SLEEP, 2025, **48**, 1–13
https://doi.org/10.1093/sleep/zsaf065
Advance access publication 15 March 2025
Original Article

Original Article

Assessment of sleep measures and their agreement: youth-reported, caregiver-reported, and Fitbit-derived data in a large early adolescent cohort

Orsolya Kiss[1,*], Adrianna Shaska[2], Eva M. Müller-Oehring[1,3], Brant P. Hasler[4], Peter L. Franzen[4], Douglas H. FitzGerald[4], Duncan B. Clark[4] and Fiona C. Baker[1,5]

[1]Center for Health Sciences, SRI International, Menlo Park, CA, USA,
[2]Physics and Astronomy, Michigan State University, East Lansing, MI, USA,
[3]Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA,
[4]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA and
[5]School of Physiology, University of the Witwatersrand, Braamfontein, South Africa

[*]Corresponding author. Orsolya Kiss, Center for Health Sciences, SRI International, 333 Ravenswood Ave, Menlo Park, CA 94025, USA. Email: orsolya.kiss@sri.com.

Abstract

Study Objectives: Adequate sleep is essential for adolescents' physical, emotional, and cognitive well-being. However, accurately capturing the complex components of sleep in this demographic is challenging, especially with retrospective self-report measures. This study aims to compare sleep data obtained from youth reports, caregiver reports, and Fitbit devices among early adolescents.

Methods: Data from 11 879 adolescents (11–14 years, 47.83% female), in year 2 of the Adolescent Brain Cognitive Development (ABCD) study, were analyzed. Adolescents self-reported their sleep characteristics using the Munich Chronotype Questionnaire, and caregivers provided data through the Children's Sleep Disturbance Scale. Additionally, a subset of participants (N = 3803) wore Fitbit Charge 2 devices for 21 days. We assessed the questionnaires' internal consistency and utilized Bland-Altman and interclass correlation analyses for comparing self-reported sleep characteristics with Fitbit measures.

Results: Youth-reported and caregiver-reported sleep questionnaires demonstrated acceptable internal consistency. Discrepancies between caregiver and adolescent reports were more pronounced when adolescents reported sleep periods of less than 7 hours. Compared to Fitbit measurements, adolescents' self-reports showed a reasonably high agreement on sleep period and bedtime, while agreement on the duration of wakefulness after sleep onset and sleep onset latency was poor.

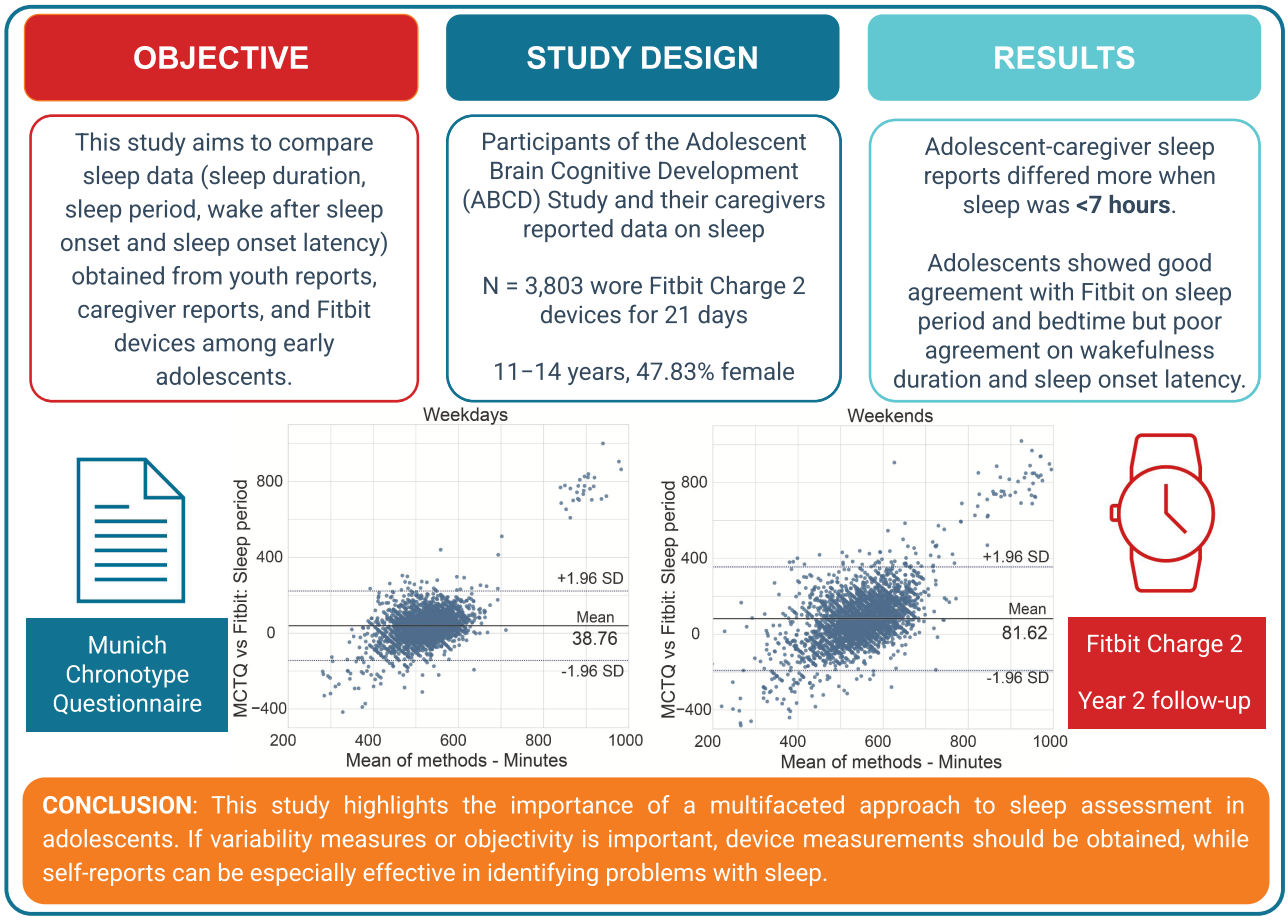
Conclusions: Results show reasonable agreement between adolescent self-reports and Fitbit measurements of sleep period and indicate their usefulness in assessing sleep behavior in adolescents. Caregivers provided valuable perspectives on the youth's sleep disturbances; however, they tended to overestimate sleep duration. These findings offer important methodological insights and highlight the necessity of adopting multi-dimensional approaches to assess sleep in adolescents.

Key words: Fitbit; self-reported sleep; adolescence

Submitted for publication: October 17, 2024; **Revised:** February 7, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of Sleep Research Society. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

Graphical Abstract



Statement of Significance

This study, using data from adolescents aged 11 – 14 years in the Adolescent Brain Cognitive Development study, offers key insights into adolescent sleep patterns. In a comparison of caregiver-reported, youth-reported, and Fitbit-measured sleep data from a subset (N = 3803) who wore Fitbit devices for 21 days, youth-reported sleep periods showed better alignment with Fitbit data than with caregiver reports. Youth-reported sleep bedtimes also show significant alignment with Fitbit data. However, challenges in estimating wakefulness and sleep onset latency suggest inherent biases in both subjective and objective measures. These findings underscore the importance of using multiple assessment methods to capture the complexity of adolescent sleep and improve understanding of sleep health.

Accurately assessing sleep patterns in adolescents presents significant challenges due to the complexity of sleep as a behavior and the unique developmental context of adolescence [1]. The multidimensional nature of sleep, encompassing duration, quality, and timing, demands diverse assessment tools that can capture its various dimensions [2].

Most research into the relationship between objective and subjective sleep assessments has focused on adults [3, 4], leaving a gap in our understanding of this relationship in adolescents [5, 6]. Studies comparing objective and subjective assessments across large cohorts have found that adults tend to overestimate their mean habitual total sleep time by about 1 hour when using sleep diaries, as compared to polysomnography (PSG) recordings [3], or when using questionnaires as opposed to actigraphy [4]. Adolescence is characterized by extensive developmental changes across biological systems. These changes encompass alterations in physiological sleep and circadian processes, evolving preferences for nighttime activities and later bedtimes [7], and lifestyle modifications that can lead to disrupted and insufficient sleep [8, 9]. These developmental changes are partly driven by the slower accumulation of homeostatic sleep drive [10] and delayed circadian phase [11]. Additionally, the increasing autonomy of adolescents, alongside shifts in social and academic obligations, introduces variability in sleep habits [7] that is challenging to measure with global questionnaires and objective measures based on short-term (often single-night) overnights in the sleep laboratory. Moreover, during adolescence, there is an increase in sleep-related problems [12], ranging from sub-clinical complaints to clinical insomnia [13, 14], with greater risk in older female adolescents [14]. The relevance of understanding adolescent sleep patterns is underscored by the critical role sleep plays in supporting the intense developmental changes occurring during this

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

life stage, including cognitive and emotional growth, and physical maturation [15].

The Sleep Disturbance Scale for Children (SDSC) [16], developed from a cohort of 1157 Italian children, and the Munich Chronotype Questionnaire (MCTQ) [17], developed from a cohort of 500 adults and high school students in Germany and Switzerland, including 142 participants younger than 21, are well-established tools designed to standardize the assessment of sleep behaviors. Both instruments had proven validity in their respective initial studies and have been extensively used in different populations around the world. Both questionnaires were included in the large, diverse Adolescent Behavior and Cognitive Development (ABCD) study, an ongoing prospective cohort study of over 11 000 adolescents in the United States, providing a unique opportunity to evaluate these measures within the rich context of adolescent development, particularly given the study's demographic and developmental breadth which extends beyond the original populations for which these questionnaires were validated. Measures of sleep from the ABCD study have already been shown to be related to brain development [18], obesity risk [19], and mental health [20, 21], showing the value of sleep assessments in the broader context of adolescent health.

Although polysomnography (PSG) remains the gold standard for quantifying sleep, research-grade and commercial wearables, which now frequently include multi-sensors along with accelerometers, are widely available and accepted proxies for estimating sleep in nonlaboratory settings. Commercial wearables such as Fitbit have shown specificity and sensitivity in detecting sleep equivalent to or better than research-grade actigraphy [22, 23]. Fitbit tracks sleep using a blend of movement data from a 3-axis accelerometer and heart rate information via a photoplethysmography sensor. These data are used to derive information about sleep and wake states, including distinguishing various sleep stages (light, deep, and REM sleep), by noting the changes in movement and heart rate patterns typically associated with each stage. Fitbit's sleep tracking feature has been the subject of several studies both in adults [24, 25] and adolescents [23, 26, 27], examining its effectiveness and accuracy. Actigraphy shows high sensitivity in detecting true sleep and overall accuracy in detecting both wake and sleep; however, it is inherently limited in detecting true wake (specificity) as it cannot differentiate motionless wakefulness from sleep [28], and its performance in distinguishing between different sleep stages is generally lower [29]. Commercial wearables are limited in reliably determining lights-off times, making it challenging to accurately assess sleep onset latency without users providing self-reported lights-out information [28]. The classification of wake-after-sleep onset (WASO) [26, 27] is poorer in cases when WASO is high. While cognizant of these limitations, wearables have the advantage of being able to track many nights of sleep, unobtrusively, at a low cost to users.

The primary objective of this study was to leverage the large and diverse ABCD study to explore the concordance between caregiver and youth-reported sleep reports and measures of sleep obtained via Fitbit Charge 2, with a particular focus on sleep period and duration, bedtime, wakefulness after sleep onset and sleep onset latency. Specifically, we assessed the agreement between different sleep measures, considering variations between weekday and weekend sleep patterns. Furthermore, we aimed to examine age and sex differences in the agreement between self-report and measured sleep parameters among adolescents. Based on prior research, we hypothesize that the agreement between caregiver reports and youth self-reports will be lower compared to the agreement between youth self-reports and Fitbit measures. Specifically, we expect statistically significant agreement between youth self-reports and Fitbit for all sleep measures, including sleep period, sleep duration, wake after sleep onset, sleep onset latency, and bedtime. Additionally, we anticipate that the agreement between youth self-reports and Fitbit will be stronger for sleep periods than for sleep duration.

Methods

Participants and procedure

Data were obtained from the US-based, multi-site ABCD study, which is a population-informed study on adolescent development that enrolled 11 875 children at baseline at 21 sites across the United States. The ABCD study implemented epidemiologically informed strategies via school systems to recruit a sample representative of the diversity within the United States. Additional information about recruitment, procedures, and measures for the ABCD study are described elsewhere [30]. Study participants provided written informed assent and parent/guardians provided written informed consent. Centralized institutional review board (IRB) approval was obtained from the University of California, San Diego (protocol number: #160091AW). Study sites obtained approval from local IRBs.

Starting in year 2 of the study, data were collected from 6078 ABCD participants using Fitbit wearable devices. Of those, 5045 subjects had data within a 21-day period after the annual assessment. Participants within the Fitbit sample were excluded if they either did not have at least 4 days of data in a week or if they did not have at least one weekend night in their data [31, 32]. This process excluded 1232 subjects for not fulfilling these criteria. An additional 173 participants were excluded for data inconsistencies, such as implausible sleep values, and eight participants were excluded for missing MCTQ or SDS data during year 2. This left a final sample of 3803 (age range: 10.58–13.67; M_{age} = 11.97 years) participants for our analysis. Participants were not excluded based on long sleep durations, unusual wake-up times, or delayed bedtimes, as these variations are expected during adolescence due to normative developmental changes and circadian shifts. A sensitivity analysis further examining these factors is presented in Section 2 of the Supplementary Material. Data are from ABCD data release 5.0.0 (released in June 2023 DOI: 10.15154/8873-zj65; the data dictionary for the ABCD Survey can be found at https://data-dict.abcdstudy.org/).

The demographic composition of the Fitbit sample differed from that of the broader ABCD study population [33]. Specifically, the Fitbit sample contained a smaller proportion of black participants (7.5% vs. 15%) and a larger proportion of white participants (60% vs. 52%). Additionally, this cohort was more likely to come from households with an annual income of $75 000 or more. Comprehensive demographic details for both the Fitbit sample and the general ABCD study population are provided in Table 1.

Measures

Demographics.

Demographic information was taken from the ABCD Study® baseline visit: sex at birth (female), race (white, black, Asian, multiple, other/not reported), ethnicity (Hispanic/Latino, Non-Hispanic/Latino, Not reported). Parental education (less than a high school diploma, high school diploma/GED, some college, bachelor's degree, postgraduate degree) was used as an indicator of family socioeconomic status. Age in months at the year 2 follow-up of the ABCD Study® was used to assess associations with chronological age.

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

Table 1. Demographics Collected at Baseline for the Subset of ABCD Study® Participants Included in the Current Analysis (Fitbit Sample) and the Complete ABCD Study Sample

Variable	Fitbit sample	Complete ABCD sample
N	3803	11 879
Biological sex		
Male	1957 (51.46%)	6196 (52.16%)
Female	1846 (48.53%)	5682 (47.83%)
Race		
White	2310 (60.73%)	6178 (52.01%)
Black	287 (7.55%)	1784 (15.02%)
Hispanic/Latino	737 (19.38%)	2411 (20.30%)
Asian	99 (2.60%)	252 (2.12%)
Other	370 (9.73%)	1248 (10.51%)
Unknown/not reported	0 (0.00%)	5 (0.04%)
Highest parental education		
<High school	107 (2.81%)	593 (4.99%)
HS diploma/GED	229 (6.02%)	1132 (9.53%)
Some college	834 (21.94%)	3079 (25.92%)
Bachelor's degree	1097 (28.85%)	3015 (25.38%)
Postgraduate degree	1530 (40.24%)	4044 (34.05%)
Unknown/not reported	4 (0.11%)	14 (0.12%)
Household income		
<$74 999	1253 (32.96%)	4722 (39.75%)
>=$75 000	2311 (60.78%)	6137 (51.67%)
Unknown/not reported	237 (6.23%)	1018 (8.57%)

Self-reported sleep measures

Munich Chronotype Questionnaire (MCTQ) [17]: As part of their year 2 annual assessment, participants reported their typical sleep over the past 4 weeks, with separate reports for school days (typically weekdays) and free days (typically weekends). The MCTQ was modified for the ABCD study to include questions about awakenings and WASO [34] that are not in the original version. We included bedtime (hours, based on the item: "I go to bed at … o'clock."), wake-up time (hours), sleep onset latency (based on the item: "I need… minutes to fall asleep"). Sleep period was defined as the period between the sleep onset ("I fell asleep at…o'clock") and wake-up time, without excluding WASO. Sleep duration was defined as the period between the time of sleep onset and wake-up time, excluding WASO ("Altogether, these awakenings last … minutes"). We also calculated relative social jetlag (hours: the difference between the middle of the sleep period on free days (e.g. weekends) and weekdays) variables in our analysis. The middle point was defined as the point halfway through an individual's sleep period, calculated as the time of sleep onset plus half of the total sleep duration.

Sleep disturbance scale for children (SDSC) [16]: The questionnaire includes 26 items administered to the parents or caregivers of adolescents to assess a broad spectrum of sleep disturbances, including disorders of initiating and maintaining sleep (seven items), sleep breathing disorders (three items), disorders of arousal (seven items), sleep–wake transition disorders (four items), disorders of excessive somnolence (four items), and sleep hyperhidrosis (excessive sweating during sleep, two items) in the past 6 months. The Likert scale used for responses (items 3 to 26) ranges from 1 (never) to 5 (always/daily), allowing for the quantification of sleep disturbance severity through summed scores. Based on the recommendations of the developers of the survey [16], a cutoff of 39 was used to indicate that a child had acute sleep disturbances.

In the interrater comparison analysis, for caregiver-reported hours of sleep, we focused on the SDSC's first question regarding children's typical sleep time per night: "How many hours of sleep does your child get on most nights." The response options were the following: (1) 9–11 hours, (2) 8–9 hours, (3) 7–8 hours, (4) 5–7 hours, (5) less than 5 hours. Due to underrepresentation in the "less than 5 hours" category, we merged it with the "5–7 hours" category to create a new, more representative category of "<7 hours." For sleep onset latency we used the second item: "How long after going to bed does your child usually fall asleep?." The response options were the following: (1) "less than 15', (2) 15–30', (3) 30–45', (4) 45–60', (5) more than 60'.

Objective sleep measures

For the year 2 assessment, ABCD study participants were invited to wear a Fitbit Charge HR 2 (Fitbit Inc., San Francisco, CA) on their wrist for 21 days [35]. Data were collated and stored through a third party (Fitabase). This device, previously validated against gold-standard research tools for children and adolescents [23], enabled the collection of objective sleep data. Key metrics assessed included total sleep duration (sum of time in light, REM, and deep sleep stages, i.e. not including WASO), sleep period (sleep duration including WASO), bedtime (estimated based on the "first minute in bed" as labeled by Fitabase), the total time spent awake during the night after sleep onset (WASO; as provided by Fitabase) and sleep onset latency (time difference between bedtime and sleep onset, defines as the first minute classified as sleep by Fitabase). Importantly, all measurements were separately evaluated for weekdays and weekends. As such, weekday and weekend averages of all the measures are reported. Additionally, we calculated the weighted weekly averages. We also calculated social jetlag, defined as the difference in the middle point of the sleep period between weekdays and weekends. The middle point was defined as sleep onset + sleep duration/2.

Psychometric properties of the self-reported sleep measures

In evaluating the reliability of the MCTQ (including the items that we assessed in the ABCD study), the calculated Cronbach's alpha for the year 2 follow-up was 0.70, with a confidence interval ranging from 0.697 to 0.714. This statistic is indicative of acceptable internal consistency, suggesting that the questionnaire items are sufficiently correlated and thus reliable for assessing sleep measures in our sample.

The SDSC demonstrated strong reliability in this dataset, with an overall Cronbach's alpha of 0.83 [CI: 0.828, 0.837], indicating high internal consistency across the scale. The subscale analysis revealed varying degrees of reliability, reflective of the diverse nature of sleep disturbances assessed. The reliability for "disorders of initiating and maintaining sleep" (7 items) was satisfactory, with a Cronbach's alpha of 0.72 [CI 0.716, 0.731], supporting its use in identifying these specific sleep issues. However, subscales assessing "sleep breathing disorders" (three items) and "disorders of arousal" (seven items) showed lower reliability, with Cronbach's alphas of 0.39 [CI 0.374, 0.413] and 0.45 [CI 0.44, 0.475], respectively. These values suggest a need for further refinement

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

or cautious interpretation when evaluating these particular disturbances. Subscales for "Sleep–Wake Transition Disorders" (four items), "Disorders of Excessive Somnolence" (four items), and "Sleep Hyperhidrosis" (two items) showed better reliability, with alphas of 0.61 [CI 0.607, 0.629], 0.73 [CI 0.722, 0.738], and 0.81 [CI 0.805, 0.819], respectively. These values indicate a satisfactory to a high level of internal consistency, enhancing confidence in these subscales' ability to accurately assess the respective sleep disturbances.

Our confirmatory factor analysis (CFA) of the SDSC indicated a root mean square error of approximation (RMSEA) of 0.05, suggesting a good fit. However, the comparative fit index (CFI) and Tucker-Lewis Index (TLI) values, at 0.845 and 0.823, respectively, fell slightly below the preferred threshold of 0.90, indicating a moderate overall model fit. These results affirm the SDSC's construct validity, with indications for potential refinement to better capture the complexity of sleep disturbances.

Statistical analysis

First, we calculated the psychometric properties of the questionnaires. Cronbach's alphas were calculated using the *pingouin* Python package [36] to assess the internal consistency of both youth and caregiver-reported surveys. Given the multidimensional structure of the Sleep Disturbance Scale for Children (SDSC), which comprises six subscales representing different sleep disturbances, we employed confirmatory factor analysis (CFA) using the *semopy* package [37].

Second, we compared caregiver-reported hours of sleep against both youth-reported sleep periods and sleep periods measured by Fitbit. This comparison utilized specific statistical tests and visualizations tailored to categorical and continuous measures, evaluating agreement from multiple perspectives. A distribution analysis with a normalized agreement matrix helped visualize the overall alignment between the different sources of categorical data (i.e. caregiver reported sleep duration vs. youth reported and Fitbit measures of sleep duration). Correlation coefficients [38] were employed to quantify the strength of linear relationships between the measures, while Wilcoxon statistics [36] tested for significant differences in the median sleep duration values across the groups, accounting for non-parametric data distributions. Cohen's Kappa [36] and Krippendorff's alpha statistics were used to assess the agreement between categorical classifications of sleep duration, with Cohen's Kappa focusing on pairwise interrater reliability, and Krippendorff's alpha offering a robust measure of reliability across multiple raters and categories. Specifically, Krippendorff's alpha statistics are designed for evaluating interrater reliability in categorical data among multiple measures. We aligned the categories for youth-reported and objective sleep duration, and sleep onset latency data with those defined by the Sleep Disturbance Scale for Children (SDSC). We used the Common Language Effect Size (CLES) to express the probability that one measure (e.g. caregiver report) is greater than another (e.g. Fitbit data), based on the distribution of values, providing an intuitive interpretation of the differences between them.

Additionally, we calculated the intraclass correlation coefficient (ICC) to assess the consistency between youth-reported continuous metrics (including the sleep period, accounting for intervening wakefulness, sleep duration, bedtime, sleep onset latency, and wake after sleep onset) and objective sleep data collected via the Fitbit Charge 2. We further extended our analysis to include Bland-Altman plots [36] for both sleep duration and bedtime, providing a detailed examination of the agreement between reported and objectively measured sleep parameters.

To assess the impact of developmental and sex differences on self-reported versus objectively measured sleep duration and bedtime, we conducted a linear regression analysis. The dependent variable for this analysis was the difference score, calculated as the discrepancy between youth-reported and objective measures of sleep duration and bedtime. Age, measured in months, and sex were included as independent variables, along with their interaction term and the research site as a random term. The analysis was implemented in R using the "lme4" package [39].

Results

Descriptive statistics for all sleep measures, including caregiver-reported sleep duration, youth self-reports, and Fitbit-derived sleep metrics, are presented in Table 3. For more in-depth descriptive statistics, please see Section 1 of the Supplementary Material.

The caregiver-reported sleep disturbances revealed that 25.2% of the sample had a total sleep disturbance score exceeding 39, indicating sleep disturbances among these participants. This subset of individuals, identified through their elevated sleep disturbance scores, exhibited an average sleep duration from Fitbit of 7.42 hours, with a mean bedtime of 11:04 pm (Fitbit). In both the youth sleep measures (MCTQ) and Fitbit-collected data, significant differences were observed between participants whose total sleep disturbance score exceeded 39 and those without sleep problems (SDSC). Measures such as sleep period, sleep duration, bedtime, and WASO consistently showed statistical significance in most comparisons, with p-values indicating meaningful differences (Table 2).

Comparing caregiver-reported hours of sleep and youth-reported sleep period

Our analysis revealed varying degrees of agreement across different sleep period categories. Notably, the highest level of agreement was observed in the 8–9 hours category, with a likelihood of 0.54, suggesting that caregivers and youth are most aligned in their reports at this range of sleep period (Figure 1). In instances where youth reported sleep periods of 9–11 hours, caregivers were equally likely to report either 9–11 hours or 8–9 hours, with a likelihood of 0.43. This suggests that while caregivers sometimes reported hours of sleep matching the youth, they were just as likely to report slightly shorter sleep, indicating potential underestimation of longer sleep periods.

Furthermore, when youth reported sleeping 7–8 hours, the most common caregiver response was also 8–9 hours, albeit with only a 0.31 likelihood for the 7–8 hours category itself. In the most pronounced cases of reported short sleep (<7 hours) by the youth, the likelihood of caregivers reporting the same category was notably low (0.13), with a higher likelihood for caregivers reporting longer sleep periods (7–8 hours at 0.37 and 8–9 hours at 0.38), further highlighting this tendency for overestimation. Although the two measures are significantly associated (Spearman's rho = 0.30, $p < .01$), the Cohen's Kappa (κ) statistics indicate a small agreement (Cohen's $\kappa = 0.14$) between the two sets of measurements. This finding is further supported by the results of the Wilcoxon signed ranked test (W-val = 540436.5, $|r| = .54$, $p < .01$), showing significant differences between the measures. The proportion of

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

Table 2. Comparison of Sleep Measures Between Participants With (N = 958) and Without Sleep Problems (N = 2845) Based on Caregiver-Reported SDSC (Total Sleep Disturbance Score > 39)

Variable	Total score of SDSC >39 (mean)	Total score of SDSC ≤ 39 (mean)	t-value (P-value)	Cohen's d
Sleep period (MCTQ; hours)	9.21	9.33	*−1.98* (.048)	−0.08
Sleep period (Fitbit; hours)	8.36	8.47	*−3.90* (<.001)	−0.15
Sleep duration (MCTQ; hours)	9.13	9.26	*−2.16* (.031)	−0.08
Sleep duration (Fitbit; hours)	7.42	7.51	*−3.71* (<.001)	−0.14
Bedtime (MCTQ; clock time)	09:44 pm	09:32 pm	*4.78* (<.001)	0.18
Bedtime (Fitbit; clock time)	11:04 pm	10:48 pm	*6.28* (<.001)	0.24
WASO (MCTQ; minutes)	7.49	4.98	*3.61* (.001)	0.14
WASO (Fitbit; minutes)	56.64	57.88	*−3.01* (.003)	−0.11
SOL (MCTQ; minutes)	25.94	18.66	*7.06* (<.001)	0.29
SOL (Fitbit; minutes)	6.94	6.82	0.56 (.575)	0.02

MCTQ, Munich Chronotype Questionnaire; WASO, wake after sleep onset; SOL, sleep onset latency.
Significant differences are highlighted with italics.

pairs when the caregiver report was higher than the youth report was 0.64, as indicated by the common language effect size (CLES = 0.64).

Comparing caregiver-reported hours of sleep with objective sleep period, measured by Fitbit

In examining the concordance between caregiver-reported hours of sleep and measures from Fitbit, we decided to compare the caregiver reports with the categorized Fitbit-defined sleep period, given that the SDSC does not account for WASO. The normalized agreement matrix reveals that the highest likelihood of agreement is observed for longer sleep durations (Figure 2), with a 0.55 likelihood for Fitbit measurements and caregiver reports within the 9–11 hours category. This likelihood decreases across categories as sleep duration shortens, highlighting a trend where agreement diminishes as reported sleep duration decreases. Specifically, for the sleep period measured by Fitbit in the 8–9 hours category, the likelihood of a corresponding caregiver report in the same category is 0.50, indicating a moderately high level of concordance. However, as the measured sleep period shortens to 7–8 hours and below 7 hours, the likelihood of caregiver agreement falls to 0.29 and 0.15, respectively, in the same categories. The sleep measures were significantly correlated (Spearman's rho = 0.31) but showed only a slight agreement (Cohen's $\kappa = 0.15$). The Wilcoxon signed rank test indicates significant differences (W-val = 826134.5 $|r| = .22$, $p < .01$). The caregiver report was higher than the Fitbit period category in 43%, as indicated by the common language effect size (CLES = 0.44).

Krippendorff's alpha of 0.23 for youth-reported sleep period (MCTQ), objectively measured sleep period (Fitbit), and caregiver-reported sleep duration (SDSC) reveals a low overall agreement among these measuress

Comparing youth-reported and objectively collected sleep measures

To quantify the agreement between continuous sleep measures (sleep duration, sleep period, bedtime, WASO, and SOL) among objective and youth-reported weekly averaged measures (weighted for weekdays and free days/weekends), we utilized Intraclass Correlation Analysis (Table 3).

Our findings indicate a moderate level of agreement between youth-reported sleep duration (excludes WASO) and the corresponding Fitbit-measured sleep duration, with an ICC of 0.186 (95% CI: [0.16, 0.22], $p < .001$). Comparison between the Fitbit-measured sleep period (includes intervening wakefulness) and youth-reported sleep period revealed a higher agreement, with an ICC of 0.217 (95% CI: [0.19, 0.25], $p < .001$). Additionally, both sleep duration and sleep period had wider limits of agreement on weekends compared to weekdays (sleep duration: −135.18 to 408.65 vs. −93.4 to 276.9; sleep period: −191.77 to 355.01 vs. −144.48 to 221.99).

For bedtime measures, the agreement was notably higher, with an ICC of 0.460 (95% CI: [0.43, 0.48], $p < .001$), indicating a substantial concordance between youth-reported and Fitbit-recorded bedtimes. Conversely, the agreement for WASO was very low (with the within-group variance greater than the between-group variance), with an ICC of −0.015 (95% CI: [−0.05, 0.02], $p = .835$), suggesting poor agreement between the reported and measured values in this domain. Similarly, the agreement between the SOL was also very low, with an ICC of −0.009 (95% CI: [−0.04, 0.02], $p = .719$).

To further explore the agreement patterns between youth-reported and Fitbit-measured sleep parameters, we conducted Bland-Altman analyses, separately considering weekdays and weekends. These analyses revealed a proportional bias in both sleep duration and sleep period measures, indicating that discrepancies between youth-reported and Fitbit-measured values were associated with the length of sleep. Specifically, participants with sleep durations shorter than 400 minutes or longer than 600 minutes exhibited greater discrepancies (Figure 3). For weekdays, the Bland-Altman figures underscored the highest level of agreement for the comparison between the youth-reported sleep period and the Fitbit-measured sleep period, with a mean discrepancy of 38.76 minutes. This discrepancy doubled during weekends, reaching an average of 81.62 minutes, suggesting a significant divergence in agreement across these temporal contexts. Moreover, the analysis of sleep duration measures paralleled this pattern, revealing substantially larger discrepancies with longer sleep duration, particularly during weekends. Additionally, both the sleep duration and sleep period had wider limits of agreement on weekends than on weekdays (sleep duration: −93.4 to 276.9 vs. −135.18 to 408.65; sleep period: −144.48 to 221.99 vs. −191.77 to 355.01).

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

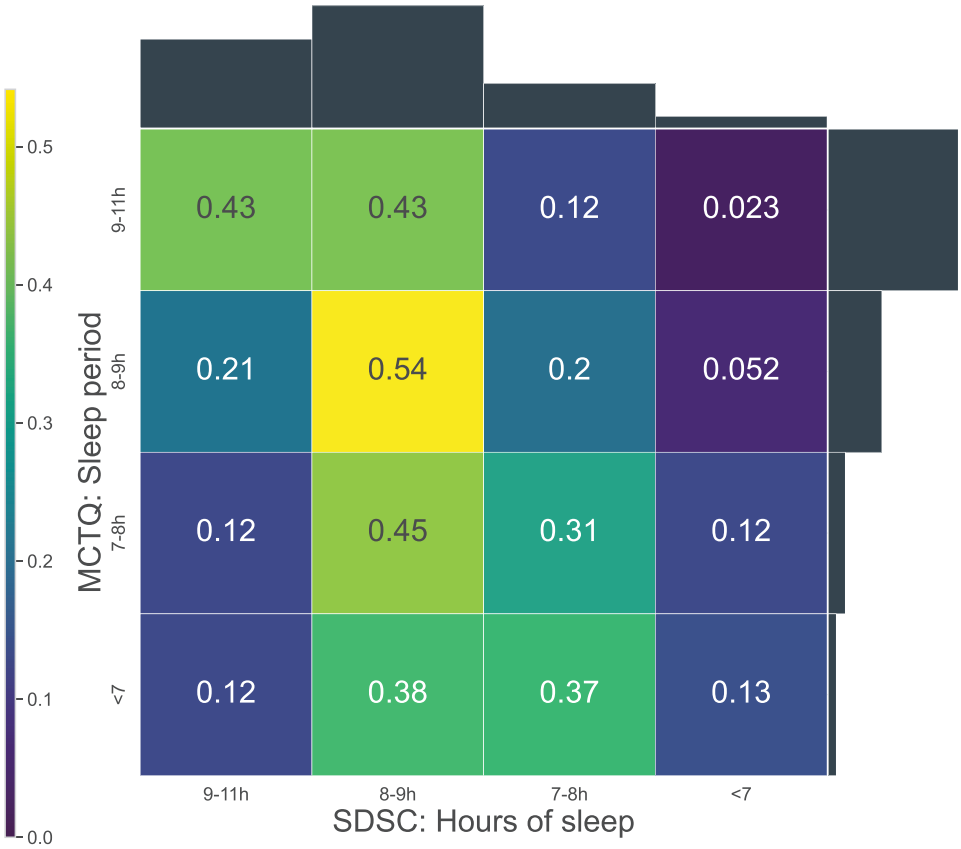


Figure 1. Agreement between Caregiver reported Hours of sleep and Youth Reported Measures. To assess the concordance between youth-reported and caregiver-reported sleep, we conducted an analysis using a normalized agreement matrix. This matrix cross-tabulated the reported sleep period categories for youth (rows) against those reported by their caregivers (columns), normalized to the youth reports, with distributions shown along each axis. Each cell within the matrix represents the normalized likelihood of agreement between youth and caregiver reports, providing insight into the concordance of perceptions regarding sleep duration. Frequencies within the matrix are indicated by a color scale from dark purple (low frequency) to light yellow (high frequency), highlighting areas of highest and lowest agreement. The diagonal emphasizes instances where reports from youth and caregivers exactly match, with the greatest concordance observed in the 8–9 hours sleep duration category.

Table 3. Descriptive Statistics and Comparison of the Continuous Measures of Sleep Duration, Sleep Period, Bedtime and WASO Between the Youth Reported and Objectively Collected Sleep Measures (Fitbit Charge 2)

Variable	Youth reported data Munich Chronotype Questionnaire (MCTQ)		Objective sleep measures Fitbit charge 2		Interclass correlation coefficient
	Weekday mean (SD)	Weekend mean (SD)	Weekday mean (SD)	Weekend mean (SD)	Value, CI [], *p*
Sleep duration (hours)	9.00 (1.59)	9.78 (2.27)	7.47 (0.68)	7.50 (0.92)	*0.186 [0.16, 0.22] p < .001*
Sleep period (sleep duration + WASO; hours)	9.08 (155)	9.83 (2.25)	8.43 (0.78)	8.47 (1.06)	*0.217, [0.19, 0.25] p < .001*
Bedtime in hours	09:11 pm (1.06)	10:33 pm (1.73)	10:45 pm (1.19)	11:08 pm (1.44)	*0.460 [0.43, 0.48] p < .001*
Wakefulness after sleep onset (WASO; minutes)	4.39 (14.20)	2.91 (10.82)	57.39 (11.37)	58.03 (13.94)	−0.015 [−0.05, 0.02] *p* = .837
Sleep onset latency (minutes)	21.63 (25.58)	17.63 (23.83)	6.92 (5.99)	6.67 (7.64)	−0.009 [−0.04, 0.02] *p* = .719
Caregiver reported hours of sleep	*9 ≤ SD*	*8 ≤ SD < 9*	*7 ≤ SD < 8*	*5 ≤ SD < 7*	*SD < 5*
%	33	45.8	16.5	4	0.1

The significant agreements are highlighted with italics.

During weekdays, the mean discrepancy in bedtime was −1.56 hours, with limits of agreement from −4.08 to 0.95 hours. Weekend data revealed a mean discrepancy of −0.59 hours and broader limits of agreement from −4.00 to 2.83 hours, indicating a larger variance in discrepancies during weekends compared to weekdays. Furthermore, the plots indicate that for bedtimes

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

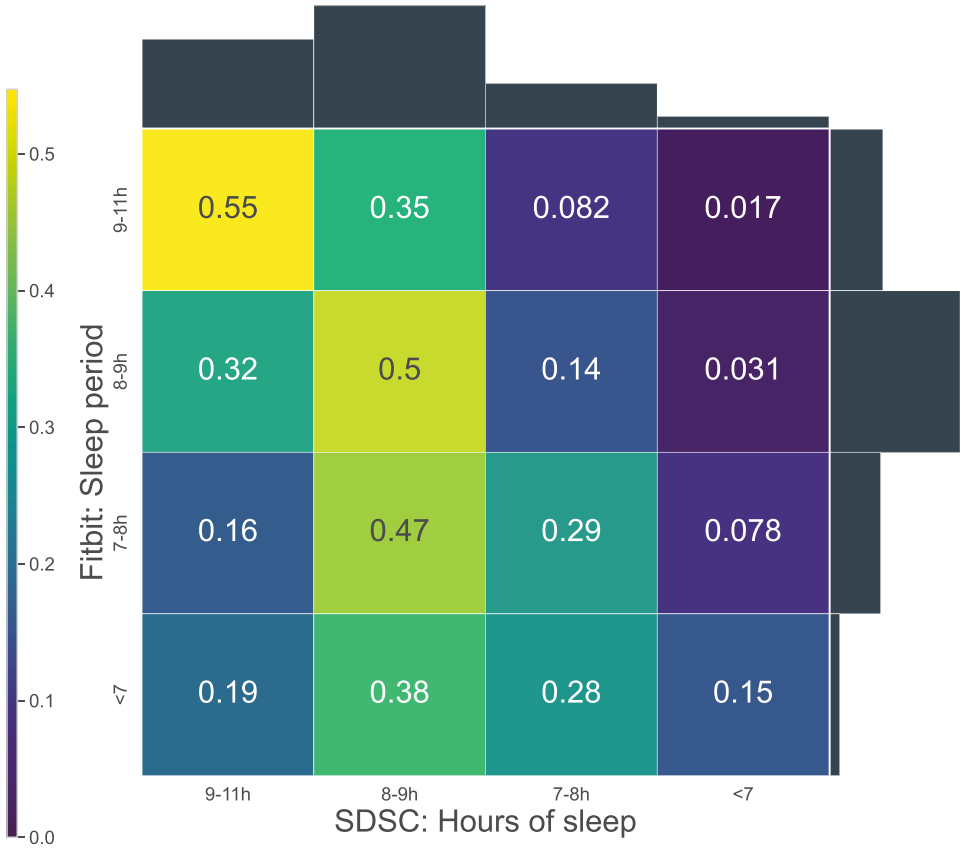


Figure 2. Agreement between caregiver-reported hours of sleep and objective measures. To assess the concordance between caregiver-reported sleep duration and objectively, we conducted an analysis using a normalized agreement matrix, normalized to Fitbit data. The matrix arranges Fitbit-measured sleep periods along the vertical axis and caregiver-reported durations along the horizontal axis, with each cell representing the likelihood of agreement normalized by rows. The likelihood values indicate the proportion of agreement between measured and reported sleep durations across four categories: 9–11 hours, 8–9 hours, 7–8 hours, and less than 7 hours.

substantially later than the central tendency (for bedtimes that were either significantly earlier or later than the typical range, which corresponds to the hours between 08:00 pm and 12:00 am), there is a trend toward increased discrepancy, potentially reflecting variability in weekend sleep schedules or reporting accuracy (Figure 4).

Assessing the effects of age and sex on congruency of youth self-report and Fitbit sleep measures

The Linear mixed-effects regression analysis revealed a significant main effect of age, where younger adolescents tended to report greater differences in sleep period compared to objective measurements ($\chi^2(1) = 6.98$, $p = .008$; Figure 5). Additionally, there was a significant age × sex interaction effect ($\chi^2(1) = 5.81$, $p = .015$), suggesting that this age-related trend in reported sleep discrepancy was driven by the lower discrepancy among older female adolescents (slope estimate = −0.94, $p < .01$). Our analysis found that differences in youth-reported and Fitbit-measured bedtimes were associated with the interaction between age and sex ($\chi^2(1) = 4.36$, $p = .03$); however, the slope estimate was not significant in either group.

Discussion

This study aimed to enhance our understanding of adolescent sleep patterns by analyzing and comparing sleep measures collected within a large and diverse sample of early adolescent participants from the ABCD study's year 2 follow-up using three distinct sources: the caregiver-reported Sleep Disturbance Scale for Children, the youth-reported Munich Chronotype Questionnaire, and Fitbit Charge 2 devices that collected sleep measures across a 21-day period.

The analysis highlighted a low agreement in sleep duration reported through the caregiver-reported Sleep Disturbance Scale for Children compared with youth-reported and objective Fitbit measures. Given that the Sleep Disturbance Scale for Children was designed to assess sleep disturbances comprehensively, our findings underscore its value within its intended context and caution against relying solely on sleep duration as an individual item. In contrast, sleep periods reported by youth showed a higher agreement with Fitbit measurements. We speculate that the stronger alignment between self-reported sleep period and Fitbit-recorded sleep period, compared to sleep duration, may be due to challenges in accurately estimating wakefulness after sleep onset.

The Bland-Altman analysis further revealed a proportional bias based on the length of sleep; agreement was relatively high among participants with average self-reported sleep durations but decreased among those with notably short or long self-reported sleep durations. This pattern aligns with previous findings that Fitbit's agreement with polysomnography decreases among poorer sleepers [24]. In our analysis, we do not have a true gold standard measure against which to compare results, therefore, while

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

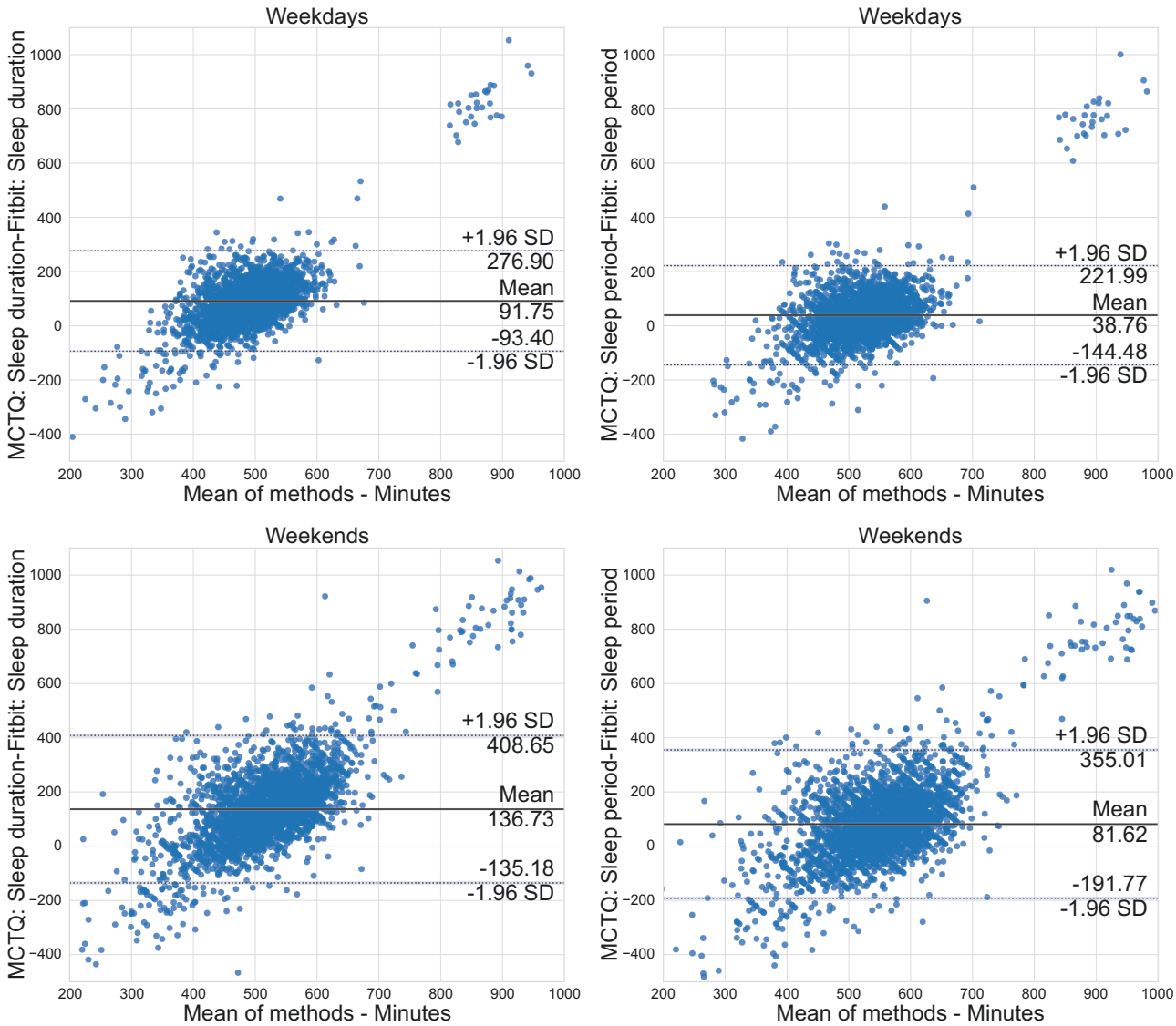


Figure 3. Bland-Altmann analysis for sleep durations and periods across weekdays and weekends. Panels A and B compare youth-reported sleep duration and period (MCTQ) with Fitbit-measured sleep duration and sleep period (SP), respectively, during weekdays. Panels C and D extend this comparison to weekends. The y-axis in each panel represents the difference scores between compared measures, illustrating the discrepancy between youth reports and Fitbit recordings. Solid lines mark the limits of agreement, defined as the mean difference ± 1.96 standard deviations (SD), encapsulating the expected range of discrepancy in 95% of cases.

we can show that there is low agreement, we cannot conclude whether Fitbit or the self-reported measure is a true measure of sleep. However, both the MCTQ and Fitbit measures differentiated between participants with clinically significant sleep disturbance and those who scored below this threshold from the sleep disturbance scale for children completed by caregivers, with the analysis showing statistically significant group differences.

Objective measures suggest later bedtimes than those reported by youth. This discrepancy could stem from either youth perception, where self-reports reflect the intention to sleep, or potential bias in objective measures, such as Fitbit, which may record later bedtimes due to continued movement after attempting to fall asleep. Bedtime changes during adolescence [40–42] are influenced by a variety of factors, including biological shifts in circadian rhythms, increased social and academic demands, and greater exposure to screen time before sleep. These changes contribute to later bedtimes and variations in sleep patterns as adolescents progress through the early, middle, and late stages of adolescence. Such transitions can lead to inconsistencies in self-reported bedtimes, as adolescents navigate these developmental milestones. Subjective-objective sleep discrepancies, particularly larger ones, may reflect underlying sleep problems such as insomnia [43–45] and could be associated with other health outcomes, such as neurodevelopmental disorders [46], highlighting an important area for future research as additional ABCD data becomes available.

Our findings indicate weak agreement between Fitbit and MCTQ reports of sleep onset latency and WASO, suggesting adolescents may struggle to accurately gauge the duration of wakefulness, yet could estimate their overall sleep period more reliably. This is consistent with previous research, which found that actigraphy estimates of wakefulness were substantially greater than sleep diary estimates [45, 47]. A WASO of over 30 minutes per night is often considered a criterion for sleep maintenance insomnia [48]. However, the discrepancies between subjective reports and objective sleep measures reflect a more complex

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

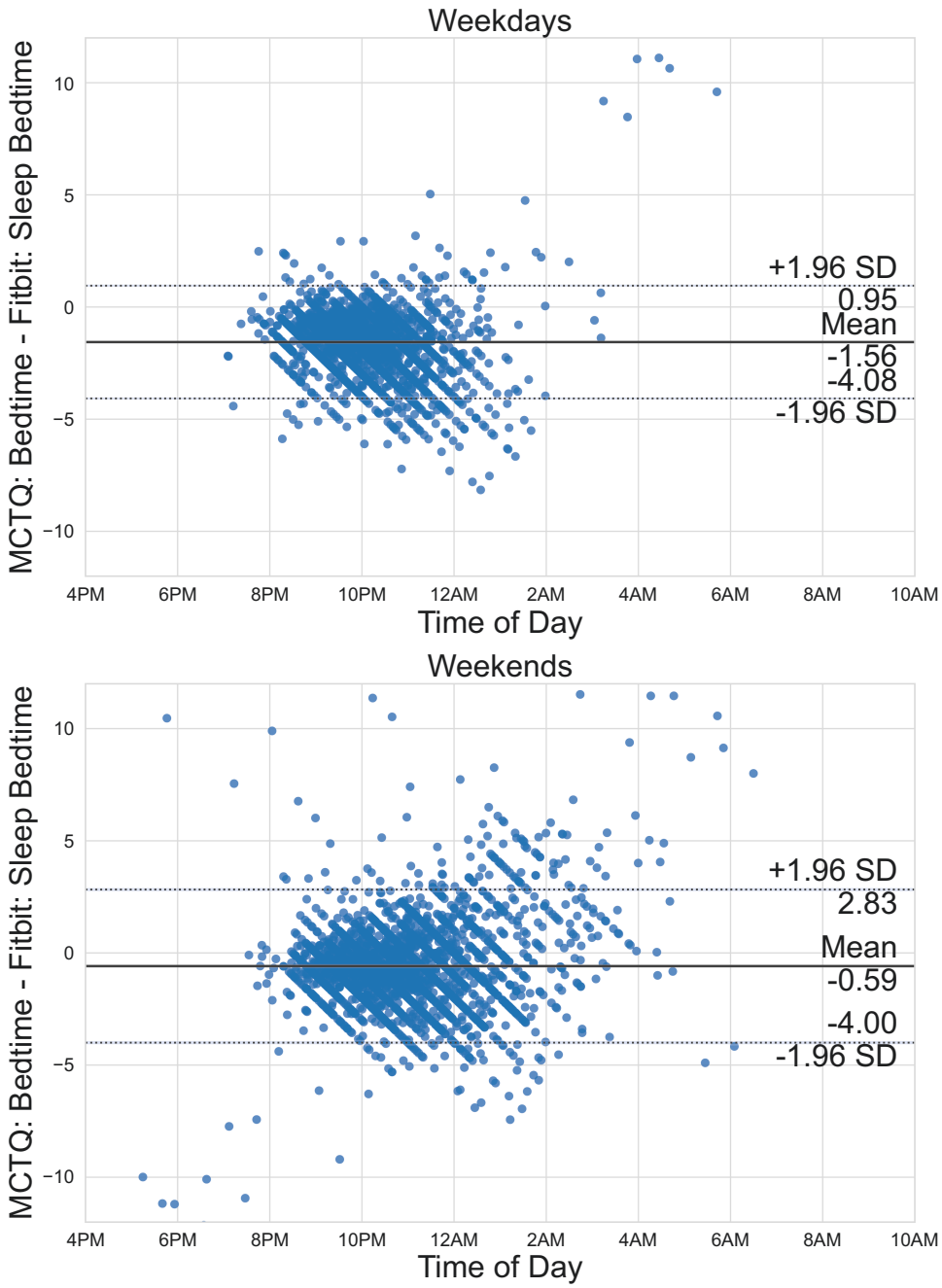


Figure 4. Bland-Altman analysis of bedtime comparisons. This set of Bland-Altman plots illustrates the agreement between youth-reported bedtimes (MCTQ) and those measured by Fitbit across different temporal contexts: weekdays (Top Panel) and weekends (Bottom Panel). The x-axis represents the bedtime in hours to capture the reported bedtimes, while the y-axis shows the differences between MCTQ and Fitbit measures (MCTQ—Fitbit). Solid lines denote the mean difference (bias), and the dashed lines indicate the limits of agreement at ±1.96 standard deviations (SD) from the mean. Notably, the mean differences suggest that youth report going to bed later than what Fitbit data indicates, with this discrepancy being more pronounced on weekends. The spread of points beyond the limits of agreement on weekends, compared to weekdays, suggests greater variability in bedtime reporting during non-school days. The proportional bias observed, where the discrepancy increases for earlier or later bedtimes, highlights a potential systematic difference that varies with bedtime.

interaction of different components. Fitbit, for example, might categorize movements during the night as awakenings even when they do not align with the individual's perception, which can contribute to an overestimation of wake time. Additionally, the fact that people often remain amnestic to brief nighttime awakenings [49] further complicates the accuracy of subjective reporting. Together, these findings emphasize the inherent differences in how sleep behaviors are captured across instruments, where overlapping terminology—such as "sleep duration"—may mask the distinct dimensions of sleep each instrument is measuring. The weaker agreement for sleep onset latency and WASO compared to sleep period may also be due to the relatively limited duration of wakefulness during the night. The total sleep period comprises a much larger proportion of the night, such that minor discrepancies don't have such a large effect. In contrast, if the proportion of WASO is low during the night, even minor discrepancies in estimates can have a proportionally greater impact on agreement.

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

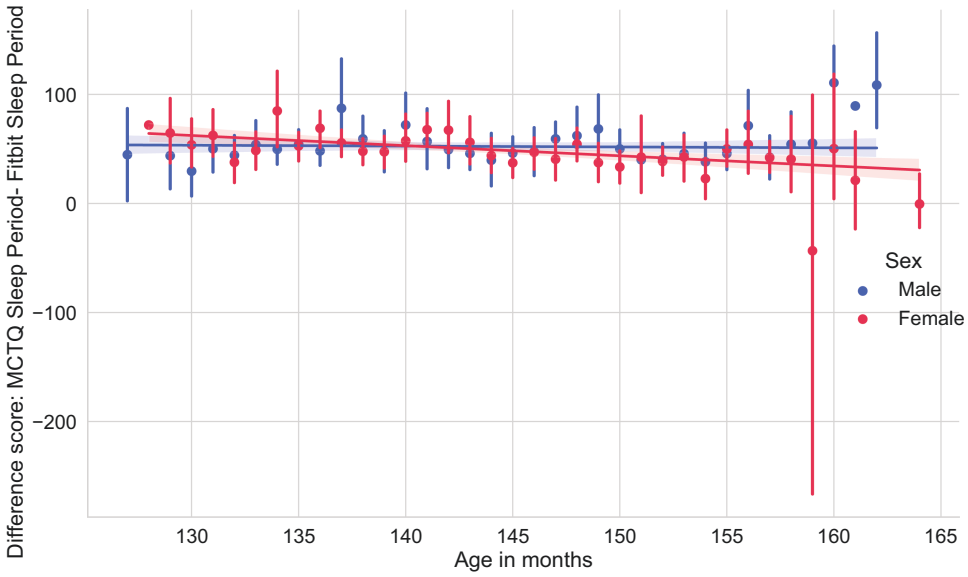


Figure 5. Association between the age in months and the difference score of the sleep period (youth-reported MCTQ Sleep period vs. Fitbit Sleep period; markers: mean ± CI), shown separately for males (blue) and females (red).

Consistently with prior studies [45, 50], our analysis revealed some indication of existing sex differences in the self-reports of sleep versus objective measures. Sex differences in sleep measures are known in adolescence, such as poorer sleep quality and greater prevalence of insomnia [13] in female adolescents. Sex differences in sleep duration and bedtimes have also been reported, although findings are inconsistent [5] and may differ according to weekdays versus weekends, as well as age [5, 51].

The results of our analysis suggest that while no single instrument provides a comprehensive view, each contributes valuable insights into different aspects of adolescent sleep [52]. Caregiver reports, for example, offer a window into potential sleep issues among younger adolescents, particularly when contextualized with additional information [53]. Youth-reported measures, while more closely aligned with objective sleep periods, may lack precision in estimating bedtimes and wakefulness and are prone to recall or social desirability bias [54]. Conversely, objective measures offer a complementary perspective on sleep, potentially enriching our understanding of sleep variability and stability, though their utility in diagnosing sleep issues remains to be fully elucidated. Objective sleep measures do not necessarily reflect diagnostic characteristics of sleep disorders, evidenced by empirical data showing that individuals with insomnia often report poor sleep quality, even when objective indices suggest their sleep patterns are relatively normal [55]. Nevertheless, our findings suggest that both self-reported and objective measures, as recorded by Fitbit, were able to differentiate between participants with and without sleep disturbances, as reported by the caregivers.

This study has several limitations that should be noted. While the ABCD sample is diverse, it is not fully representative of the US population, and the Fitbit substudy sample differs from the complete ABCD sample, particularly in terms of SES and demographic characteristics. These differences may limit the generalizability of our findings. Additionally, our analysis did not incorporate sleep disorders, clinical sleep assessments, or direct physiological sleep measures, such as polysomnography, which is considered the gold standard in sleep research but was not available in the ABCD study dataset. Caregiver and youth reports were not fully aligned in their time frames, as caregivers reported on the past 6 months while youth provided estimates of their typical sleep over the past 4 weeks. We acknowledge that caregivers were not specifically instructed to exclude WASO in their estimates of sleep duration, which may have introduced variability in how sleep duration was interpreted and reported. Fitbit measures were collected after the self-reported sleep measures, which may introduce misalignment between the two sources. However, our approach aimed to capture typical sleep patterns rather than specific nights, and we addressed this by using weighted averages of Fitbit sleep measures covering up to 21 days of sleep. In this way, we attempted to make our Fitbit measures more aligned with the self-report Munich Chronotype Questionnaire, which assesses "typical" sleep behavior over the past 4 weeks. This difference in time frames likely introduces discrepancies, which might be reduced if nightly sleep diaries were collected over the same period as the Fitbit data. Future studies with more synchronized data collection are needed to improve comparability. Additionally, although our focus was on sleep timing and duration measures, future investigations could benefit from exploring sleep quality measures to gain a more comprehensive understanding of sleep behaviors in adolescents. Additionally, while the Fitbit Charge 2 is used as an objective measure of sleep and has been shown to have adequate performance against polysomnography in adolescents [23], it is important to acknowledge that some of its derived measures, such as sleep onset latency, exhibit low accuracy. Commercial wearables are limited in their reliability for identifying intention to sleep (lights-off) and therefore for calculating sleep onset latency. Additionally, variability in age, gender, and sleep architecture introduces further inaccuracies of wearables in detecting transitions between wakefulness and sleep, including at sleep onset, and also during the night [56, 57]. Some researchers have suggested that pairing wearable data with self-reported lights-off times can lead to better estimates of sleep onset latency [28, 58]; however, due to feasibility constraints, the data collection did not include sleep diaries and daily self-reported lights-off times were not available in the ABCD study. There are also limitations for self-reported retrospective recall of sleep onset latency as well as other metrics [3]. Thus, we cannot say which measure is the true gold standard in this analysis and can only highlight the levels

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

of agreement and discrepancies between these different methods. This limitation suggests caution in interpreting these specific aspects of the Fitbit data and highlights the need for multimodal assessment strategies to capture the full spectrum of sleep patterns and disturbances accurately.

Conclusions

In conclusion, this study highlights the importance of a multifaceted approach to sleep assessment in adolescents. If variability measures or objectivity is important, device measurements should be obtained, while self-reports can be especially effective in identifying problems with sleep. Recognizing the limitations and strengths of each instrument allows for a more holistic understanding of adolescent sleep patterns, underscoring the value of integrating multiple sources of information to capture the complex nature of sleep among this population.

Supplementary material

Supplementary material is available at *SLEEP* online.

Disclosure Statement

Financial disclosure: O.K. and F.B. were supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, R01MH128959, and R01MH135492. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. *Nonfinancial disclosure:* The authors declare no competing interests.

Data Availability

The datasets analyzed during the current study are from the ABCD data repository, available at the NIMH Data Archive Digital Object Identifier (DOI) 10.15154/8873-zj65. DOIs can be found at https://dx.doi.org/10.15154/1518688.

References

1. Owens J; Adolescent Sleep Working Group. Insufficient sleep in adolescents and young adults: an update on causes and consequences. *Pediatrics*. 2014;**134**(3):e921–e932. doi:10.1542/peds.2014-1696
2. Tavernier R, Hill GC, Adrien TV. Be well, sleep well: an examination of directionality between basic psychological needs and subjective sleep among emerging adults at university. *Sleep Health*. 2019;**5**(3):288–297. doi:10.1016/j.sleh.2019.02.007
3. Silva GE, Goodwin JL, Sherrill DL, *et al*. Relationship between reported and measured sleep times: the sleep heart health study (SHHS). *J.Clin Sleep Med*. 2007;**3**(6):622–630.
4. Cespedes EM, Hu FB, Redline S, *et al*. Comparison of self-reported sleep duration with actigraphy: results from the Hispanic Community Health Study/Study of Latinos Sueño Ancillary Study. *Am J Epidemiol*. 2016;**183**(6):561–573. doi:10.1093/aje/kwv251
5. Albinni B, Baker FC, Javitz H, *et al*. Morning perception of sleep, stress, and mood, and its relationship with overnight physiological sleep: findings from the National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) study. *J Sleep Res*. 2023;**32**(5):e13886. doi:10.1111/jsr.13886
6. Mazza S, Bastuji H, Rey AE. Objective and subjective assessments of sleep in children: comparison of actigraphy, sleep diary completed by children and parents' estimation. *Front Psychiatry*. 2020;**11**:495. doi:10.3389/fpsyt.2020.00495
7. Colrain IM, Baker FC. Changes in sleep as a function of adolescent development. *Neuropsychol Rev*. 2011;**21**(1):5–21. doi:10.1007/s11065-010-9155-5
8. Feinberg I, Campbell IG. Sleep EEG changes during adolescence: an index of a fundamental brain reorganization. *Brain Cogn*. 2010;**72**(1):56–65. doi:10.1016/j.bandc.2009.09.008
9. Feinberg I. Changes in sleep cycle patterns with age. *J Psychiatr Res*. 1974;**10**(3-4):283–306. doi:10.1016/0022-3956(74)90011-9
10. Carskadon MA, Acebo C, Jenni OG. Regulation of adolescent sleep: implications for behavior. *Ann N Y Acad Sci*. 2004;**1021**(1):276–291. doi:10.1196/annals.1308.032
11. Crowley SJ, Wolfson AR, Tarokh L, Carskadon MA. An update on adolescent sleep: new evidence informing the perfect storm model. *J Adolesc*. 2018;**67**:55–65. doi:10.1016/j.adolescence.2018.06.001
12. Meltzer LJ, Mindell JA. Behavioral sleep disorders in children and adolescents. *Sleep Med Clin*. 2008;**3**(2):269–279. doi:10.1016/j.jsmc.2008.01.004
13. Yuksel D, Kiss O, Prouty DE, Baker FC, de Zambotti M. Clinical characterization of insomnia in adolescents – an integrated approach to psychopathology. *Sleep Med*. 2022;**93**:26–38. doi:10.1016/j.sleep.2022.03.010
14. de Zambotti M, Goldstone A, Colrain IM, Baker FC. Insomnia disorder in adolescence: diagnosis, impact, and treatment. *Sleep Med Rev*. 2018;**39**:12–24. doi:10.1016/j.smrv.2017.06.009
15. Tarokh L, Saletin JM, Carskadon MA. Sleep in adolescence: physiology, cognition and mental health. *Neurosci Biobehav Rev*. 2016;**70**:182–188. doi:10.1016/j.neubiorev.2016.08.008
16. Bruni O, Ottaviano S, Guidetti V, *et al*. The Sleep Disturbance Scale for Children (SDSC) Construct ion and validation of an instrument to evaluate sleep disturbances in childhood and adolescence. *J Sleep Res*. 1996;**5**(4):251–261. doi:10.1111/j.1365-2869.1996.00251.x
17. Roenneberg T, Wirz-Justice A, Merrow M. Life between clocks: daily temporal patterns of human chronotypes. *J Biol Rhythms*. 2003;**18**(1):80–90. doi:10.1177/0748730402239679
18. Hernandez LM, Kim M, Hernandez C, *et al*. Decoupling sleep and brain size in childhood: an investigation of genetic covariation in the adolescent brain cognitive development study. *Biol Psychiatry Glob Open Sci*. 2023;**3**(1):139–148. doi:10.1016/j.bpsgos.2021.12.011
19. Kiss O, Baker FC, Palovics R, Dooley EE, Pettee Gabriel K, Nagata JM. Using explainable machine learning and fitbit data to investigate predictors of adolescent obesity. *Sci Rep*. 2024;**14**(1):12563. doi:10.1038/s41598-024-60811-2
20. Karcher NR, Barch DM. The ABCD study: understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology*. 2021;**46**(1):131–142. doi:10.1038/s41386-020-0736-6
21. Zink J, O'Connor SG, Blachman-Demner DR, Wolff-Hughes DL, Berrigan D. Examining the bidirectional associations between sleep duration, screen time, and internalizing symptoms in the ABCD study. *J Adolesc Health*. 2024;**74**(3):496–503. doi:10.1016/j.jadohealth.2023.09.001
22. Depner CM, Cheng PC, Devine JK, *et al*. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. *Sleep*. 2020;**43**(2). doi:10.1093/sleep/zsz254
23. Godino JG, Wing D, de Zambotti M, *et al*. Performance of a commercial multi-sensor wearable (Fitbit Charge HR) in measuring physical activity and sleep in healthy children. *PLoS One*. 2020;**15**(9):e0237719. doi:10.1371/journal.pone.0237719

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025

24. De Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. *Chronobiol Int*. 2018;**35**(4):465–476. doi:10.1080/07420528.2017.1413578
25. Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. *Sleep Breath*. 2012;**16**:913–917. doi:10.1007/s11325-011-0585-y
26. Meltzer LJ, Walsh CM, Traylor J, Westin AM. Direct comparison of two new actigraphs and polysomnography in children and adolescents. *Sleep*. 2012;**35**(1):159–166. doi:10.5665/sleep.1608
27. de Zambotti M, Baker FC, Willoughby AR, *et al*. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiol Behav*. 2016;**158**:143–149. doi:10.1016/j.physbeh.2016.03.006
28. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. *Med Sci Sports Exerc*. 2019;**51**(7):1538–1557. doi:10.1249/MSS.0000000000001947
29. Lee XK, Chee NI, Ong JL, *et al*. Validation of a consumer sleep wearable device with actigraphy and polysomnography in adolescents across sleep opportunity manipulations. *J Clin Sleep Med*. 2019;**15**(9):1337–1346. doi:10.5664/jcsm.7932
30. Garavan H, Bartsch H, Conway K, *et al*. Recruiting the ABCD sample: design considerations and procedures. *Develop Cogn Neurosci*. 2018;**32**:16–22. doi:10.1016/j.dcn.2018.04.004
31. Wing D, Godino JG, Baker FC, *et al*. Recommendations for identifying valid wear for consumer-level wrist-worn activity trackers and acceptability of extended device deployment in children. *Sensors (Basel)*. 2022;**22**(23):9189. doi:10.3390/s22239189
32. Lau T, Ong JL, Ng BK, *et al*. Minimum number of nights for reliable estimation of habitual sleep using a consumer sleep tracker. *Sleep Adv*. 2022;**3**(1):zpac026.
33. Kim EH, Jenness JL, Miller AB, *et al*. Association of demographic and socioeconomic indicators with the use of wearable devices among children. *JAMA Netw Open*. 2023;**6**(3):e235681–e235681. doi:10.1001/jamanetworkopen.2023.5681
34. Hasler BP, Wallace ML, Graves JL, *et al*. Pre-pandemic circadian phase predicts pandemic alcohol use among adolescents. *J Sleep Res*. 2024;**33**(1):e13905. doi:10.1111/jsr.13905
35. Bagot K, Matthews SA, Mason M, *et al*. Current, future and potential use of mobile and wearable technologies and social media data in the ABCD study to increase understanding of contributors to child health. *Develop Cogn Neurosci*. 2018;**32**:121–129. doi:10.1016/j.dcn.2018.03.008
36. Vallat R. Pingouin: statistics in Python. *J Open Source Softw*. 2018;**3**(31):1026. doi:10.21105/joss.01026
37. Igolkina AA, Meshcheryakov G. semopy: a python package for structural equation modeling. *Struct Equ Modeling*. 2020;**27**(6):952–963. doi:10.1080/10705511.2019.1704289
38. Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011;**12**(null):2825–2830.
39. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. sparse matrix methods; linear mixed models; penalized least squares; Cholesky decomposition. *J Stat Softw*. 2015;**67**(1):48. doi:10.18637/jss.v067.i01
40. Carskadon MA. Sleep in adolescents: the perfect storm. *Pediatr Clin North Am*. 2011;**58**(3):637–647. doi:10.1016/j.pcl.2011.03.003
41. Carskadon MA. Evolution of sleep and daytime sleepiness in adolescents. *Sleep/wake disorders: Natural history, epidemiology, and long-term evolution*. 1983:201–216.
42. Carskadon MA, Kim H, Paula D, Anders TF, Litt IF, Dement WC. Pubertal changes in daytime sleepiness. *Sleep*. 1980;**2**(4):453–460.
43. Herbert V, Pratt D, Emsley R, Kyle SD. Predictors of nightly subjective-objective sleep discrepancy in poor sleepers over a seven-day period. *Brain Sci*. 2017;**7**(3):29. doi:10.3390/brainsci7030029
44. Trimmel K, Eder HG, Böck M, Stefanic-Kejik A, Klösch G, Seidel S. The (mis) perception of sleep: factors influencing the discrepancy between self-reported and objective sleep parameters. *J Clin Sleep Med*. 2021;**17**(5):917–924. doi:10.5664/jcsm.9086
45. Short MA, Gradisar M, Lack LC, Wright H, Carskadon MA. The discrepancy between actigraphic and sleep diary measures of sleep in adolescents. *Sleep Med*. 2012;**13**(4):378–384. doi:10.1016/j.sleep.2011.11.005
46. Munkhjargal O, Oka Y, Tanno S, *et al*. Discrepancy between subjective and objective sleepiness in adolescents. *Sleep Med*. 2022;**96**:1–7. doi:10.1016/j.sleep.2022.04.025
47. Tremaine RB, Dorrian J, Blunden S. Measuring sleep habits using the Sleep Timing Questionnaire: a validation study for school-age children. *Sleep Biol Rhythms*. 2010;**8**(3):194–202. doi:10.1111/j.1479-8425.2010.00446.x
48. American Psychiatric Association D, American Psychiatric Association D. *Diagnostic and statistical manual of mental disorders: DSM-5*. vol **5**. American psychiatric association, WA, DC; 2013.
49. Bonnet MH. Memory for events occurring during arousal from sleep. *Psychophysiology*. 1983;**20**(1):81–87. doi:10.1111/j.1469-8986.1983.tb00906.x
50. Albinni B, de Zambotti M, Iacovides S, Baker FC, King CD. The complexities of the sleep-pain relationship in adolescents: a critical review. *Sleep Med Rev*. 2023;**67**:101715. doi:10.1016/j.smrv.2022.101715
51. Lin LN, Chang L-Y, Hurng B-S, Wu C-C, Yen L-L, Chang H-Y. Sex differences in sleep patterns and changes in 7th to 12th graders: a longitudinal follow-up study in Taiwan. *Sleep*. 2018;**41**(3). doi:10.1093/sleep/zsx211
52. Bensen-Boakes D-B, Lovato N, Meaklim H, Bei B, Scott H. "Sleep-wake state discrepancy": toward a common understanding and standardized nomenclature. *Sleep*. 2022;**45**(10). doi:10.1093/sleep/zsac187
53. Mayne SL, Mitchell JA, Virudachalam S, Fiks AG, Williamson AA. Neighborhood environments and sleep among children and adolescents: a systematic review. *Sleep Med Rev*. 2021;**57**:101465. doi:10.1016/j.smrv.2021.101465
54. Wade NE, Ortigara JM, Sullivan RM, *et al*.; ABCD Novel Technologies Workgroup. Passive sensing of preteens' smartphone use: an adolescent brain cognitive development (abcd) cohort substudy. *JMIR Mental Health*. 2021;**8**(10):e29426. doi:10.2196/29426
55. Fernandez-Mendoza J, Calhoun SL, Bixler EO, *et al*. Sleep misperception and chronic insomnia in the general population: role of objective sleep duration and psychological profiles. *Psychosom Med*. 2011;**73**(1):88–97. doi:10.1097/PSY.0b013e3181fe365a
56. Danzig R, Wang M, Shah A, Trotti LM. The wrist is not the brain: Estimation of sleep by clinical and consumer wearable actigraphy devices is impacted by multiple patient-and device-specific factors. *J Sleep Res*. 2020;**29**(1):e12926. doi:10.1111/jsr.12926
57. Moreno-Pino F, Porras-Segovia A, López-Esteban P, Artés A, Baca-García E. Validation of Fitbit Charge 2 and Fitbit Alta HR against polysomnography for assessing sleep in adults with obstructive sleep apnea. *J Clin Sleep Med*. 2019;**15**(11):1645–1653. doi:10.5664/jcsm.8032
58. Van Hees VT, Sabia S, Anderson KN, *et al*. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. *PLoS One*. 2015;**10**(11):e0142533. doi:10.1371/journal.pone.0142533

Downloaded from https://academic.oup.com/sleep/article/48/7/zsaf065/8078870 by University of Witwatersrand user on 27 August 2025