Y after tracking their sleep habits for an entire week, which may have influenced their responses. Improvements in participant recall offered by the SASS-Y over the SASS may have been diluted by participants' increased attention towards their sleep habits over the previous week, regardless of the condition to which they were randomized. Second, an important assumption in the current study is that sleep diaries were completed prospectively. However, some participants may have completed their sleep diaries at one time, retrospectively, at the end of the week. To minimize this problem, participants were specifically discouraged from filling in data if they forgot. Although this instruction may have decreased the likelihood of participants completing multiple days of their sleep diaries at one time, it likely contributed to increased missing data. Third, all participants began and ended the study on different days. This means that individuals in the SASS-Y condition may have completed the SASS-Y between one and five days after the weekend sleep they were actually reporting on ended. Ideally, the day participants started and ended the study would have been standardized to reduce variability in recall bias.

Fourth, an objective (eg, PSG) or inferred measure of sleep (eg, actigraphy) was not available for the current study. Matthews et al. [33], suggest that it is vital to assess sleep parameters such as sleep duration using multiple methods of assessment to ensure confidence in the results. A replication study including objective or inferred measures of sleep would allow for comparison that is less reliant on participant adherence to instructions and not prone to self-report bias, and this is an important next step in this line of research. Fifth, an additional underlying assumption contributing to the rationale for the current study is that participants would have different sleep schedules on weekdays and weekends due to changes in school and/or work responsibilities throughout the week. More than 35% of participants worked or had class on weekends, which suggests many participants in the study had schedules which deviated from the traditional workweek that likely impacted their sleep schedules. Ultimately, as a consequence of this study's use of college students, who may be less likely to have traditional work schedules (eg, Monday through Friday, 9am-5pm), the variability between weekday and weekend may have been lessened, which may have resulted in less variance overall to find significant differences. Our results also indicated that some of the mean differences between the SASS/SASS-Y and weighted sleep diaries were not normally distributed (SASS SOL, SASS SE, SASS WASO, and SASS-Y SOL were all leptokurtic), which also may have biased Bland-Altman results and restricted the ability to find differences.

In future research, it would be important to replicate the current study in adults with more traditional work schedules, as well as other populations (eg, older adults, clinical populations, individuals with children/families). In addition, it would be best to extend the question time frame in the SASS-Y to “weekday/work days” and “weekend/off days”. However, as shown above, this is a relatively smaller percent of the population (eg, 3%–20% of our sample). As currently designed, the SASS-Y would likely not be valid in shift-working populations, but neither would the SASS or the PSQI. Another important future direction in this area is to examine whether certain participant characteristics may be responsible for differences in accuracy of retrospective reporting of sleep information. A recent study by Mathews et al. [33], demonstrated that personal factors such as hostility, depressive symptoms, age, insomnia symptoms, and perceptions of poor health were factors that partially explained differences between prospectively- and retrospectively-assessed sleep parameters.

4.2. Summary and conclusion

In summary, the results of this study suggested asking participants to retrospectively estimate their sleep on weekdays and weekends separately offers some small advantages for estimating some sleep parameters (TWAK, SOL, SE, and QUAL) compared to asking participants to retrospectively estimate their sleep for the entire week or past month using the PSQI. The separate recall of weekday and weekends may lead to slight overestimation of TST but less variability in accuracy. Sleep diaries remain preferable over retrospective questionnaires whenever possible but use of a questionnaire corresponding to the Consensus Sleep Diary, particularly one that asks about weekdays and weekends separately, offers clear benefit above existing retrospective questionnaires of sleep. As with any questionnaire, researchers and clinicians must balance validity and reliability with ease of administration and completion and consider the specific needs of the population of interest when choosing a retrospective questionnaire over other measures.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: https://doi.org/10.1016/j.sleep.2019.05.015.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sleep.2019.05.015.

References

[1] Buysse DJ, Yu L, Moul DE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. Sleep 2010;33:781.
[2] Lauderdale DS. Survey questions about sleep duration: does asking separately about weekdays and weekends matter? Behav Sleep Med 2014;12:158–68.
[3] Hasler BP, Dahl RE, Holm SM, et al. Weekend-weekday advances in sleep timing are associated with altered reward-related brain function in healthy adolescents. Biol Psychol 2012;91:334–41.
[4] Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012;35:287–302.
[5] Bei B, Wiley JF, Trinder J, et al. Beyond the mean: a systematic review on the correlates of daily intraindividual variability of sleep/wake patterns. Sleep Med Rev 2016;28:108–24.
[6] Lichstein KL, Stone KC, Donaldson J, et al. Actigraphy validation with insomnia. Sleep 2006;29:232–9.
[7] Monk TH, Reynolds CF, Kupfer DJ, et al. The Pittsburgh sleep diary. J Sleep Res 1994;3:111–20.
[8] Coates TJ, Killen JD, George J, et al. Estimating sleep parameters: a multitrait-multimethod analysis. J Consult Clin Psychol 1982;50:345–52.
[9] Vallières A, Morin CM. Actigraphy in the assessment of insomnia. Sleep 2003;26:902–6.
[10] Lockley SW, Skene DJ, Arendt J. Comparison between subjective and actigraphic measurement of sleep and sleep rhythms. J Sleep Res 1999;8:175–83.
[11] Martin JL, Hakim AD. Wrist actigraphy. Chest 2011;139:1514–27.
[12] Moul DE, Hall M, Pilkonis PA, et al. Self-report measures of insomnia in adults: rationales, choices, and needs. Sleep Med Rev 2004;8:177–98.
[13] Buysse DJ, Reynolds III CF, Monk TH, et al. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28:193–213.
[14] Besedovsky L, Lange T, Born J. Sleep and immune function. Pflueg Arch Eur J Physiol 2012;463:121–37.
[15] Wolfson AR, Carskadon MA. Sleep schedules and daytime functioning in adolescents. Child Dev 1998;69:875–87.
[16] Wolfson AR, Carskadon MA, Acebo C, et al. Evidence for the validity of a sleep habits survey for adolescents. Sleep 2003;26:213–6.
[17] Babkoff H, Weller A, Lavidor M. A comparison of prospective and retrospective assessments of sleep. J Clin Epidemiol 1996;49:455–60.
[18] Bernert RA, Turvey CL, Conwell Y, et al. Association of poor subjective sleep quality with risk for death by suicide during a 10-year period: a longitudinal, population-based study of late life. JAMA Psychiatry 2014;71:1129–37.
[19] Bardwell WA, Berry CC, Ancoli-Israel S, et al. Psychological correlates of sleep apnea. J Psychosom Res 1999;47:583–96.
[20] Ağargün MY, Kara H, Solmaz M. Sleep disturbances and suicidal behavior in patients with major depression. J Clin Psychiatry 1997;58:249–51.
[21] Dietch JR, Taylor DJ, Sethi K, et al. Psychometric evaluation of the PSQI in US college students. J Clin Sleep Med 2016;12:1121–9.
[22] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Int J Nurs Stud 2010;47:931–6.
[23] Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60.
[24] de Zambotti M, Claudatos S, Inkelis S, et al. Evaluation of a consumer fitness-tracking device to assess sleep in adults. Chronobiol Int 2015;32:1024–8.
[25] Werner H, Molinari L, Guyer C, et al. Agreement rates between actigraphy, diary, and questionnaire for children's sleep patterns. Arch Pediatr Adolesc Med 2008;162:350–8.
[26] Watson NF, Badr MS, Belenky G, et al. Joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society on the recommended amount of sleep for a healthy adult: methodology and discussion. Sleep 2015;38:1161–83.
[27] Lufi D, Tzischinsky O, Hadar S. Delaying school starting time by one hour: some effects on attention levels in adolescents. J Clin Sleep Med 2011;7:137–43.
[28] Varughese J, Allen RP. Fatal accidents following changes in daylight savings time: the American experience. Sleep Med 2001;2:31–6.
[29] Leproult R, Deliens G, Gilson M, et al. Beneficial impact of sleep extension on fasting insulin sensitivity in adults with habitual sleep restriction. Sleep 2015;38:707–15.
[30] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
[31] Lehnert B. BlandAltmanLeh: plots (slightly extended) Bland-Altman plots. R package. version 0.3. 1st ed. 2015.
[32] Libman E, Fichten CS, Bailes S, et al. Sleep questionnaire versus sleep diary: which measure is better? Int J Rehabil Health 2000;5:205–9.
[33] Matthews KA, Patel SR, Pantesco EJ, et al. Similarities and differences in estimates of sleep duration by polysomnography, actigraphy, diary, and self-reported habitual sleep in a community sample. Sleep Health 2018;4:96–103.