



OPEN

Estimation bias and agreement limits between two common self-report methods of habitual sleep duration in epidemiological surveys

Maria Korman^{1✉}, Daria Zarina¹, Vadim Tkachev¹, Ilona Merikanto^{2,3}, Bjørn Bjorvatn^{4,5}, Adrijana Koscec Bjelajac⁶, Thomas Penzel⁷, Anne-Marie Landtblom^{8,9}, Christian Benedict¹⁰, Ngan Yin Chan¹¹, Yun Kwok Wing¹¹, Yves Dauvilliers¹², Charles M. Morin¹³, Kentaro Matsui¹⁴, Michael Nadorff¹⁵, Courtney J. Bolstad^{15,16}, Frances Chung¹⁷, Sérgio Mota-Rolim¹⁸, Luigi De Gennaro^{19,20}, Giuseppe Plazzi^{21,22}, Juliana Yordanova²³, Brigitte Holzinger²⁴, Markku Partinen^{25,26} & Cátia Reis^{27,28✉}

Accurate measurement of habitual sleep duration (HSD) is crucial for understanding the relationship between sleep and health. This study aimed to assess the bias and agreement limits between two commonly used short HSD self-report methods, considering sleep quality (SQ) and social jetlag (SJL) as potential predictors of bias. Data from 10,268 participants in the International COVID Sleep Study-II (ICOSS-II) were used. Method-Self and Method-MCTQ were compared. Method-Self involved a single question about average nightly sleep duration (HSD_{self}), while Method-MCTQ estimated HSD from reported sleep times on workdays (HSD_{MCTQwork}) and free days (HSD_{MCTQfree}). Sleep quality was evaluated using a Likert scale and the Insomnia Severity Index (ISI) to explore its influence on

¹Department of Occupational Therapy, Faculty of Health Sciences, Ariel University, Ariel, Israel. ²SleepWell Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland. ³Orton Orthopaedics Hospital, Helsinki, Finland. ⁴Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway. ⁵Norwegian Competence Center for Sleep Disorders, Haukeland University Hospital, Bergen, Norway. ⁶Institute for Medical Research and Occupational Health, Zagreb, Croatia. ⁷Sleep Medicine Center, Charité Universitätsmedizin Berlin, Berlin, Germany. ⁸Department of Medical Sciences, Neurology, Uppsala University, Uppsala, Sweden. ⁹Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden. ¹⁰Department of Pharmaceutical, Uppsala University, Uppsala, Sweden. ¹¹Li Chiu Kong Family Sleep Assessment Unit, Department of Psychiatry, The Chinese University of Hong Kong, Sha Tin, Hong Kong SAR, China. ¹²Sleep-Wake Disorders Unit, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier, INSERM Institute of Neurosciences of Montpellier, University of Montpellier, Montpellier, France. ¹³Centre de Recherche CERVO/Brain Research Center, École de Psychologie, Université Laval, Quebec, QC, Canada. ¹⁴Department of Clinical Laboratory, National Center Hospital, National Center of Neurology and Psychiatry, Kodaia, Japan. ¹⁵Department of Psychology, Mississippi State University, Starkville, MS, USA. ¹⁶South Texas Veterans Health Care System, San Antonio, TX, USA. ¹⁷Department of Anesthesia and Pain Management, Toronto Western Hospital, University Health Network, University of Toronto, Toronto, ON, Canada. ¹⁸Brain Institute, Physiology and Behavior Department and Onofre Lopes University Hospital, Federal University of Rio Grande do Norte, Natal, Brazil. ¹⁹Department of Psychology, Sapienza University of Rome, Roma, Lazio, Italy. ²⁰RCCS Fondazione Santa Lucia, Rome, Italy. ²¹Ircs Istituto Delle Scienze Neurologiche di Bologna, Bologna, Italy. ²²Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio-Emilia, Modena, Italy. ²³Institute of Neurobiology, Bulgarian Academy of Sciences, Sofia, Bulgaria. ²⁴Institute for Consciousness and Dream Research, Medical University of Vienna, Vienna, Austria. ²⁵Department of Clinical Neurosciences, University of Helsinki Clinicum Unit, Helsinki, Finland. ²⁶Helsinki Sleep Clinic, Terveystalo Healthcare Services, Helsinki, Finland. ²⁷Católica Research Centre for Psychological - Family and Social Wellbeing, Universidade Católica Portuguesa, Lisbon, Portugal. ²⁸Instituto de Medicina Molecular João Lobo Antunes, Universidade de Lisboa, Lisbon, Portugal. ✉email: maria.korman@ariel.ac.il; catia.reis@medicina.ulisboa.pt

estimation bias. HSD_{self} was on average 42.41 ± 67.42 min lower than $HSD_{MCTQweekr}$ with an agreement range within ± 133 min. The bias and agreement range between methods increased with poorer SQ. $HSD_{MCTQwork}$ showed less bias and better agreement with HSD_{self} compared to $HSD_{MCTQfree}$. Sleep duration irregularity was -43.35 ± 78.26 min on average. Subjective sleep quality predicted a significant proportion of variance in HSD_{self} and estimation bias. The two methods showed very poor agreement and a significant systematic bias, both worsening with poorer SQ. Method-MCTQ considered sleep intervals without adjusting for SQ issues such as wakefulness after sleep onset but accounted for sleep irregularity and sleeping in on free days, while Method-Self reflected respondents' interpretation of their sleep, focusing on their sleep on workdays. Including an SQ-related question in surveys may help bidirectionally adjust the possible bias and enhance the accuracy of sleep-health studies.

Habitual Sleep Duration (HSD) is a widely investigated parameter due to the number of highly reproducible associations to physical and psychological health outcomes^{1,2}. It is common to find that health outcomes of interest deteriorate as self-reported HSD deviates from the reference sleep norm interval³⁻⁷. Choosing the right tools to estimate HSD is challenging in epidemiological sleep research. The best method to self-report HSD is a sleep diary⁸, but it is generally non-applicable in surveys. Majority of the validated (*vis-a-vis* polysomnography (PSG)) sleep questionnaires, that are routinely used in clinical evaluation to reliably distinguish between individuals with and without sleep disorders, are relatively long⁹. To ensure good compliance and high response rates, tools that have minimal number of items are therefore prioritized in epidemiological surveys¹⁰.

Assessment of HSD in epidemiological surveys can include single questions such as "How many hours do you usually sleep at night?" (e.g., Pittsburgh Sleep Quality Index—PSQI, Self-Assessment of Sleep Survey—SASS)^{11,12}, which assumes that adults provide an accurate global and retrospective approximation of their sleep length. Other HSD estimation methods use two questions about sleep onset and offset times to estimate the sleep interval (e.g., Karolinska Sleep Questionnaire—KSQ, Basic Nordic Sleep Questionnaire—BNSQ, Munich Chronotype Questionnaire—MCTQ); these questions are asked separately for work and free days¹³⁻¹⁵. This method estimates sleep timing and crucial sleep metrics like social jetlag (SJL) and irregular sleep¹⁶. For example, inconsistent sleep timing is an important risk factor for metabolic abnormalities, even more significant than sleep duration¹⁷.

Various studies found weak-to-moderate correlations between single items of HSD and objectively measured sleep, however the agreement between different methods is poor—ranging between 2.0 and 3.5 h above and below the difference between the means^{1,18-22}. Also, sleep diaries and single-question HSDs, displayed either non-significant or weak associations¹. Self-assessment and time-in-bed duration calculated from habitual bedtime and wake time (rather than sleep onset and offset times), were recently reported to show disagreement with actigraphy-based sleep duration. Specifically, the single question provided a significant underestimate of HSD while the bed-wake interval agreed well with Time-in-Bed (TIB) but overestimated Total Sleep Time (TST)¹⁸. These biases and disagreements pose a significant challenge in the accurate assessment of contribution of HSD to physical and psychological health in survey research. Further, a recent methodological review showed that the variability in the questions relating to sleep, such as event definitions (e.g., "go to bed" vs. "fall asleep"), context (e.g., "habitual" vs. "work/free days") and timeframe ("typical night" vs. "recently") leads to discrepancies in HSD estimation by different self-report methods²³. Additionally, perceived sleep quality, insomnia symptoms and social schedules are important factors that can affect self-reported HSD¹⁹, but the extent of these effects have not been systematically quantified in large cohorts.

Sleep quality refers to the subjective experience of sleep, reflecting a number of quantifiable components of physiological sleep, such as depth of sleep (i.e., amount of slow-wave sleep), sleep continuity (i.e., wake after sleep onset, percentage of time awake, and number of awakenings) and additional internal or external factors (i.e., circadian profile, pain, stress)²⁴. Poor sleep quality can lead to overestimation or underestimation of sleep duration²⁵. A single question of overall sleep quality using a Likert scale is common in both experimental and epidemiological studies, with a verbal scale providing more stable estimation compared to a numerical scale^{10,12}. The Insomnia Severity Index (ISI) is sometimes also used as a proxy for sleep quality^{26,27}. Social time pressure refers to the demands and constraints of social obligations that may limit the sleep duration²⁸. In industrialized societies, people often experience a high social time pressure on workdays, and a large mismatch between internal biological and social times. This mismatch can be quantified by the difference between mid-sleep point on free and workdays and reflects irregularity of sleep timing, called Social Jet Lag (SJL)²⁹. Because self-report questions always encompass more than physiological sleep duration alone, evaluating the differences between common self-report methods used to assess HSD in surveys focusing on the potential predictors of the bias is important. The first objective of this study was to evaluate within-subjects estimation bias and the limits of agreement between two short self-report methods used to assess HSD in a large, global, heterogeneous sample of the International Covid Study II (ICOSS-II) project³⁰. The second objective of this study was to address the contribution of subjective Sleep Quality and Social Time Pressure to estimate the HSD bias. The contribution of Sleep Quality was validated *vis-à-vis* Insomnia Severity Index (ISI)—one of the most widely used tools to assess sleep problems in clinical and community samples²⁷.

Results

The sample consisted of 10,268 participants with a mean age of 43.16 ± 16.80 years (Mean \pm standard deviation) and 68.3% were female. Demographic descriptive in Table 1.

Variables	Sample total n = 10,268
Age, years	43.2 ± 16.8
18–34	3636 (35.4%)
35–39	2771 (27.0%)
50–64	2522 (24.6%)
65–99	1339 (13.0%)
Gender, female	7012 (68.3%)
Country	
Austria	527 (5.1%)
Brazil	197 (1.9%)
Bulgaria	341 (3.3%)
Canada	464 (4.5%)
Croatia	477 (4.6%)
France	305 (3.0%)
Germany	445 (4.3%)
Finland	1181 (11.5%)
Hong Kong	243 (2.4%)
Israel	352 (3.4%)
Italy	786 (7.7%)
Japan	2581 (25.1%)
Norway	491 (4.8%)
Portugal	408 (4.0%)
Sweden	688 (6.7%)
USA	744 (7.2%)
Other	38 (0.4%)
Ethnicity	
White (Caucasian)	6626 (64.9%)
Asian	2713 (26.6%)
African	153 (1.5%)
Hispanic	212 (2.1%)
Other	503 (4.9%)
Marital status	
Single	3351 (32.6%)
Married/relationship	6026 (58.7%)
Divorce/separated	707 (6.9%)
Widowed	179 (1.7%)
Education	
Primary/elementary/lower secondary school	295 (2.9%)
Secondary/high/vocational school	3184 (31.9%)
University/college or above	6512 (65.2%)
Present work	
Student	1903 (18.5%)
Regular day work	5119 (49.9%)
Irregular day work/freelancer/artist/research	989 (9.6%)
Unemployed	356 (3.5%)
Retired	1205 (11.7%)
At home (no salary)	605 (5.9%)
Temporary laid off	91 (0.9%)
Financial burden	
Not at all	4794 (4.8%)
A little/somewhat	3825 (3.8%)
Much/very much	1467 (1.4%)
Body Mass Index (BMI)	25.0 ± 6.3
Insomnia Severity Index (ISI)	8.5 ± 6.1
0–7; no clinical insomnia	5136 (50.3%)
8–14; subthreshold insomnia	3249 (31.8%)
15–21; moderate insomnia	1502 (14.7%)
Continued	

Variables	Sample total n = 10,268
22–28; severe insomnia	320 (3.1%)
Sleep quality	
Well	2059 (20.1%)
Rather well	2994 (29.2%)
Neither well nor badly	2658 (25.9%)
Rather badly	1958 (19.1%)
Badly	599 (5.8%)
Habitual sleep duration self-report (HSD _{self}), min	418.9 ± 77.2
Habitual sleep duration MCTQweek (HSD _{MCTQweek}), min	461.4 ± 75.1
Habitual sleep duration MCTQwork (HSD _{MCTQwork}), min	449.0 ± 81.1
Habitual sleep duration MCTQfree (HSD _{MCTQfree}), min	492.3 ± 87.7
Social jetlag (SJL), min	56.5 ± 62.2

Table 1. Socio-demographic characteristics and sleep measures of the sample. Mean ± SD or frequency (% of group total).

Estimation of habitual sleep duration bias and the agreement between methods

Distributions of HSDs from both methods are shown in Fig. 1a, with mean HSD_{self} being shorter (418.9 ± 77.2) than HSD_{MCTQweek} (461.4 ± 75.1). A paired t-test was used to quantify the within-subject difference between methods. A systematic HSD estimation bias was observed ($t = -63.07$, $df = 10,267$, $p < 0.001$). The mean bias was -42.41 ± 67.42 min (95% CI of the difference: -43.72 to -41.11) and had a normal distribution (Fig. 1b), though HSD_{self} and HSD_{MCTQweek} were significantly positively correlated ($\rho = 0.604$, $p < 0.001$, weighted by age).

The level of agreement between the two HSD assessment methods is visualized using the Bland–Altman plot in Fig. 1c. As neither of the two methods is a “reference”, the bias was compared with the means of the HSD_{self} and the HSD_{MCTQweek} values. To assess whether the bias (represented by the gap between the X axis, and the mean line (blue)) is stable through the whole range of values, a linear regression line (red) was fit to the HSD data points. A Pearson test demonstrated a significant negligible slope ($k = 0.034$, $\text{Beta} = 0.02$, $p = 0.03$). Finally, the limits of agreement between methods were calculated as: Upper limit [$\bar{d} - 1.96 s$] = $-42.41 - (1.96 \times 67.42) = 175$;

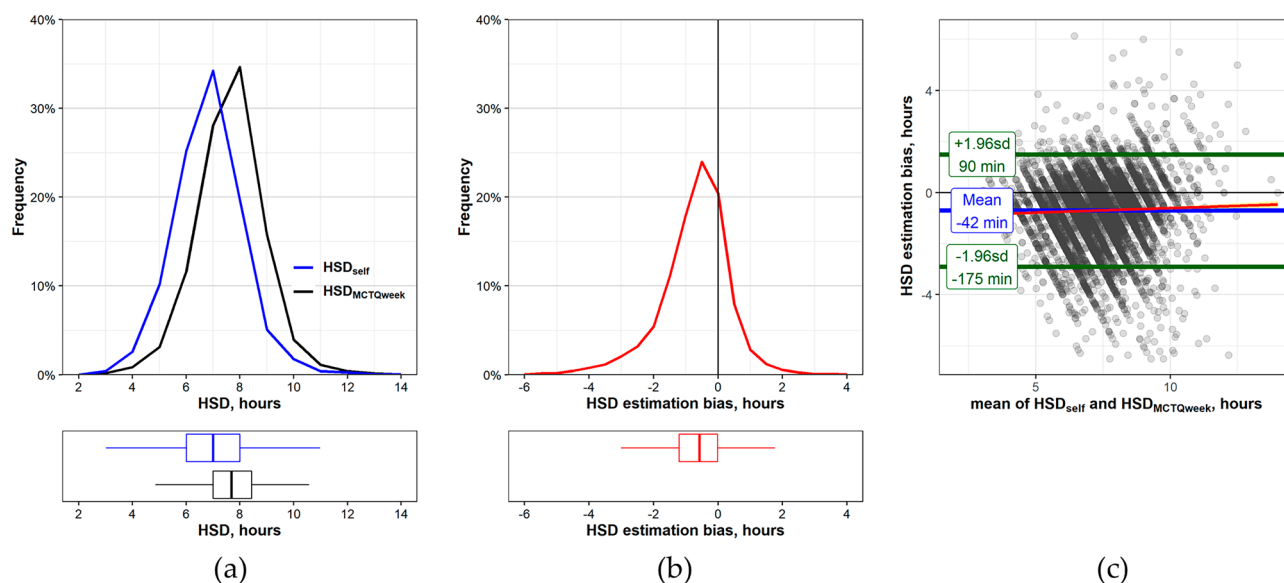


Figure 1. Habitual sleep duration (HSD) by Method-Self and Method-MCTQweek. (a) Upper panel—HSD distributions, percent from group total by method: blue line—HSD_{self}, black line—HSD_{MCTQweek}, 1-h bin. Lower panel—Boxplots of individual HSD by method. Whiskers—max and min values, box borders—75th and 25th percentiles, line through the box—median. (b) Upper panel—HSD estimation bias values distribution, percent from group total, 30-min bin. Lower panel—Boxplots of individual HSD estimation bias values. (c) Bland–Altman plot comparing Method-Self and Method-MCTQweek. The blue line indicates that the Method-Self sleep duration estimates are on average 42 min shorter than Method-MCTQ estimates. The green lines indicate the 95% limits of agreement ($\pm 1.96SD$ s). The linear regression line (red) shows that the HSD estimation bias is stable through the whole range values. The two methods only agree to within ± 2.2 h.

Lower limit $[\bar{d} + 1.96\text{ s}] = -42.41 + (1.96 \times 67.42) = 90$. Altogether, the two methods only agreed within ± 133 min, in other words, the HSD_{self} may be 90 min above or 175 min below the $\text{HSD}_{\text{MCTQweek}}$.

A simple regression model using weighted joint distribution of gender and age by country showed that age was not a significant predictor of the HSD bias ($F(1, 10,256) = 2.77, p = 0.096, \text{Beta} = 0.016$). However, women had significantly larger HSD bias than men ($t = 4.55, p < 0.001, \text{mean difference} = 6.6$ min), but with a negligibly small effect size (Cohen's $d = 0.097$).

Sleeping well? The HSD estimation bias and the agreement of the methods depend on subjective sleep quality

HSD estimated by both methods negatively correlated with participants' subjective Sleep Quality, with sleep quality demonstrating a stronger relation to HSD_{self} (Pearson correlations weighted by age: $\rho = -0.334, p < 0.01, \rho = -0.134, p < 0.01; \text{HSD}_{\text{self}}$ and $\text{HSD}_{\text{MCTQweek}}$, respectively). Although the two methods are presumably estimating the same construct, using the Fisher r -to- z transformation we found that the two correlation coefficients were also significantly different ($z = -15.71, p < 0.01$). The correlation between HSD estimation bias and subjective Sleep Quality was also significant ($\rho = -0.207, p < 0.01$).

To quantify the dependence of the agreement between the two methods in reference to subjective sleep quality, given the large sample size of the ICOS-2 study, HSD bias for each 5 Sleep Quality groups was separately analyzed. One-way ANOVA showed that the estimation bias became more negative as the sleep quality decreased ($F(4, 10,256) = 105.16, p < 0.001$). The results are summarized in Fig. 2. The minimal HSD estimation bias value (-26.69 ± 58.10 min) and the narrowest range of agreement between methods (± 114 min) were in the group sleeping "well". The estimation bias and range of agreement became progressively larger with poorer sleep quality. HSD estimation bias in the group sleeping "badly" reached a maximum value of (-79.97 ± 97.29 min) with a range of agreement of ± 191 min. Post-hoc pairwise comparisons with Bonferroni corrections demonstrated significant distinctions between each of the five sleep quality groups (see supplementary information SI-Table S.1), suggesting underestimation of HSD_{self} relative to $\text{HSD}_{\text{MCTQweek}}$ increases incrementally.

Workdays or freedays? The HSD estimation bias and the agreement of methods depends on social time pressure (workdays/free days)

Most participants reported irregular sleep durations across the week. The mean difference between $\text{HSD}_{\text{MCTQwork}}$ and $\text{HSD}_{\text{MCTQfree}}$ was -43.35 ± 78.26 min (449.0 ± 81.1 and 492.3 ± 87.7 min, respectively; paired t -test,

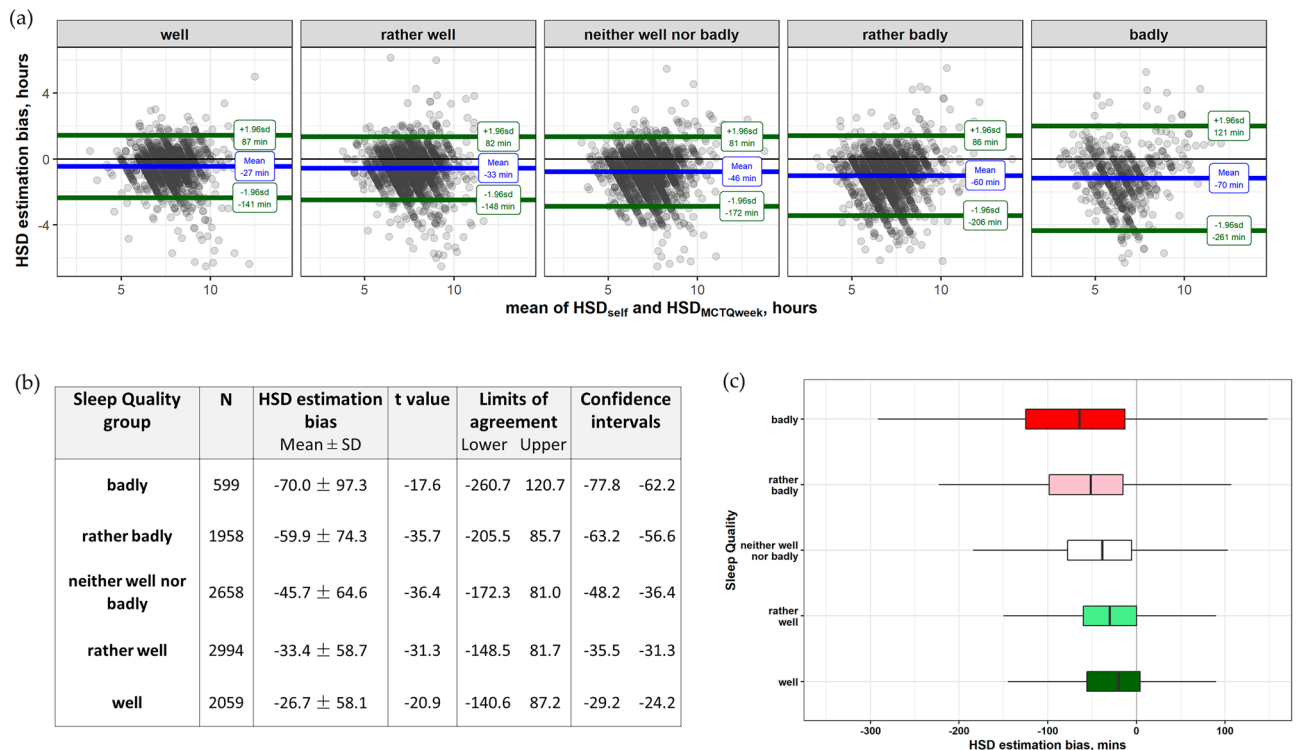


Figure 2. HSD estimation bias by Sleep Quality. (a) Bland–Altman plots comparing Method-Self and Method-MCTQweek in five Sleep Quality groups. The blue lines (mean per Sleep Quality group) indicate that underestimation of HSD_{self} relative to $\text{HSD}_{\text{MCTQweek}}$ increased incrementally as the Sleep Quality worsened: from -27 min in the “well” sleeping group to -70 min in the “badly” sleeping group. The 95% limits of agreement (± 1.96 SDs, green lines) also become progressively further apart. (b) Statistics of the Bland and Altman plots. (c) Boxplots of HSD estimation bias by Sleep Quality. Notations as in Fig. 1c.

$t(10,267) = -56.13, p < 0.001$). Accordingly, the distribution of the difference between $HSD_{MCTQwork}$ and $HSD_{MCTQfree}$, with majority of respondents reporting longer sleep duration during free days (percentiles in minutes: 25th = 0, 50th = 30, 75th = 75).

Next, we tested the hypothesis that $HSD_{MCTQwork}$ would demonstrate a smaller estimation bias and better agreement with HSD_{self} as compared to $HSD_{MCTQfree}$. The mean estimation bias for the $HSD_{MCTQwork}$ was smaller than the $HSD_{MCTQfree}$ (-30 min, and -73 min, respectively, Fig. 3a). Further, the agreement limits with the HSD_{self} were similar to the limits of the $HSD_{MCTQweek}$ but better than in $HSD_{MCTQfree}$ (± 140 min vs. ± 169 min, respectively, Fig. 3b,c). The observation that Sleep Quality groups were significantly different from each other was replicated also in $HSD_{self} - HSD_{MCTQwork}$ and $HSD_{self} - HSD_{MCTQfree}$ comparisons (SI-Tables S.2, S.3).

The mean SJL of the sample was 56.5 ± 62.2 min (SJL percentiles, in minutes: 25th = 15, 50th = 45, 75th = 90). There were no significant differences in SJL between the Sleep Quality groups (One-way ANOVA $p = 0.205$).

The combined contribution of sleep quality and social time pressure on HSD estimation bias

Having established the effects of Sleep Quality and Social Time Pressure on HSD estimation bias, we presumed that their combination may demonstrate conditions under which the bias is minimal and the agreement between the methods is most reliable. One-way ANOVAs showed that the estimation bias became more negative in both methods as the sleep quality decreased ($F(4, 10,263) = 84.312, p < 0.001$; $F(4, 10,263) = 79.65, p < 0.001$; Method- $MCTQ_{work}$ and Method- $MCTQ_{free}$, respectively). Post-hoc pairwise comparisons with Bonferroni corrections for $HSD_{MCTQwork}$ showed that “well” and “rather well” Sleep Quality groups did not differ, while all other groups showed significant differences (SI-Table S.4). In contrast, for $HSD_{MCTQfree}$, “rather badly” and “badly” Sleep Quality groups were not significantly different from each other, while all other groups showed significant differences (SI-Table S.5). The “well” and “rather well” sleeping groups during workdays showed the best parameters: the mean HSD estimation bias was only -15.81 ± 62.77 min and the two methods agreed within ± 114 min (Fig. 4a,b).

Weighted least squares stepwise regressions were conducted to examine the extent to which Sleep Quality and Social Time Pressure (represented by SJL) explained the variance in different HSDs and the HSD estimation bias itself. The main model had 5 predictors: Sleep Quality, SJL, age, gender, and BMI. Gender and age by country distribution was used for weighting. The model explained 13.7% of the HSD_{self} variance, 4.2% of the $HSD_{MCTQweek}$ variance, 3.6% of the $HSD_{MCTQwork}$ variance, 10.8% of the $HSD_{MCTQfree}$ variance and 6.9% of the variance in the HSD estimation bias. Leading predictor in all models, except $HSD_{MCTQfree}$, was Sleep Quality, with HSD_{self} demonstrating the largest dependence (12.5% vs. 2.1% vs. 2.1% and 6.2%; $HSD_{self} - HSD_{MCTQweek}$ and $HSD_{MCTQwork}$ and HSD estimation bias, respectively). Leading predictor of $HSD_{MCTQfree}$ was SJL (7.4%). Age and gender were significant predictors in most models but explained less than 1% of the variance for all (statistical details in supplementary information SI-Table S.6).

Comparison between the contributions of sleep quality and ISI score to HSD estimation bias

The contribution of subjective Sleep Quality to the models was assessed using the ISI score, a clinical index of insomnia symptoms severity. Weighted least squares stepwise regressions were re-run with the ISI score used instead of the Sleep Quality and the other four predictors similar to the original model. The variance in HSD_{self} , $HSD_{MCTQweek}$ and $HSD_{MCTQwork}$ was primarily explained by the ISI score but the models were less robust (8.4%, 1.4% and 1.5%, respectively, (see details in supplementary information SI-Table S.7). See full statistical details in SI-Table S.7 and SI-Fig. S.1 for the distribution of the HSD estimation bias values by ISI categories. Finally, a model including both Sleep Quality and ISI continuous score as predictors (and SJL, gender, age, and BMI), explained 6.9% of the variance in HSD estimation bias. Note that the ISI score was the least robust contributor

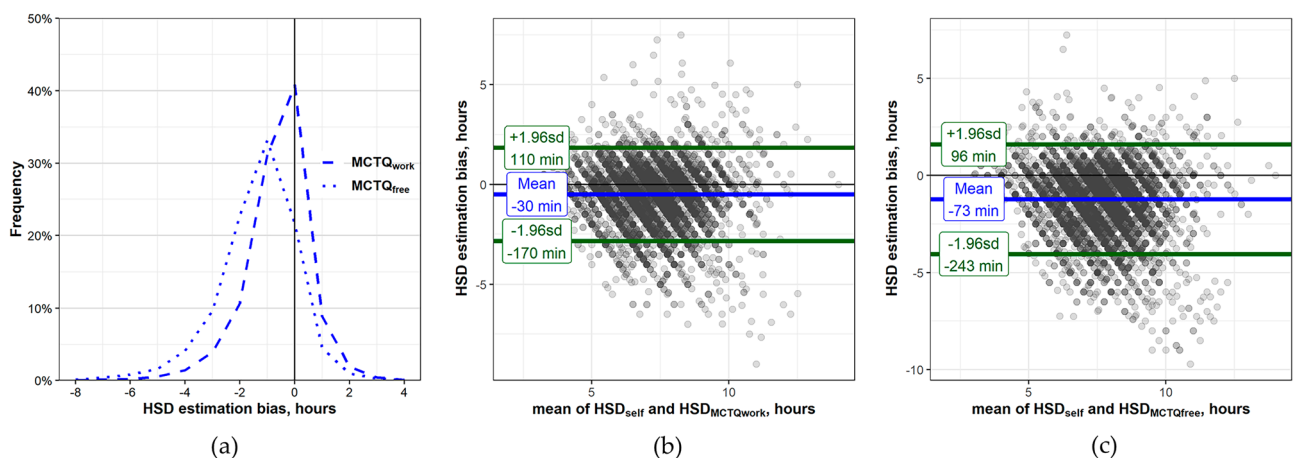


Figure 3. Estimation bias differences between Method-MCTQwork and Method-MCTQfree. (a) Habitual sleep duration estimation bias values distribution for workdays and free days, percent from group total. Dotted line— $HSD_{MCTQfree}$, dashed line— $HSD_{MCTQwork}$. (b) Bland–Altman plot comparing Method-Self and Method-MCTQwork. Notations as in Fig. 1c. The two methods agree within ± 2.3 h. (c) Bland–Altman plot comparing Method-Self and Method-MCTQfree. The two methods agree within ± 2.8 h. Notations as in Fig. 1c.

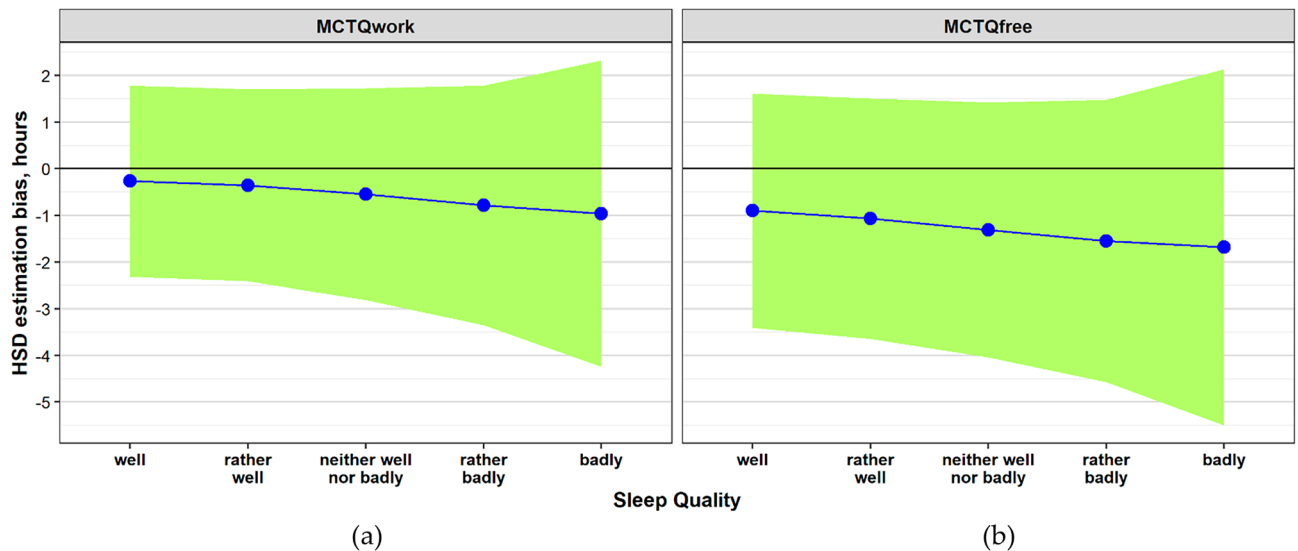


Figure 4. HSD estimation bias as a function of Sleep Quality by (a) Method-MCTQwork versus (b) Method-MCTQfree. HSD estimation bias values are smaller (closer to zero line) in the Method-MCTQwork as compared with the Method-MCTQfree in all Sleep Quality groups. Green areas around the means—the 95% limits of agreement (± 1.96 SDs). Note that the Method-MCTQwork narrower agreement ranges in all Sleep Quality groups as compared to the Method-MCTQfree.

accounting only for 0.1% of the variance (SI-Table S.8), demonstrating that ISI score was practically redundant as a predictor of the HSD estimation bias.

Discussion

It is not clear which self-report method to measure sleep duration can be advised to be used with confidence in large online surveys, since great discrepancies are systematically observed between different methods. Our findings in a large international sample of 10,268 participants also showed poor agreement range (± 133 min), and also indicated systematic and high estimation bias (42.41 ± 67.42 min) between HSD derived from sleep onset and offset and a single question. Thus, for a given person, self-reported sleep duration (HSD_{self}) will be almost always lower than self-reported sleep interval (according to $HSD_{MCTQweek}$). For example, if somebody says they sleep 7.5 h a night that means that he/she would estimate their sleep interval as $\sim 8h12min$ ($+ 42$ min), on average, but the accuracy of this estimation will be very low (± 133 min).

While inaccuracy and problems with face validity of different methods are well recognized in the literature, differences in the dimensionality of the self-report methods, factors that contribute to the poor agreement between them and explain the bias, at least partially, were less studied^{18,19,23}. If HSD is systematically under- or overestimated depending on the question, the associations of the health outcomes with sleep duration will also be systematically inflated or flattened³¹. Our findings showed that subjective sleep quality was a strong driver of the estimation bias, the bias almost tripled from the best to worst Sleep Quality group (from 26.69 ± 58.10 to -79.97 ± 97.29 min). Furthermore, estimation bias changed incrementally with decreasing sleep quality. We also showed that a single question addressing sleep quality contributed to the model explaining the HSD estimation bias more than a multi-item insomnia symptoms severity score. Moreover, having both Sleep Quality and ISI scores as predictors of HSD estimation bias was, in fact, redundant. Sleep quality was also a leading predictor of HSD_{self} , $HSD_{MCTQweek}$ and $HSD_{MCTQwork}$, while SJL was a leading predictor of $HSD_{MCTQfree}$. The quantitative estimation of the bias between methods can be used bi-directionally to estimate HSD from one method to the other, if a subjective sleep quality parameter is available.

Our findings therefore indicate that assessing HSD with a single question, or HSD from sleep onset and offset, may capture distinct aspects of sleep duration. The $HSD_{MCTQweek}$ was only subtly influenced by sleep quality, while HSD_{self} and the estimation bias were profoundly sensitive to it. Conversely, the single-question method accounts for poor sleep, but lacks sensitivity to sleep rebound on free days. This may happen because people tend to report the most representative days of the week (i.e., workdays), and lower sleep satisfaction during workdays. This makes the single-question method more susceptible to sleep misperception. Sleep misperception has been found to vary a lot in people from the general population, in patients with insomnia³², hypersomnia³³ and obstructive sleep apnea³⁴. These results are in agreement with previous findings, where single questions about sleep duration and sleep quality using the PSQI tool were shown to represent workdays, whereas when the same PSQI questions were asked separately, participants from the general population³⁵ had better sleep during free days as well as in clinical populations, and this difference was mediated by SJL³⁶. Women had a slightly higher HSD estimation bias compared to men (~ 6 min), and this finding may be explained by the fact that women tend to report lower sleep quality³⁷. Interestingly, although sleep duration changes through life³⁸, age had no effect on the HSD estimation bias, suggesting that underestimation of HSD_{self} relative to $HSD_{MCTQweek}$ is a stable phenomenon across ages related to sleep quality.

Several limitations exist when interpreting our results. Among those, it was a convenience sample that was collected during COVID-19 pandemic, included unusual participants with a novel health profile of long COVID, and had a clear overrepresentation of women (68.3%). In particular, the data collection period was associated with many changes in the social and personal lives of people across participating countries but note that data was not collected during confinement. Sleep–wake habits during the pandemic were adaptively changing worldwide, with many people working and studying from home^{39–41}. Additionally, this study was designed to engage participants who may have had COVID-19 and suffer from symptoms of long COVID^{25,30}. Indeed, 9.1% of the sample reported symptoms of long COVID when enrolled in the ICOSS-II study. However, the sensitivity analyses in a sub-group of participants with long-COVID symptoms and in a subgroup of older adults supported the conclusion that HSD bias between methods is a stable trait primarily related to Sleep Quality (see details in the “Methods” and Supplementary Materials sections). Altogether, the web-based survey’s generalizability is limited, but may be partially offset by the large sample size and uniform data acquisition period.

Concerns about self-reported sleep duration accuracy in surveys are longstanding^{19,42,43}, even prompting suggestions to exclude it from epidemiological studies⁴⁴. Nevertheless, in large-scale field sleep studies the use of self-report tools is often the only possible option, like in the case of the COVID-19 pandemic^{28,30}. Over the last years, many studies showed associations between self-report measures with chronic diseases and mental health^{5–7,45}, identifying risk factors, screening for sleep disorders, monitoring changes in the population habits, and understanding the broader public health implications. We believe that researchers using measures of sleep duration based on self-reports should be aware of the meanings and limitations associated with each method, as well as about their disagreement without assuming that all of them reflect physiological sleep to the same extent and strive to add objective measurements of sleep duration or sleep diary when possible.

To conclude, the two methods showed very poor agreement and a significant systematic bias, both worsening with poorer subjective sleep quality. The method using self-reported sleep onset and offset times provides a “raw” calculation of the sleep intervals for work and free days, accounts for irregularities in sleep duration and timing but is inherently insensitive to the frequency and length of awakenings^{46,47}. The accuracy of sleep intervals estimations would benefit from inclusion of a wakefulness after sleep onset item, as in Evanger et al.⁴⁸. The single-question sleep duration assessment was found to be associated with sleep quality, and thus may reflect in part how respondents perceive their sleep. However, this method is inherently insensitive to the sleep rebound that occurs on days off^{31,49}. We suggest that assessing sleep duration and subjective sleep quality separately for workdays and free days may improve the design of future studies^{35,36}. This can be done using either single or two-question approach, in accordance with the specific objectives of the study and, when possible, should include objective measures of sleep. Future studies should evaluate whether including items assessing sleep quality (e.g., single question) and wakefulness after sleep onset may facilitate the implementation of adjustments accounting for potential biases between HSD estimation methods.

Methods

Data collection

This study used data from the International Covid Study II (ICOSS-II)³⁰, which is an international collaboration between sleep and circadian rhythm experts. Using a web-based anonymous survey, ICOSS-II took place between May to December 2021 in parallel across the following 16 countries using translations to local languages: Austria, Brazil, Bulgaria, Canada, Hong Kong/China, Croatia, Finland, France, Germany, Israel, Italy, Japan, Norway, Portugal, Sweden, USA. The survey used Qualtrics and Redcap platforms. The study conforms to recognized standards by the Declaration of Helsinki. After a brief explanation of the study, the survey was available to participants after obtaining their informed consent to be part of the study. All investigators obtained local ethical committee (REB) approval when applicable (detailed list in supplementary material Table S.8). Due to the anonymous nature of the survey, REB permissions were exempted in some countries.

A total of 16,899 participants opened the link to the ICOSS questionnaire, and 15,859 had valid data. For this study we excluded shift/night workers and subjects reporting severe health conditions (atrial fibrillation, heart failure, stroke, other heart conditions, chronic obstructive pulmonary disease, kidney failure, cancer, immunosuppressive treatment, ongoing Covid-19). For quality control reasons, we excluded participants with HSD < 2.5 h or > 16 h (in either HSD_{self} and HSD_{MCTQfree}), with discrepancy in sleep duration estimation of more than 400 min between the two methods, or with missing data in sleep duration and sleep quality parameters. We had a final sample of 10,268 individuals.

Sleep assessment items and measures

HSD times were assessed twice for each participant using two methods: Method-Self assessment was based on a single-question (i.e., “How many hours per night you have been sleeping on average CURRENTLY?”) in the format hh:mm (HSD_{self}). The Method-MCTQ used an adapted version of the Munich Chronotype Questionnaire (μMCTQ). The questions were referring to sleep onset and offset timings (reported in 24 h local time format) (i.e., “At what time do you usually fall asleep at work/free days CURRENTLY?”, “At what time do you usually wake up at work/free days CURRENTLY?”). Separate reports were obtained for workdays and free days, enabling calculation of HSD during workdays and free days (HSD_{MCTQwork}, HSD_{MCTQfree}) and a weighted weekly average HSD, assuming 5 workdays (HSD_{MCTQweek})⁵⁰. The resolution of the answers was 15 min. Sleep mid-points (between reported sleep onset and offset times) on work- and free days were used to calculate SJL (absolute difference between sleep mid-points on free and workdays)²⁹.

Subjective Sleep Quality was reported by participants on a 5-point Likert scale (i.e., well, rather well, neither well or badly, rather badly and badly) as in the BNSQ, in response to the question “How well have you been sleeping CURRENTLY?”. We used these categories to stratify the sample by Sleep Quality groups. Symptoms of

insomnia were assessed using the Insomnia Severity Index (ISI), a 7-item questionnaire assessing the nature, severity, and impact of insomnia during “the last month”. A 5-point Likert scale is used to rate each item (0 = no problem to 4 = very severe problem), which provided a total score ranging from 0 to 28. The total score was interpreted as follows: absence of insomnia (0–7); sub-threshold insomnia (8–14); moderate insomnia (15–21); and severe insomnia (22–28)²⁷.

Statistical analysis

Data are reported as mean \pm SD or frequency (% of group total). The agreement between the two methods for assessment of HSD (Method-Self and Method-MCTQ) was analyzed using the approach proposed by Bland and Altman⁵¹. Mean differences between the methods [$\text{HSD}_{\text{self}} - \text{HSD}_{\text{MCTQweek}}$], or [$\text{HSD}_{\text{self}} - \text{HSD}_{\text{MCTQwork}}$], or [$\text{HSD}_{\text{self}} - \text{HSD}_{\text{MCTQfree}}$] were valued as a measure of systematic bias using paired t-tests. The upper and lower limits of agreement were defined as mean difference $\pm 1.96 \times$ standard deviation with corresponding 95% confidence interval (95% CI). The difference between limits of agreement represents the range of HSD values covering the agreement between the two methods for ~95% of the individuals as a measure of precision. Sleep Quality groups were compared using Mann–Whitney or t-tests for continuous variables, according to the type and variables distribution. A simple regression model with weighted joint distribution of gender and age by country was used to estimate the contribution of these demographics to the HSD bias. Multiple regressions were run to evaluate the extent to which Sleep Quality and social time pressure (given by SJL) explained the variance in different HSDs and the HSD estimation bias itself. The main model included a set of 5 predictors: Sleep Quality, SJL, and potential demographic confounders previously linked to HSD—including age, gender, and Body Mass Index (BMI). In the validation analysis, ISI score was also used as a predictor. Collinearity tests showed no multicollinearity concerns with the predictors.

The sensitivity analyses to explore potential plausible biases were performed in a sub-group of participants with long-COVID symptoms (SI-Table S.8) and in a subgroup of older adults (>65 years old, majority after retirement, SI-Table S.9): (1) As the ICOSS-II data were collected 15–21 months after the onset of the COVID-19 pandemic, the first subgroup for sensitivity analysis included 934 (9.1% from total) individuals who met the WHO criteria for long COVID-19⁵². COVID-19 is a recent disorder that impacts sleep and may change the perception of sleep duration with the two estimates. We performed a sensitivity analysis focusing on the HSD estimation and agreement between Method-Self and Method-MCTQ to investigate potential bias in a sub-sample of participants with symptoms of long COVID. (2) Since age and retirement play a major role in sleep habits, sleep quality and social time pressure, the second subgroup for sensitivity analysis included 1187 participants (11.5% from total). The mean age of this group was 71.22 ± 3.68 years old. The data were analyzed using SPSS 29.0 (IBM Corp., Armonk, NY, USA) and R (version 4.0.5).

Data availability

We included all the data needed for the evaluation of the conclusions in the “Results” section or in the Supplementary Information file. Additional data related to this article may be requested from the authors.

Received: 19 November 2023; Accepted: 29 January 2024

Published online: 10 February 2024

References

1. Benz, F. *et al.* How many hours do you sleep? A comparison of subjective and objective sleep duration measures in a sample of insomnia patients and good sleepers. *J. Sleep Res.* **32**, e13802 (2023).
2. Chaput, J.-P. *et al.* Sleep duration and health in adults: An overview of systematic reviews. *Appl. Physiol. Nutr. Metab.* **45**, S218–S231 (2020).
3. Zhu, G. *et al.* Exploration of sleep as a specific risk factor for poor metabolic and mental health: A UK biobank study of 84,404 participants. *Nat. Sci. Sleep* **13**, 1903–1912 (2021).
4. Kósa, K., Vincze, S., Veres-Balajti, I. & Bába, É. B. The pendulum swings both ways: Evidence for U-shaped association between sleep duration and mental health outcomes. *Int. J. Environ. Res. Public Health* **20**, 5650 (2023).
5. Cappuccio, F. P., Cooper, D., D’Elia, L., Strazzullo, P. & Miller, M. A. Sleep duration predicts cardiovascular outcomes: A systematic review and meta-analysis of prospective studies. *Eur. Heart J.* <https://doi.org/10.1093/eurheartj/ehr007> (2011).
6. Cappuccio, F. P., D’Elia, L., Strazzullo, P. & Miller, M. A. Sleep duration and all-cause mortality: A systematic review and meta-analysis of prospective studies. *Sleep* **33**, 585–592 (2010).
7. Reis, C. *et al.* Sleep duration, lifestyles and chronic diseases: A cross-sectional population-based study. *Sleep Sci.* **11**, 217–230 (2018).
8. Carney, C. E. *et al.* The consensus sleep diary: Standardizing prospective sleep self-monitoring. *Sleep* **35**, 287–302 (2012).
9. Shahid, A., Wilkinson, K., Marcu, S. & Shapiro, C. M. *STOP, THAT and One Hundred Other Sleep Scales* (Springer, 2012). <https://doi.org/10.1007/978-1-4419-9893-4>.
10. Croy, I., Smith, M. G., Gidlöf-Gunnarsson, A. & Persson-Waye, K. Optimal questions for sleep in epidemiological studies: Comparisons of subjective and objective measures in laboratory and field studies. *Behav. Sleep Med.* **15**, 466–482 (2017).
11. Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193–213 (1989).
12. Dietch, J. R., Sethi, K., Slavish, D. C. & Taylor, D. J. Validity of two retrospective questionnaire versions of the Consensus Sleep Diary: The whole week and split week Self-Assessment of Sleep Surveys. *Sleep Med.* **63**, 127–136 (2019).
13. Nordin, M., Åkerstedt, T. & Nordin, S. Psychometric evaluation and normative data for the karolinska sleep questionnaire. *Sleep Biol. Rhythms* **11**, 216–226 (2013).
14. Roenneberg, T., Wirz-Justice, A. & Mrosovsky, M. Life between clocks: Daily temporal patterns of human chronotypes. *J. Biol. Rhythms* **18**, 80–90 (2003).
15. Partinen, M. & Gislason, T. Basic Nordic Sleep Questionnaire (BNSQ): A quantitated measure of subjective sleep complaints. *J. Sleep Res.* **4**, 150–155 (1995).
16. Monk, T. H. *et al.* Measuring sleep habits without using a diary: The sleep timing questionnaire. *Sleep* **26**, 208–212 (2003).

17. Huang, T. & Redline, S. Cross-sectional and prospective associations of actigraphy-assessed sleep regularity with metabolic abnormalities: The multi-ethnic study of atherosclerosis. *Diabetes Care* **42**, 1422–1429 (2019).
18. Lauderdale, D. S. Commentary on “Agreement between simple questions about sleep duration and sleep diaries in a large online survey”. *Sleep Health* **1**, 138–139 (2015).
19. Miller, C. B. *et al.* Agreement between simple questions about sleep duration and sleep diaries in a large online survey. *Sleep Health* **1**, 133–137 (2015).
20. Silva, G. E. *et al.* Relationship between reported and measured sleep times. *J. Clin. Sleep Med.* **03**, 622–630 (2007).
21. Matthews, K. A. *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* **4**, 96–103 (2018).
22. Lee, P. H. Validation of the National Health And Nutritional Survey (NHANES) single-item self-reported sleep duration against wrist-worn accelerometer. *Sleep Breath.* **26**, 2069–2075 (2022).
23. Robbins, R. *et al.* Self-reported sleep duration and timing: A methodological review of event definitions, context, and timeframe of related questions. *Sleep Epidemiol.* **1**, 100016 (2021).
24. McCarter, S. J. *et al.* Physiological markers of sleep quality: A scoping review. *Sleep Med. Rev.* **64**, 101657 (2022).
25. Santos, R. B. *et al.* Prevalence and predictors of under or overestimation sleep duration in adults: The ELSA-Brasil study. *Sleep Epidemiol.* **1**, 100013 (2021).
26. Muzni, K., Groeger, J. A., Dijk, D. & Lazar, A. S. Self-reported sleep quality is more closely associated with mental and physical health than chronotype and sleep duration in young adults: A multi-instrument analysis. *J. Sleep Res.* **30**, e13152 (2021).
27. Morin, C. M., Belleville, G., Bélanger, L. & Ivers, H. The insomnia severity index: Psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep* **34**, 601–608 (2011).
28. Korman, M. *et al.* COVID-19-mandated social restrictions unveil the impact of social time pressure on sleep and body clock. *Sci. Rep.* **10**, 22225 (2020).
29. Roenneberg, T., Pils, L. K., Zerbini, G. & Winnebeck, E. C. Chronotype and social jetlag: A (self-) critical review. *Biology (Basel)* **8**, 54 (2019).
30. Merikanto, I. *et al.* Disturbances in sleep, circadian rhythms and daytime functioning in relation to coronavirus infection and Long-COVID—A multinational ICOSS study. *J. Sleep Res.* **31**, e13542 (2022).
31. Jackson, C. L. *et al.* 0694 Concordance between self-reported and objectively-assessed sleep duration among African-American adults: Findings from the Jackson Heart Sleep Study. *Sleep* **42**, A278–A278 (2019).
32. Fernandez-Mendoza, J. *et al.* Sleep misperception and chronic insomnia in the general population: Role of objective sleep duration and psychological profiles. *Psychosom. Med.* **73**, 88–97 (2011).
33. Evangelista, E. *et al.* Characteristics associated with hypersomnia and excessive daytime sleepiness identified by extended polysomnography recording. *Sleep* **44**, zsa264 (2021).
34. Choi, S. J., Suh, S., Ong, J. & Joo, E. Y. Sleep misperception in chronic insomnia patients with obstructive sleep apnea syndrome: Implications for clinical assessment. *J. Clin. Sleep Med.* **12**, 1517–1525 (2016).
35. Pils, L., Keller, L., Lenssen, D. & Roenneberg, T. Time to rethink sleep quality: PSQI scores reflect sleep quality on workdays. *Sleep* **2**, zsy029 (2018).
36. Reis, C., Pils, L. K., Keller, L. K., Paiva, T. & Roenneberg, T. Social timing influences sleep quality in patients with sleep disorders. *Sleep Med.* **71**, 8–17 (2020).
37. Fatima, Y., Doi, S. A. R., Najman, J. M. & Al Mamun, A. Exploring gender difference in sleep quality of young adults: Findings from a large population study. *Clin. Med. Res.* **14**, 138–144 (2016).
38. Hirshkowitz, M. *et al.* National Sleep Foundation’s updated sleep duration recommendations: Final report. *Sleep Health* **1**, 233–243 (2015).
39. Leone, M. J. & Sigman, M. Effects of lockdown on human sleep and chronotype during the COVID-19 pandemic. *Curr. Biol.* **30**, R905–R931 (2020).
40. Scarpelli, S. *et al.* Subjective sleep alterations in healthy subjects worldwide during COVID-19 pandemic: A systematic review, meta-analysis and meta-regression. *Sleep Med.* **100**, 89–102 (2022).
41. Brandão, L. E. M. *et al.* Social jetlag changes during the COVID-19 pandemic as a predictor of insomnia—A multi-national survey study. *Nat. Sci. Sleep* **13**, 1711–1722 (2021).
42. Bliwise, D. L. & Young, T. B. The parable of parabola: What the U-shaped curve can and cannot tell us about sleep. *Sleep* **30**, 1614–1615 (2007).
43. Lauderdale, D. S., Knutson, K. L., Yan, L. L., Liu, K. & Rathouz, P. J. Self-reported and measured sleep duration. *Epidemiology* **19**, 838–845 (2008).
44. Bianchi, M. T., Thomas, R. J. & Westover, M. B. An open request to epidemiologists: Please stop querying self-reported sleep duration. *Sleep Med.* **35**, 92–93 (2017).
45. Schurhoff, N. & Toborek, M. Circadian rhythms in the blood–brain barrier: Impact on neurological disorders and stress responses. *Mol. Brain* **16**, 5 (2023).
46. Zavada, A., Gordijn, M. C. M., Beersma, D. G. M., Daan, S. & Roenneberg, T. Comparison of the Munich Chronotype Questionnaire with the Horne–Östberg’s morningness–eveningness score. *Chronobiol. Int.* **22**, 267–278 (2005).
47. Roenneberg, T., Daan, S. & Mrosovsky, M. The art of entrainment. *J. Biol. Rhythms* **18**, 183–194 (2003).
48. Evanger, L. N. *et al.* Later school start time is associated with longer school day sleep duration and less social jetlag among Norwegian high school students: Results from a large-scale, cross-sectional study. *J. Sleep Res.* **32**, e13840 (2023).
49. St-Onge, M.-P. *et al.* Information on bedtimes and wake times improves the relation between self-reported and objective assessments of sleep in adults. *J. Clin. Sleep Med.* **15**, 1031–1036 (2019).
50. Ghotbi, N. *et al.* The μ MCTQ: An ultra-short version of the Munich ChronoType Questionnaire. *J. Biol. Rhythms* <https://doi.org/10.1177/0748730419886986> (2019).
51. Bland, J. M. & Altman, D. G. Comparing methods of measurement: Why plotting difference against standard method is misleading. *Lancet* **346**, 1085–1087 (1995).
52. Soriano, J. B., Murthy, S., Marshall, J. C., Relan, P. & Diaz, J. V. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect. Dis.* **22**, e102–e107 (2022).

Acknowledgements

We acknowledge Ying Huang (Germany), Harald Hrubos-Strøm (Norway), Colin A. Espie (Great Britain) and Yuichi Inoue (Japan) for being instrumental in giving inputs for study design or providing data to this study. This material is partially the result of work supported with resources at the South Texas Veterans Health Care System in San Antonio, TX, USA. The contents of this publication do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

Author contributions

Conceptualization: M.K., C.R.; Data curation: M.K. (Israel), I.M., M.P. (Finland), B.B. (Norway), A.K.B. (Croatia), T.P. (Germany), A.L., C.B. (Sweden), N.Y.C., Y.K.W. (Hong Kong), Y.D. (France), C.M.M. (Canada), K.M. (Japan), M.N., F.C. (US), S.M.R. (Brazil), L.G., G.P. (Italy), J.Y. (Bulgaria), B.H. (Austria), C.R. (Portugal); Formal analysis: M.K., D.Z., V.T.; Analysis discussion: M.K., C.R.; Methodology: M.K., C.R.; Project administration: M.P., I.M.; Writing—original draft: M.K., D.Z., V.T., C.R.; and all authors review, edit and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-53174-1>.

Correspondence and requests for materials should be addressed to M.K. or C.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024