



ORIGINAL ARTICLE

Measuring sleep regularity: theoretical properties and practical usage of existing metrics

Dorothee Fischer^{1,2,3,*}, Elizabeth B Klerman^{2,3,4}, Andrew JK Phillips⁵

¹Department of Sleep and Human Factors Research, Institute for Aerospace Medicine, German Aerospace Center, Cologne, Germany ²Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, USA ³Division of Sleep Medicine, Harvard Medical School, Boston, MA, USA ⁴Department of Neurology, Massachusetts General Hospital, Boston, MA, USA ⁵Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Melbourne, VIC, Australia

*Corresponding author. Dorothee Fischer, Sleep and Human Factors Research, Institute of Aerospace Medicine, German Aerospace Center, Linder Hoehe 15, 51147 Cologne, Germany. Email: Dorothee.Fischer@dlr.de.

Abstract

Study Objectives: Sleep regularity predicts many health-related outcomes. Currently, however, there is no systematic approach to measuring sleep regularity. Traditionally, metrics have assessed deviations in sleep patterns from an individual's average; these traditional metrics include intra-individual standard deviation (StDev), interdaily stability (IS), and social jet lag (SJL). Two metrics were recently proposed that instead measure variability between consecutive days: composite phase deviation (CPD) and sleep regularity index (SRI). Using large-scale simulations, we investigated the theoretical properties of these five metrics.

Methods: Multiple sleep-wake patterns were systematically simulated, including variability in daily sleep timing and/or duration. Average estimates and 95% confidence intervals were calculated for six scenarios that affect the measurement of sleep regularity: “scrambling” the order of days; daily vs. weekly variation; naps; awakenings; “all-nighters”; and length of study.

Results: SJL measured weekly but not daily changes. Scrambling did not affect StDev or IS, but did affect CPD and SRI; these metrics, therefore, measure sleep regularity on multi-day and day-to-day timescales, respectively. StDev and CPD did not capture sleep fragmentation. IS and SRI behaved similarly in response to naps and awakenings but differed markedly for all-nighters. StDev and IS required over a week of sleep-wake data for unbiased estimates, whereas CPD and SRI required larger sample sizes to detect group differences.

Conclusions: Deciding which sleep regularity metric is most appropriate for a given study depends on a combination of the type of data gathered, the study length and sample size, and which aspects of sleep regularity are most pertinent to the research question.

Statement of Significance

Sleep regularity may be as important for health as sleep duration and timing. Traditional metrics for quantifying sleep regularity compare each day's sleep-wake pattern to the average pattern, whereas other, recently-developed metrics compare sleep between consecutive days. We examined five sleep regularity metrics under different simulated scenarios and study lengths, including common sources of day-to-day variability. We found that there are differences between metrics in their sensitivity to specific sources of variability. We also found that there are differences between metrics in the requirements for study length and sample size. These findings provide a set of guidelines for the selection and use of sleep regularity metrics, for a time in which sleep regularity is becoming a major focus for the field.

Key words: intra-individual variability; inter-individual variability; sleep variability; sleep stability; circadian misalignment; circadian disruption

Submitted: 24 November, 2020; Revised: 18 February, 2021

© Sleep Research Society 2021. Published by Oxford University Press on behalf of the Sleep Research Society.

All rights reserved. For permissions, please email: journals.permissions@oup.com

Introduction

Day-to-day variability in sleep-wake patterns has emerged in recent years as an important factor for health and safety [1, 2]. Studies have found associations of irregular sleep with multiple adverse outcomes, including cardiac autonomic modulation [3], inflammation [4], metabolism [5–7], mental health [8, 9], and performance and cognitive function [10–12]. These studies, however, have used a variety of metrics to quantify sleep regularity, and the vast majority used metrics that quantify sleep regularity by comparing each day's sleep-wake pattern to the individual's average pattern. These *overall metrics* include intra-individual standard deviation (StDev) and interdaily stability (IS [13]). The daily variation around an individual's mean, as quantified by these metrics, has also been referred to as intra-individual variability (IIV) [1]. A related metric is social jet lag (SJL [14]), which quantifies the difference in average sleep timing between workdays and free days.

Recently, two novel metrics have been developed to instead capture variability in sleep-wake patterns between consecutive days, based on the hypothesis that day-to-day changes may cause circadian disruption or misalignment. These *consecutive metrics* include the composite phase deviation (CPD [15]) and the sleep regularity index (SRI [16]). Higher CPD values, indicating lower regularity, have been associated with later chronotype [15], poorer mood/well-being and more irregular class/event schedules in college students [17], lower stability of meal timing [18], and more variable caloric intake in Type 2 diabetes patients with melatonin MT_2 receptor mutations [19]. Lower SRI values, indicating lower regularity, have been associated with delayed circadian phase and lower academic performance [16, 20], impaired daytime function [21, 22], poorer mood/well-being [23, 24], depression [25], insomnia and PTSD [26], and impaired cardiometabolic outcomes [27].

This study was designed to theoretically compare these five sleep regularity metrics, and to help researchers make informed decisions about which sleep regularity metric(s) may be appropriate for the specific features of their study, including study length and sample size. While overall metrics other than StDev and IS have been used to quantify variability in sleep-wake patterns, most notably IS's counterpart intra-daily variability (IV), we chose StDev and IS because (1) they specifically quantify variability between days, not within days (which is the purpose of the IV metric), and (2) they are the most commonly used measures in sleep and circadian rhythm research [1]. To allow the theoretical properties of each metric to be established, we simulated sleep-wake patterns using scenarios that included different sources of variability: "scrambling" the order of days; daily vs. weekly variation in sleep timing and duration; daytime naps; nocturnal awakenings; and "all-nighters" (nights with no sleep). We also compared the performance of the five metrics in response to sleep-wake patterns of different lengths (2–28 days) and calculated the sample size required to detect a significant difference between two groups by each metric. These findings can guide the planning of future studies and provide context for interpreting existing research.

Methods

Metrics for quantifying sleep regularity

The metrics can be calculated from any multi-day recording of sleep (e.g. actigraphy, sleep diaries, PSG).

Overall metrics

Overall metrics assess sleep regularity by comparing features of each day's sleep pattern to the individual's average sleep pattern (StDev, IS), or by comparing an individual's average timing of sleep on workdays with their average timing of sleep on free days (SJL). These metrics each quantify variability relative to a mean, rather than comparing sleep patterns between consecutive days. The three selected overall metrics are the most commonly used to quantify sleep regularity but are not exhaustive. Other measures exist, for example, the coefficient of variation (CoV), which is closely related to standard deviation.

Standard deviation (StDev). StDev is a statistical measure of the amount of variation or dispersion. A lower number indicates a more regular pattern. It can be applied to features of daily sleep (e.g. sleep onset, sleep offset, midsleep, sleep duration) (Figure 1, C):

$$\text{StDev} = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}} \quad (1)$$

where N is the number of days, X_i is the value of the sleep variable on day i , and \bar{X} is the mean.

Interdaily stability (IS). IS was developed to measure the stability of rest-activity rhythms over multiple days, comparing the pattern of activity each day to the average pattern across days. Higher IS values reflect more regular patterns of activity. IS originally used activity count data but can also be defined using binary sleep-wake data (i.e. sleep (=0) or wake (=1) state assigned to each epoch). It is calculated as the ratio of the variance within the same time interval each day and the overall variance (Figure 1, D):

$$\text{IS} = \frac{N \sum_{h=1}^p (\bar{X}_h - \bar{X})^2}{p \sum_{i=1}^N (X_i - \bar{X})^2} \quad (2)$$

where N is the total number of data points; p is the number of data points per day (e.g. 1-h epochs = 24 data points per day); \bar{X} is the mean of all data; \bar{X}_h are the means of every epoch across all days; and X_i represent the individual data points. A higher number indicates a more regular pattern.

Social jet lag (SJL). SJL is the mismatch in average midsleep timing between workdays and free days (Figure 1, B). Midsleep on weekends can be either earlier (negative SJL values) or later (positive SJL values) than midsleep on weekdays:

$$\text{SJL} = \text{MSF} - \text{MSW}$$

where MSF is the average midsleep on work-free days and MSW is the average midsleep timing on workdays. A lower number indicates a more regular pattern.

Consecutive metrics

Consecutive metrics have been recently proposed to specifically measure variability in sleep-wake patterns between consecutive days. The rationale for these metrics is that changes from one day to the next (on a circadian timescale) are challenging for the circadian system to accommodate, due to limits on the

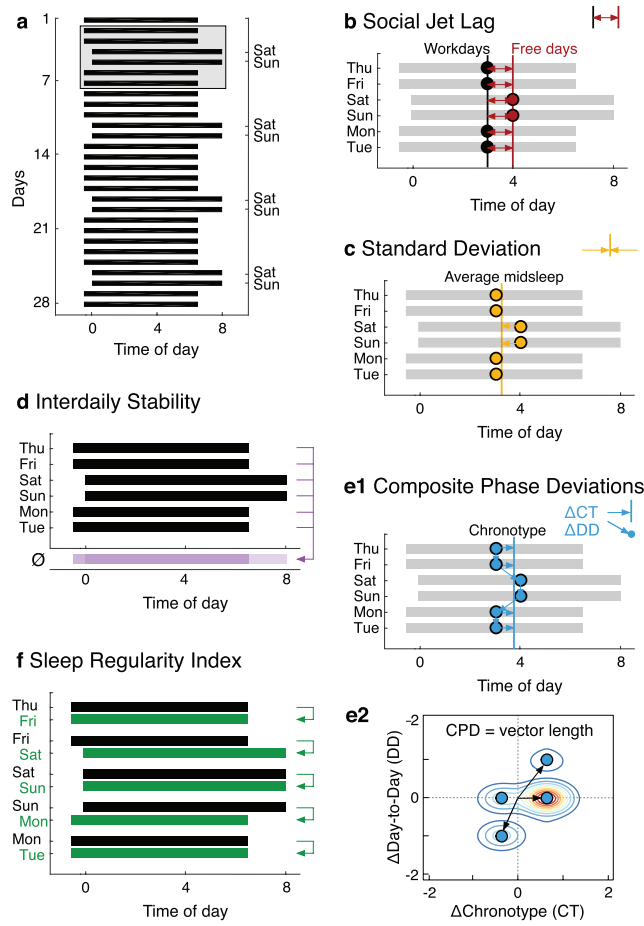


Figure 1. Sleep regularity metrics. (A) Raster plot of a 28-day sleep-wake pattern with a weekly variation of 1 h later and 1 h longer sleep on weekends than weekdays. The gray box marks six days of the pattern used to illustrate the five sleep regularity metrics in panels B-F. (B) Social jet lag (SJL). (C) Standard deviation (StDev). (D) Interdaily stability (IS). (E1 and E2) Composite phase deviation (CPD). (F) Sleep regularity index (SRI). Note: Only clock times 0-8 are shown for space reasons.

rate of phase resetting. Measuring this type of variability may therefore be a useful proxy for circadian disruption associated with irregular sleep patterns.

Composite phase deviation (CPD). The CPD was developed to quantify circadian disruption in the context of shift work, where sleep-wake behavior is often both mistimed (e.g. sleep during the daytime) and irregular (e.g. rotating schedules where working times alternate). CPD uses an individual's chronotype to determine the optimal timing of sleep; chronotype is assessed using midsleep on free days, corrected for potential sleep loss on workdays (MSF_{sc}) [28]. The “mistiming” component is quantified by calculating the difference in hours between midsleep timing on one day and the individual's chronotype (Δ Chronotype); the “irregularity” component is computed as the difference between midsleep timing on one day from that on the previous day (Δ Day-to-Day) (Figure 1, E1). The two components are plotted against each other and the vector length from data point to the origin (i.e. perfectly aligned and regular sleep) is calculated for N sleep episodes and averaged (Figure 1, E2):

$$\Delta\text{Chronotype}_i (\Delta\text{CT}_i) = \text{MSF}_{\text{sc}} - \text{MS}_i \quad (4a)$$

$$\Delta\text{Day-to-Day}_i (\Delta\text{DD}_i) = \text{MS}_{i-1} - \text{MS}_i \quad (4b)$$

$$\text{CPD}_i = \sqrt{\Delta\text{CT}_i^2 + \Delta\text{DD}_i^2} \quad (4c)$$

$$\text{CPD} = \frac{1}{N} \sum_{i=1}^N \text{CPD}_i \quad (4d)$$

Calculating CPD requires as input the designation of one main sleep episode per day. CPD can also be derived for other sleep variables, such as daily sleep durations. A lower number indicates a more regular pattern.

Sleep regularity index (SRI). The SRI measures the similarity of an individual's sleep-wake patterns from one day to the next, based on binary sleep-wake state classifications. It calculates the percentage probability of an individual being in the same state (sleep vs. wake) at any two time points (e.g. 30-s epochs) 24-h apart, averaged across the study. The SRI is scaled to range from 0 (random) to 100 (perfectly regular):

$$\text{SRI} = -100 + \frac{200}{M(N-1)} \sum_{j=1}^M \sum_{i=1}^{N-1} \delta(s_{i,j}, s_{i+1,j}) \quad (5)$$

where M is the number of daily epochs, N is the number of days, $s_{i,j} = 0$ for sleep and $s_{i,j} = 1$ for wake, and $\delta(s_{i,j}, s_{i+1,j}) = 1$ if $s_{i,j} = s_{i+1,j}$ and 0 otherwise. The metric is defined so that two

individuals with the same standard deviation in midsleep time but different sleep durations (e.g. 4 h vs. 8 h) have approximately the same SRI scores. As a result, the SRI tends to be uncorrelated with total sleep time [16, 21, 27], despite the fact that the SRI would tend to a value of 100 if an individual hypothetically slept 0% or 100% of the time. We note that the SRI is sometimes stated to have a range of 0–100, as this corresponds to its practical range. In fact, the full theoretical range is -100 – 100 , but virtually all real-world sleep/wake patterns fall in the range 0–100. SRI values less than 0 are therefore theoretically possible (e.g. sleep for 24 h, wake for 24 h, etc.), though uncommon in practice [16]. The SRI makes no assumptions about the structure of sleep (e.g. whether there is one main sleep episode or naps). A higher number indicates a more regular pattern.

Pairing of metrics

Among the overall metrics, SJL conceptually differs from StDev and IS. SJL does not compare sleep on each day to the average sleep–wake pattern, but instead compares two averages (workdays and free days). The remaining four metrics can be paired according to how they integrate the sleep–wake data (= signal) into a sleep regularity score (Table 1). In this respect, StDev and CPD both use extracted daily features (e.g. midsleep time) that are derived from the recorded sleep–wake episode data, with a single value per day. We label these *daily-value* metrics. Conversely, IS and SRI both use the information in all recorded epochs (i.e. multiple values per day). We label these *whole-signal* metrics.

Scenarios

Our principal aim was to simulate sleep–wake patterns that closely resemble real-life patterns. To this end, the basic aspects of the simulated sleep–wake patterns were based on values reported in the literature; namely, longer and later sleep on weekends [29], the amount of social jetlag (i.e. ≥ 1 h in 80% of the population) [30], and the variation in sleep duration observed for the general population (i.e. 60 min on average) [31, 32]. These assumptions were the same across scenarios where applicable. Within each scenario, we varied specific parameters to examine their impact on sleep regularity metrics (e.g. number of all-nighters). In certain cases, where we discovered a non-monotonic relationship for at least one of the sleep regularity metrics, we explored the full theoretical range, in some cases going beyond what may be seen in the real world (e.g. 28 days of no sleep). These cases are included because, though they extend beyond observable behaviors, they provide insights into the theoretical underpinnings of each sleep regularity metric.

Sleep–wake patterns were generated for 28 days in MATLAB (MathWorks, Natick MA, USA), consisting of binary time series

of sleep (=0) or wake (=1), starting randomly on any day of the week. Day-to-day sleep variability was introduced by sampling daily midsleep times and/or sleep durations from random Normal distributions. The random Normal distributions were limited to $\pm 3\sigma$ (i.e. 99.7% of the distribution) to avoid cases of potentially overlapping or negative duration sleep episodes. We simulated six scenarios, illustrated in Figure 2, to compare how the sleep regularity metrics performed.

Scenario 1: daily variation

Average sleep timing was set to start and end at 0:00 and 8:00, respectively (average midsleep at 4:00, 8 h sleep duration). We note that the choice of sleep timing is arbitrary, as the same simulated schedules could be shifted earlier or later with no effect on sleep regularity metrics. Daily sleep variability was added by randomly varying (1) midsleep, (2) sleep duration, or (3) both. Standard deviation (σ) of the random Normal distributions ranged in 15-min steps from 0 to 120 min for (1) and (2), and from 0 to 90 min for (3). These ranges were selected to encompass a range of regular to irregular sleep behaviors.

Scenario 1A: “scrambling”

To examine how sensitive the metrics were to changes in the day-to-day sleep sequence, one sleep–wake pattern from Scenario 1 was randomly generated for each level of variation in midsleep (0–120 min). Days were then randomly re-ordered (scrambled) 1,000 times. This scenario serves to demonstrate whether each metric is sensitive to changes in sleep from one day to the next; if so, scrambling the order of sleep episodes would affect the metrics’ values. Scrambling the order of the data, therefore, allows us to appreciate the timescale on which each metric measures variability. This has important practical implications, which are covered in the Discussion.

Scenario 2: weekly variation

Real-life sleep patterns commonly show later and longer sleep on weekends (usually a free day) than on weekdays (usually a workday) [29]. To investigate this type of variation, we added a weekday-weekend difference in sleep timing and/or duration to the daily variation from Scenario 1. Average sleep timing was set to start and end at 23:30 and 6:30, respectively, on weekdays (midsleep at 3:00, 7 h sleep duration) and at 0:00 and 8:00, respectively, on weekends (midsleep at 4:00, 8 h sleep duration). The difference between weekdays and weekends was thus 1h in both midsleep and sleep duration, which is the minimum level of weekday-weekend difference experienced by 80% of the general population [30]. Examples of sleep–wake patterns in Scenario 2 with different variabilities are shown in Supplement A, Figure S1.

Scenario 3: naps

To assess how multiple blocks of sleep per day affect the metrics, naps were added to sleep–wake patterns of Scenario 2, with daily variation of the main sleep episode set to $\sigma = 60$ min for both midsleep and sleep duration. Day-to-day variability in naps was generated by randomly varying both the midpoint of the nap (“midnap”) and nap duration. Daily naps were drawn

Table 1. Classification of metrics

Metrics	Calculation	
	Daily-value (one value per day)	Whole-signal (multiple values per day)
Overall	StDev	IS
Consecutive	CPD	SRI

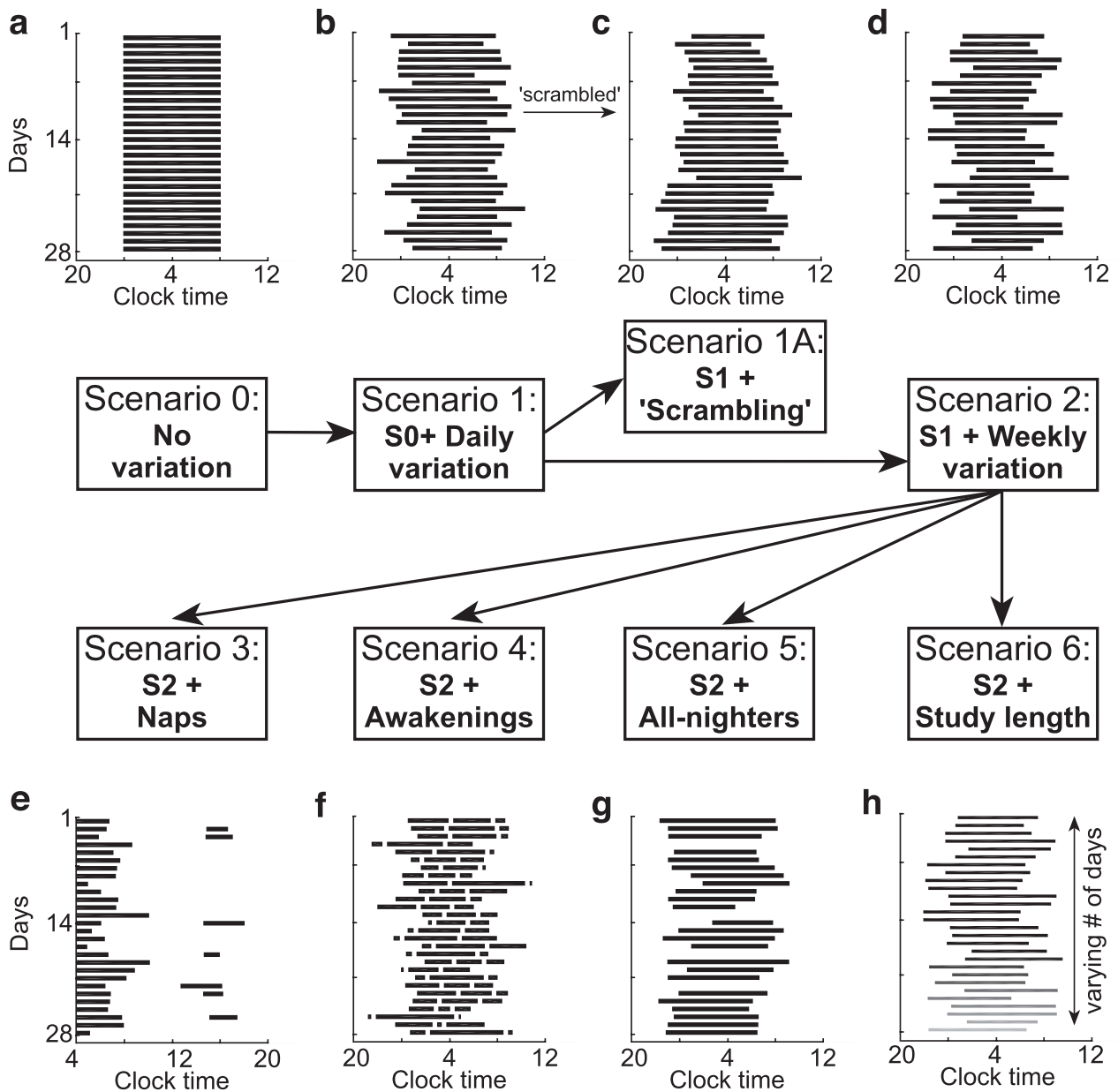


Figure 2. Scenarios of sleep regularity. (A) Scenario 0 illustrates a perfectly regular sleep–wake pattern over 28 days (zero variation). (B) Scenario 1 adds daily variation in midsleep, sleep duration, or both. (C) Scenario 1A uses sleep–wake patterns from Scenario 1, but with re-arranged (“scrambled”) order of days. (D) Scenario 2 adds weekly variation to sleep–wake patterns from Scenario 1. Scenarios 3 to 5 further add (E) daytime naps, (F) nocturnal awakenings, or (G) “all-nighters” (nights with no sleep) to Scenario 2. (H) Scenario 6 uses sleep–wake patterns of Scenario 2, varying the number of days (2–28) to simulate varying study lengths, illustrated by the gradual shading.

from a random Normal distribution with average midnap timing and duration of 15:00 and 2 h, respectively (average nap onset and offset: 14:00–16:00) and standard deviation σ that increased from 0 to 30 min in 10-min steps. The distribution width was again limited to $\pm 3\sigma$. The total number of days with naps was systematically varied from 0–28 in the 28-day simulation.

Scenario 4: Awakenings

To assess how fragmentation of the main sleep episode affects the metrics, nocturnal awakenings (wake after sleep onset, WASO) were added to sleep–wake patterns of Scenario 2, with daily variation of the main sleep episode set to $\sigma = 60$ min in

both midsleep and sleep duration. The number of Wake bouts per sleep episode varied from 1 to 60 (1–10 bouts and then in 10-bout steps) and the total duration of WASO from 10 to 240 min in 10-min steps. Accordingly, the length of a single Wake bout varied between 10 s and 4 h, with fixed length across all sleep episodes within a simulation. The wide range of number and total duration of Wake bouts was chosen to compare insomnia-like sleep–wake patterns (long but few bouts) with highly fragmented (short but many bouts) sleep–wake patterns. The position of Wake bouts within a sleep episode was drawn from a random uniform distribution; that is, Wake bouts were equally likely to occur throughout the sleep episode but were not allowed to overlap with one another.

Scenario 5: “all-nighters”

To assess how total sleep deprivation affects the metrics, nights with no sleep were generated by replacing Sleep with Wake in sleep-wake patterns of Scenario 2, with daily variation of the main sleep episode set to $\sigma = 0$ –120 min in sleep duration and $\sigma = 60$ min in midsleep. For nights with no sleep, midsleep values were treated as missing and sleep durations were set to zero. The number of all-nighters was systematically varied from 0 to 28 in the 28-day simulation.

Scenario 6: study length

Sleep-wake patterns were generated for study lengths from 2 to 28 days to determine the statistical dependence of each metric on study length. Sleep-wake patterns were simulated in the same way as for Scenario 2, with two differences: (1) to increase comparability with real data, we introduced missing data, with a one in seven chance of each day being missing, similar to real levels of missingness [1]; and (2) to simplify interpretability, daily variation increased from 15 to 120 min in 15-min steps for midsleep timing, but was set to 60 min for sleep duration, representing an average level of variability in healthy adult populations [31, 32].

Data processing

For each simulation within each scenario, we generated 1,000 28-day sleep-wake patterns with random daily variation. For example, in Scenario 3 where naps were added to the main sleep-wake pattern, a total of $29 \times 4 = 116$ combinations were simulated (i.e. 0–28 naps and 0, 10, 20, 30 min variation in midnap timing) with 1,000 iterations for each. Average estimates and 95% confidence intervals (CIs) were calculated for each metric across each batch of 1,000 iterations.

Sleep regularity metrics

The metrics SRI and IS were computed identically in all scenarios. Since the other three metrics (SJL, StDev, and CPD) are reliant on a specific feature (e.g. midsleep) and insensitive to all other features, their calculation was performed in multiple ways. For simulations that varied midsleep, the three metrics were calculated using daily midsleep values (CPD calculating deviations from chronotype and from previous midsleep). For simulations that varied sleep duration, the three metrics were calculated using daily sleep durations (CPD calculating deviations from average sleep duration and from previous sleep duration).

For sleep-wake patterns with naps (Scenario 3), variants of StDev and CPD were calculated using both midpoints of main sleep episodes (midsleeps) and midpoints of naps (midnaps): (1) $\text{StDev}^{\text{Nap}}$ was calculated as the standard deviation of all midsleeps and midnaps, and (2) CPD^{Nap} was calculated as the average of CPD using midsleeps and CPD using midnaps, weighted by the number of naps (e.g. if one nap occurred, the contribution of CPD using midnaps to CPD^{Nap} would be weighed by $1/28$).

Calculation of sample sizes. In Scenario 6, the average estimates and standard deviations calculated for each metric were used to determine sample sizes needed to detect a statistically significant difference between two groups of different sleep regularity at a given power of 0.8. Sample sizes were calculated using Welch’s t-test to account for unequal variances between

groups. Sample sizes were calculated in R using the package “pwrAB” [33].

Results**SJL is a measure of weekly but not daily sleep regularity****Scenario 1: daily variation**

SJL was the only metric whose average values did not change with increasing daily variability (Figure 3, A), highlighting the fact that SJL operates on a weekly but not a daily timescale. With increasing daily variation, StDev and CPD increased, while IS and SRI decreased (Figure 3, B–E).

Overall vs. consecutive metrics measure sleep regularity on multi-day vs. day-to-day timescales**Scenario 1A: “scrambling”**

Scrambling did not have any effect on values of the three overall metrics: SJL, StDev, and IS maintained identical average estimates and zero-width 95% CIs, despite changes in the day-to-day sequence of sleep episodes (Figure 3, A, B, D). In contrast, scrambling did affect values of the consecutive metrics CPD and SRI, slightly affecting average estimates and widening 95% CIs with increasing daily variation (Figure 3, C, E). Confidence intervals were wider for SRI than CPD, indicating a higher sensitivity of SRI to re-ordering the daily sleep sequence. This is due to CPD combining a mistiming and an irregularity component, only the latter of which is sensitive to re-ordering of days.

Daily-value vs. whole-signal metrics integrate single vs. multiple sources of variability**Scenario 2: daily + weekly variation**

When weekly variation was added to the daily variation, all metrics’ values changed to a more irregular value (Figure 3, F–J). SJL equaled 1 h, the exact amount of introduced weekly variation (Figure 3, F). Though average SJL remained the same across different levels of daily variation, 95% CIs increased, showing that individual sleep-wake patterns with the same underlying weekday-weekend difference can have different SJL values. Because of its limitations in reflecting daily changes in sleep, SJL was not included in further analyses.

StDev and CPD could capture only variation in either midsleep timing or sleep duration, depending upon how they were defined (Figure 3, G, H, K, L). In contrast, IS and SRI values were sensitive to both forms of variation, with greater sensitivity to variation in midsleep timing than sleep duration (Figure 3, I, J).

When midsleep and sleep duration both contributed to daily variability, StDev and CPD values increased in one direction only: when based on midsleep, StDev and CPD varied only by midsleep once a sleep duration variation was chosen, and vice versa; there was no interaction observed (Figure 3, K, L). IS and SRI values showed an interaction, with the lowest values when variation was greatest in both sources (Figure 3, M, N). Hence, the major difference between daily-value vs. whole-signal metrics is the number of sources of variability they reflect in their sleep regularity scores. StDev and CPD quantify variability from a single source (i.e. midsleep or sleep duration), whereas IS and

SRI reflect variability from several sources (i.e. midsleep and sleep duration).

Key points for Scenarios 1, 1A, and 2

- SJL measures weekly but not daily changes in sleep.
- Only consecutive metrics (CPD, SRI) are sensitive to re-ordering of a sleep recording.
- Daily-value metrics (StDev, CPD) quantify variability from a single, specified source (i.e. either sleep timing or duration).
- Whole-signal metrics (IS, SRI) combine variability from multiple sources (i.e. both sleep timing and duration).

IS and SRI (but not StDev and CPD) are sensitive to napping and awakenings

Scenario 3: naps

As the number of days with naps increased from 0% to 50%, IS and SRI decreased, reflecting lower regularity. As the number of

days with naps increased from 50% to 100%, IS and SRI increased, reflecting higher regularity (Figure 4, D, E). If there was no variation in the timing or duration of naps, IS and SRI eventually returned to their original values; that is, values of sleep regularity were the same for 0% and 100% naps. The extent to which naps decreased IS and SRI values compared to sleep-wake patterns without naps was up to ~10% of the scale range. StDev and CPD in their original calculations do not include naps, only main sleep. When variants of these metrics were computed that included naps, $StDev^{Nap}$ increased with number of naps but did not differ between variations in midnap/nap duration (Figure 4, B). CPD^{Nap} values showed slight changes with increasing number of naps and variations in midnap/nap duration, yet the differences were overall very small (Figure 4, C).

Scenario 4: awakenings

StDev and CPD are not sensitive to fragmented sleep (i.e. WASO), and so did not vary under this Scenario. The number of Wake bouts had virtually no impact on the IS and SRI values (Figure

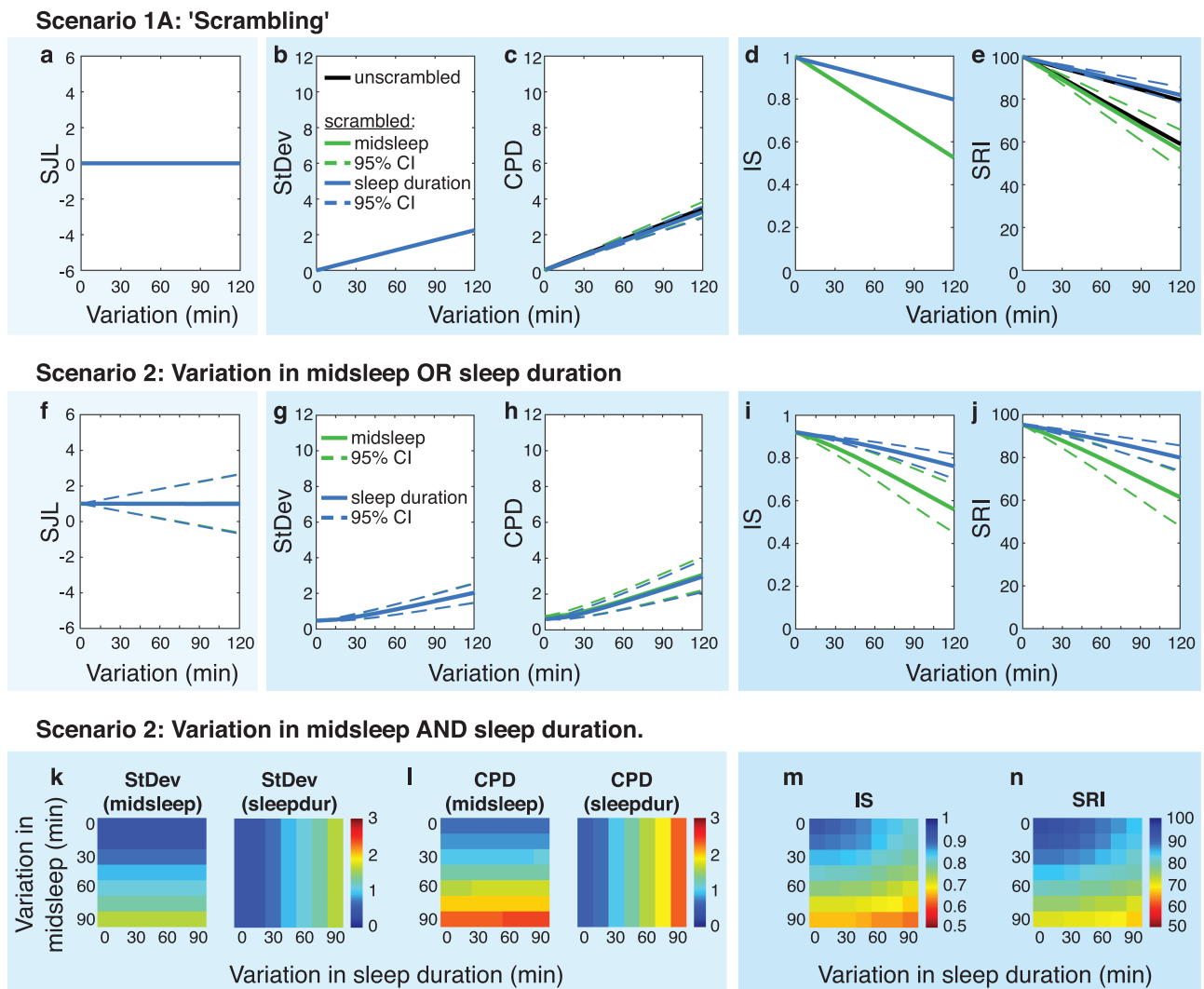


Figure 3. Sources of variability: “scrambling,” daily variation, and weekly variation (Scenarios 1A + 2). Panels A–E show the five metrics for randomly re-ordered (“scrambled”) sleep-wake patterns in Scenario 1A. Note that the metrics SJL, StDev, and IS are identical for unscrambled and scrambled sleep patterns (i.e. zero-width 95% CIs, colored lines overlay). Panels F–J show the five metrics calculated for sleep-wake patterns of Scenario 2, with daily variation in midsleep OR sleep duration. Panels K–N show the metrics for the same patterns but with daily variations in midsleep AND sleep duration. Warmer colors in heat maps indicate “more irregular” values in all metrics. Note that SJL is not shown in the bottom row due to its non-response to daily variation (see panel A).

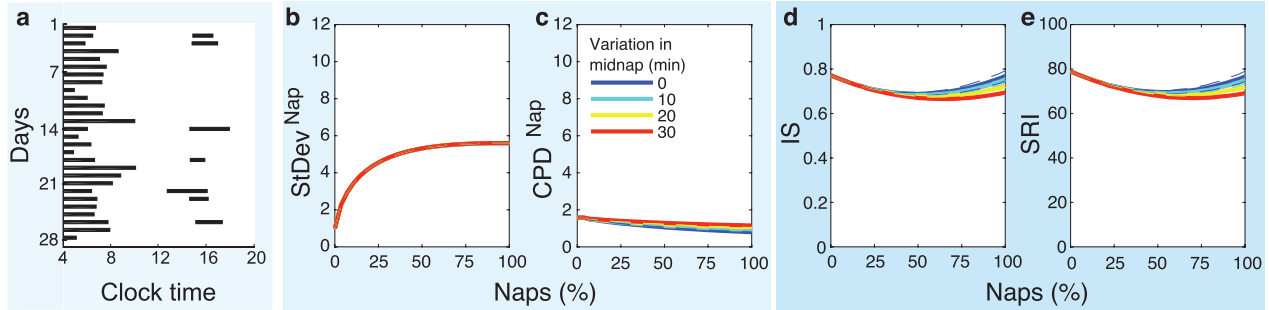
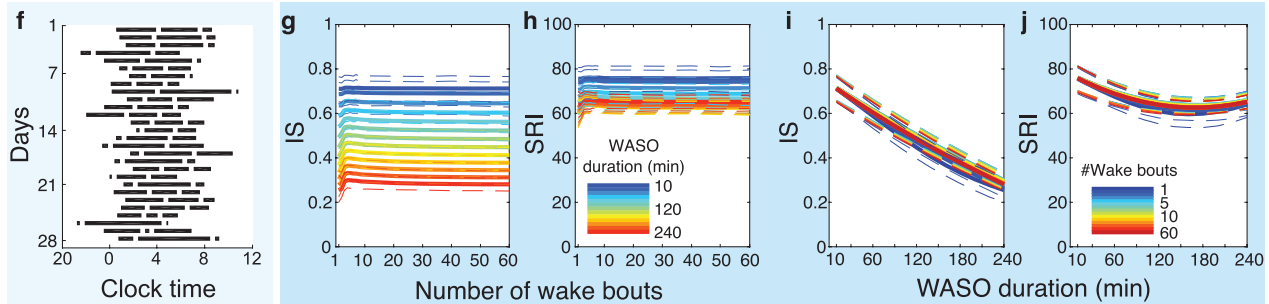
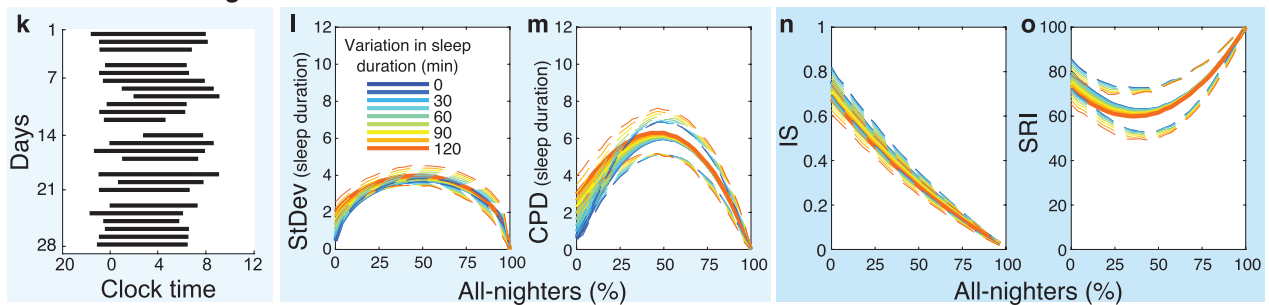
Scenario 3: Naps**Scenario 4: Awakenings (IS and SRI only)****Scenario 5: 'All-nighters'**

Figure 4. Sources of variability: naps, awakenings, and “all-nighters” (Scenarios 3, 4, and 5). The far-left panel of each row shows example raster plots of 28-day sleep-wake patterns with (A) daytime naps, (F) nocturnal awakenings, and (K) nights with no sleep (“all-nighters”). Panels (B–E) StDev, CPD, IS, and SRI for sleep-wake patterns fragmented by daytime naps. For clearer illustration, only variations in midnap timing are shown, with metrics’ values averaged across variations in nap duration. Panels (G–J) IS and SRI for sleep-wake patterns fragmented by nocturnal awakenings (WASO, wake after sleep onset). Panels (L–O) StDev, CPD, IS, and SRI for sleep-wake patterns with all-nighters.

4, G, H). With increasing total WASO duration, both IS and SRI decreased at first, reflecting less regular sleep (Figure 4, I, J); yet, for total WASO durations of >160 min, SRI increased again, reflecting “more regular” values, due to increasing night-to-night overlap of awakenings, whereas IS yielded monotonically lower values. The impact of WASO was large on IS and moderate on SRI. Compared with a consolidated sleep-wake pattern (no awakenings), WASO reduced IS and SRI by as much as 47% and 16% of the scale range, respectively. Increasing wakefulness seemed to affect IS and SRI differently, which we explored more deeply in Scenario 5.

IS requires alternating sleep and wake states for a perfectly regular score

Scenario 5: “all-nighters”

As the number of all-nighters (i.e. nights with no sleep) increased from 0% to 50%, the StDev and CPD (based on daily sleep durations) increased while SRI decreased, reflecting “less

regular” scores for each metric (Figure 4, L, M, O). As the number of all-nighters increased from 50% to 100%, these three metrics returned to more regular scores, reaching a perfectly regular score for constant wake (100% all-nighters), due to each day having an identical pattern (no sleep at all). The behavior of the IS metric was notably different, as it decreased monotonically with increasing number of all-nighters from 0% to 100% (Figure 4, N). Although every day was identical with 100% all-nighters, the lack of daily rhythmicity (i.e. low or zero amplitude) was reflected by a low IS score. A combination of all-nighters and naps was simulated to examine the impact of (compensatory) naps occurring during the daytime following an all-nighter. Naps rescued rhythmicity as measured by IS, with values returning to more regular scores for >50% naps (Supplement B, Figure S2).

Key points for Scenarios 3, 4, and 5

- StDev and CPD are not designed to assess regularity of sleep-wake patterns that include naps or awakenings.

- IS and SRI take into account both naps and awakenings, assigning higher scores when the sleep–wake pattern is more consistent between days.
- IS is the only metric that quantifies rhythmicity/amplitude in addition to regularity of sleep–wake patterns, as indicated by its monotonic reduction with increasing awakenings or all-nighters.

Overall metrics overestimate sleep regularity when based on ≤ 7 days of data

Scenario 6: study length

As simulated study length was increased from 2 to 28 days, StDev and IS values underwent systematic changes up to ~ 7 days of data (Figure 5, B, D), indicating a bias in overall metrics toward overestimating sleep regularity (therefore, underestimating sleep irregularity) for short study lengths, compared with what would be calculated using the same metric in the same individual simply from a longer recording. We note that this bias was not driven by day of the week, since the starting day of the week for simulations was randomized. This difference was particularly pronounced for the comparison IS vs. SRI, and less so for StDev vs. CPD (see Supplement B, Figure S2, C, F). The change in IS values over 2–7 days was larger for higher daily variations, suggesting that IS particularly overestimated highly irregular sleep–wake patterns (Figure 5, D, Supplement B, Figure S2, D). SRI displayed the least dependence on study length of all metrics (Figure 5, E).

The 95% CIs widened with increasing levels of daily variability and narrowed with an increasing number of days for all

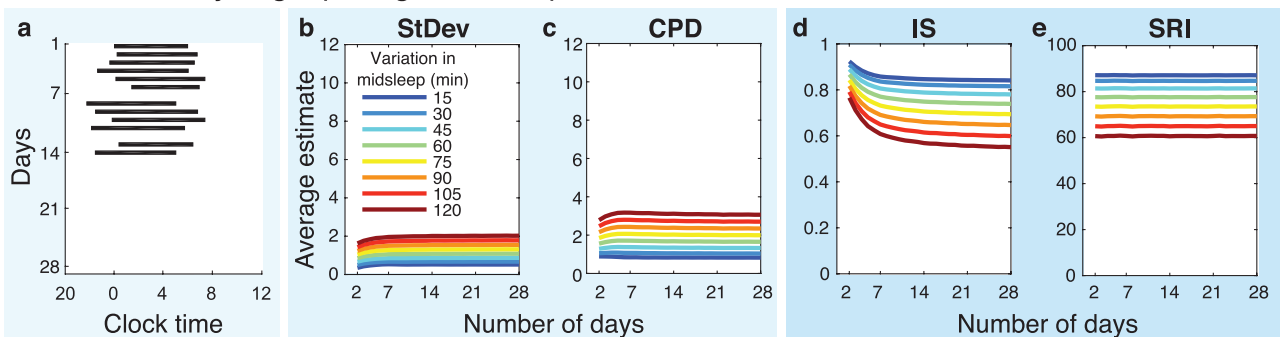
metrics. The CIs of consecutive metrics were generally wider than those of the corresponding overall metrics (Figure 5, G–J). CIs of CPD were between 1.6- and 2.4-times wider than those of StDev, irrespective of number of days or level of variability (Figure 5, G, H). The difference between CIs of IS and SRI decreased with more days (Figure 5, I, J): 95% CIs of SRI were on average 39% larger for ≤ 7 days, 20% larger for 8–14 days, and 14% larger for >14 days, than those of IS.

Consecutive metrics require larger sample sizes than overall metrics

Scenario 6: sample size

For all metrics, estimated required sample sizes decreased with an increasing number of days, and were largest when comparing groups with small differences in sleep regularity (see Supplement C, Figure S3, A–P), as expected. To achieve the same level of statistical power, consecutive metrics required generally larger samples sizes than overall metrics. Averaging across all simulated study lengths, using CPD required on average 21% more participants than StDev; using SRI required on average 20% more participants than IS. These differences declined as more days of data were available. For example, using SRI required on average a 26% larger sample than IS with 7 days of data, but only 12% with 28 days of data (illustrated in Supplement C, Figure S3, U, X). The main factor driving this difference between overall and consecutive metrics was effect size: when the difference in sleep regularity between groups was large (e.g. ≥ 45 min difference in daily variation), estimated sample sizes differed negligibly between metrics (Supplement C, Figure S3, Q–X).

Scenario 6: Study length (average estimates)



Scenario 6: Study length (width of 95% CIs)

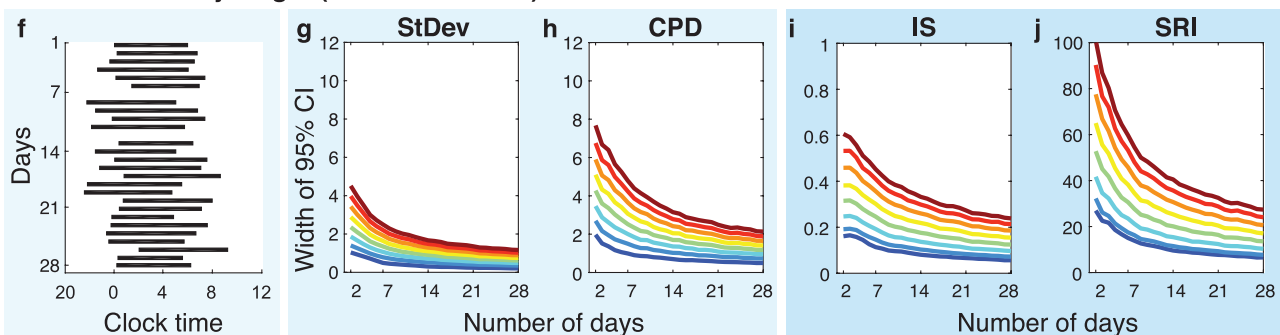


Figure 5. Impact of study length on sleep regularity metrics (Scenario 6). The far-left panel of each row shows example raster plots of the same sleep–wake pattern, based on (A) 14 days or (F) 28 days of data. Panels (B–E) Average estimates and panels (G–J) width of 95% confidence intervals (CIs) for sleep–wake patterns based on 2–28 days of data.

In [Supplementary Materials \(Tables S1 and S2\)](#), we provide detailed data tables with average estimates, 95% CIs, standard deviations, and sample size for each sleep regularity metric based on 2, 7, 14, 21, and 28 days of data, to help guide investigators' choice of metric, study length, and sample size.

Key points for Scenario 6

- Overall metrics, especially IS, overestimate sleep regularity, when based on ≤ 7 days.
- Consecutive metrics are more stable across study lengths, especially SRI, but require on average larger sample sizes to achieve the same confidence interval.
- Required sample sizes become more similar between metrics in longer studies or when the difference (Δ) in sleep regularity is large between groups ($\Delta \geq 45$ min).

Correspondence between overall and consecutive metrics depends on the source of variability

To determine how well the metrics correspond with one another, we plotted values of StDev against CPD (the paired daily-value metrics) and values of IS against SRI (the paired whole-signal metrics) for Scenarios 2–6 in [Figure 6](#). StDev and CPD corresponded closely under most conditions, with generally larger CPD than StDev values by a factor of ~ 1.4 ($\approx \sqrt{2}$, as expected given CPD's use of vector lengths) ([Figure 6, A](#)). However, there was no correspondence between the two metrics for naps ([Figure 6, B](#)): while StDev^{Nap} distinguished between numbers of naps (values moving from bottom to top with increasing number of naps), CPD^{Nap} distinguished between variations in midnap timing (values moving horizontally for a given number of naps). Agreement between StDev and CPD was overall good for all-nighters ([Figure 6, D](#)); however, values were least similar for very low or high numbers of nights with no sleep and most similar for numbers of all-nighters in the middle range (20–80%).

IS and SRI had almost perfect agreement for variations in midsleep, sleep duration ([Figure 6, G](#)), or naps ([Figure 6, H](#)). For awakenings ([Figure 6, I](#)) and all-nighters ([Figure 6, J](#)), IS and SRI showed a crescent-shaped relationship: IS and SRI values were initially similar, for example, for relatively short total WASO durations (i.e. ≤ 60 min) or a relatively low number of all-nighters (i.e. $\leq 10\%$ or three nights in four weeks), but diverged with increasing wakefulness, as IS decreased while SRI increased.

For Scenario 6, average estimates and width of 95% CIs generally became more similar between the metric pairs with an increasing number of days ([Figure 6, E, F, K, L](#)). The largest differences were observed for study lengths of ≤ 7 days: average estimates of StDev and CPD corresponded at a ratio of ~ 1.6 after 7 days, and at their usual ratio of ~ 1.4 after 14 days ([Figure 6, E](#)); correspondence between IS and SRI values was achieved after 7 days ([Figure 6, K](#)). Ninety-five percent CIs were approximately twice as wide for CPD as for StDev after 7 days ([Figure 6, F](#)); they were generally wider for SRI than IS, but were within 15% of each other when based on >14 days of data ([Figure 6, L](#)). For sample sizes, the average number of additional participants required when using consecutive vs. overall metrics gradually decreased with an increasing number of days ([Supplement D, Figure S4](#)).

Key points for metric correspondence

- The relationship between overall and consecutive metrics is not always linear. For example, SRI values cannot be inferred from IS values, except in specific circumstances.
- Differences between overall and consecutive metrics are largest for short study lengths (≤ 1 week) and become gradually smaller with more days of data.

Discussion

Using large-scale simulation of sleep–wake patterns, we compared the performance of five sleep regularity metrics—SJI, StDev, IS, CPD, and SRI—under conditions that allowed us to determine their sensitivity to factors that affect measurements of sleep regularity. Our findings have important implications for usage of these metrics and the design of studies that prospectively employ them. We found that each metric captures different aspects of sleep regularity and can be classified accordingly. SJI is a measure of weekly but not daily sleep regularity. StDev and IS are overall metrics that assess sleep regularity across multiple days, whereas CPD and SRI are consecutive metrics that assess sleep regularity from one day to the next. StDev and CPD are daily-value metrics that quantify regularity in a single aspect of sleep (e.g. midsleep time), whereas IS and SRI are whole-signal metrics that characterize regularity of the entire sleep–wake pattern. Only the whole-signal metrics (IS and SRI) are sensitive to naps and awakenings. Only the SRI is unbiased with respect to the number of days of study. And while StDev, CPD, and SRI measure the regularity of sleep–wake patterns, IS additionally assesses rhythmicity/amplitude. In the following sections, we discuss the theoretical and practical implications of these findings, to help guide investigators in choosing and appropriately using sleep regularity metrics.

Main characteristics of sleep regularity metrics

Weekly vs. daily metrics: SJI vs. StDev/IS/CPD/SRI

SJI reflects weekly but not daily changes in sleep While SJI was originally developed as a sleep-based metric for “living against one's biological clock” [28], it can also be framed as a measure of sleep regularity, given its comparison of sleep on workdays vs. free days [34]. SJI does indeed appear to be the best metric for purely assessing weekly changes in sleep. However, it does not capture daily changes within workdays or within free days. In this respect, it can be considered a complementary metric to the other four.

Overall vs. consecutive metrics: StDev/IS vs. CPD/SRI

Multi-day vs. day-to-day timescales The observation that overall metrics (StDev and IS) were not affected by scrambling the order of days highlights that they operate on a multi-day timescale. This is because these two metrics compare each day to the individual's average sleep–wake pattern, meaning the ordering of days is irrelevant. In contrast, consecutive metrics (CPD and SRI) operate on a day-to-day basis, specifically quantifying changes between consecutive days. The use of different timescales has important biological implications. Cases where a consecutive sleep regularity metric is predictive of an outcome but an overall metric is not are potentially informative

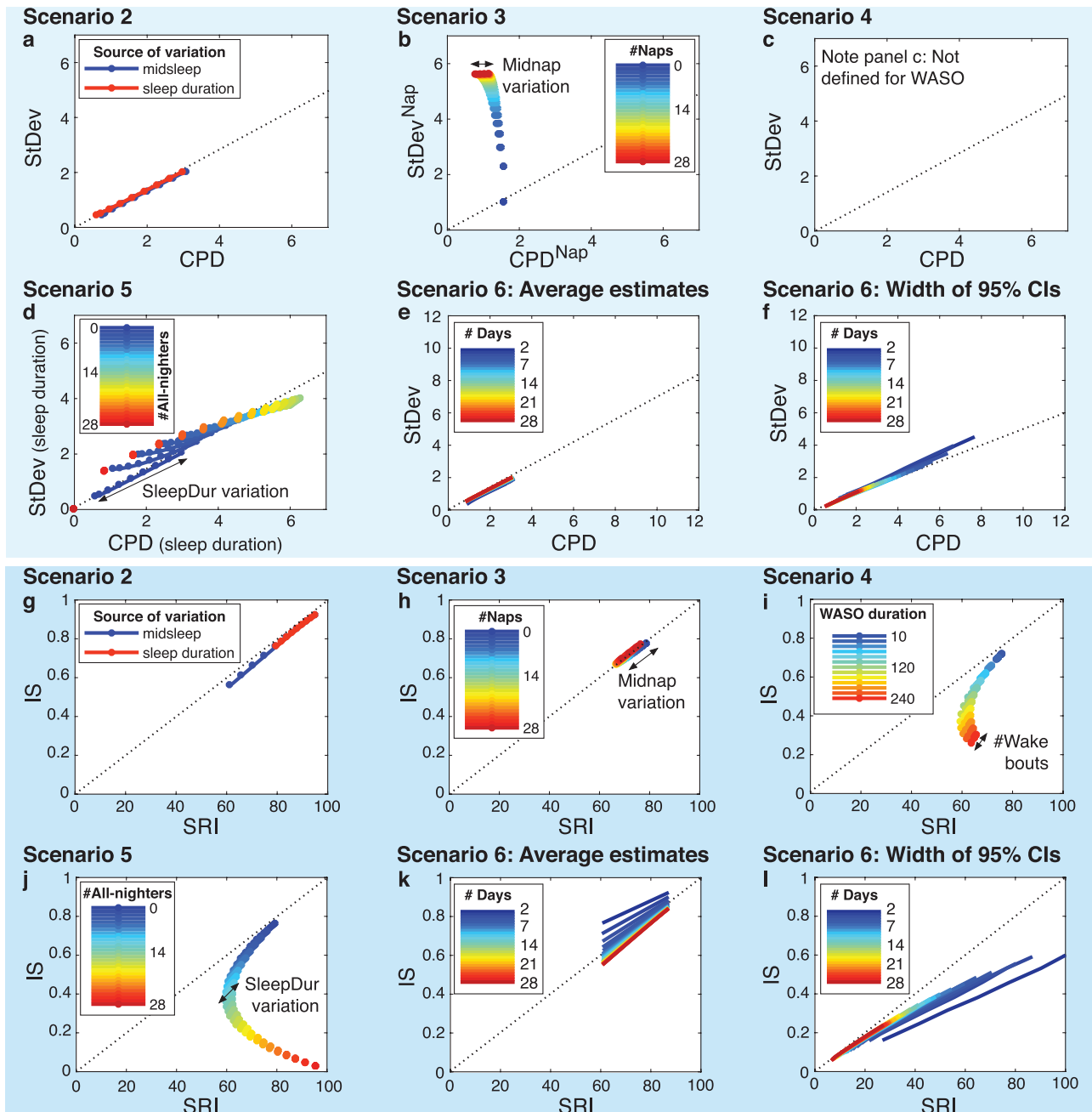


Figure 6. Correspondence between sleep regularity metrics: StDev vs. CPD and IS vs. SRI. The upper six panels show values of StDev and CPD plotted against each other. The lower six panels show values of IS and SRI plotted against each other. Results are shown for Scenario 2 (A, G), Scenario 3 (naps) (B, H), Scenario 4 (awakenings) (C, I), Scenario 5 (all-nighters) (D, J), and Scenario 6 (study length) (C-F, K-L). Note that StDev and CPD cannot be calculated for sleep-wake patterns with awakenings, so panel C is empty. The dotted lines all have a zero-intercept but the slope equals 1 in the lower six panels (G-L) and 0.7/0.5 in the upper six panels (panels A-E/panel F), corresponding to the factors by which values of StDev are generally smaller than those of CPD. The colored lines in each panel represent the different levels in a given variable, for example, number of naps in panel B. The arrows represent the effect of varying a second variable, for example, in panel B the spread in points of the same color (for each number of naps) is due to variation in midnap timing.

to the underlying mechanism (i.e. circadian disruption). For instance, the health impacts of sleep patterns with low IS values vs. sleep patterns with low SRI values are likely to mechanistically differ. The consecutive metrics CPD and SRI have also been used to examine effects of irregular sleep on multiple timescales, disaggregating within-subject (daily changes) from between-subject effects (average differences) in regression models [35]. In a recent study of 223 US college students, CPD averaged across ~30 days (but not average sleep duration) was a

predictor of average well-being, while daily sleep duration (but not daily CPD) was a predictor of students' daily well-being [17]. The findings suggest that effects of sleep regularity and duration on well-being may act on different timescales, with some effects accumulating over longer timescales.

Study length Because overall metrics are based on comparisons to the mean, their estimates can be biased when based on short study lengths. IS values showed the largest bias of all

metrics over the first 7 days, whereas SRI was independent of study length. IS tended to quantify short recordings as more regular than they would be scored based on a longer recording, especially when sleep-wake patterns were highly irregular. The use of IS in studies with relatively few days of data (≤ 7 days) should therefore be considered carefully, especially when comparing results between studies of differing lengths. StDev and CPD followed a similar, albeit less pronounced pattern: StDev values changed slightly more over the first 7 days than those of CPD. The finding that overall metrics tend to overestimate sleep regularity when based on ≤ 7 days has potentially far-reaching implications. Many epidemiological studies, where sleep assessments over long durations are not feasible, include study lengths of ≤ 7 days [36, 37]. Using overall metrics in these studies may underestimate the associations between irregular sleep and health outcomes [36]. While it may be possible to derive corrections to group-averages of StDev and IS values for short-duration studies, it would be more difficult to correct for the bias on an individual level. Future studies should explore potential remedies for the short-duration bias in overall metrics or consider this limitation in study and analysis planning.

Sample size In general, we found that consecutive metrics require larger sample sizes than overall metrics to achieve equivalent statistical power. We have provided detailed data tables in the [Supplemental Material](#) to help investigators determine appropriate values for their planned or finished studies, including mean values, 95% CIs, standard deviations, and sample sizes. The tables can be used for several purposes. For instance, most previous studies have used overall metrics, in particular StDev. Investigators may know what StDev to expect in their study population based on previous reports (e.g. sleep-wake variability has been consistently reported to be higher in younger than in older populations) [31, 38], but may be uncertain about the corresponding SRI value. Data in [Supplemental Table S1](#) provide such transformations among metrics, including matching standard deviations needed to determine sample sizes. [Supplemental Table S2](#) can be used in study design when investigating sleep regularity prospectively, allowing the study to be appropriately powered by balancing sample size, study length, and the type of metric. It can also guide post-hoc selection of metrics. We note that these power calculations are based on detecting a significant difference in sleep regularity between two groups. The required number of participants will likely be different when sleep regularity is not the outcome but the exposure variable. We do, however, still expect required sample sizes to be larger for consecutive than overall metrics, based on their wider 95% CIs.

Daily-value vs. whole-signal metrics: StDev/CPD vs. IS/SRI

Multiple vs. single sources of variability IS and SRI are whole-signal metrics, meaning they use the entire sleep-wake time series. As a consequence of this approach, they combine variability from both sleep timing and duration, as well as other sources of variability, including naps and WASO. Very distinct sleep-wake patterns can result in identical values of sleep regularity for these metrics, for example, an SRI value of 70 can be achieved by highly fragmented but otherwise stable sleep (same sleep onset and offset) or by non-fragmented sleep that varies from day to day. The fact that SRI and IS are whole-signal

measures possibly explains why they are useful predictors for a wide range of health outcomes that may have very different underlying etiology. This is both a strength and a weakness: they can detect lower regularity associated with a range of phenotypes but are consequently non-specific.

We found that both IS and SRI are more sensitive to changes in midsleep (with constant sleep duration) than to changes in sleep duration (with constant midsleep). This can be attributed to the fact that a 1-h change in sleep duration (constant midsleep) reduces the amount of overlap between two sleep episodes by 0.5 h at each end, whereas a 1-h change in midsleep (constant sleep duration) reduces the overlap by 1 h at each end.

Daily-value metrics, such as StDev and CPD, measure variability in a single aspect of sleep (e.g. midsleep time). As a result, they do not capture other forms of variability (e.g. naps and WASO), but their interpretation is consequently more straightforward. For example, a recent study examined the impact of irregular sleep on cardiovascular events by calculating StDev for sleep onset and sleep duration, and found that both were similarly associated with the outcome [36]. These metrics can also be more readily applied to data other than sleep; for example, calculating CPD when the timing but not duration of events is known, such as repeated cognitive performance, fatigue ratings, or caloric intake diaries [18]. Studies using daily-value metrics need to make a priori assumptions about where variability is expected or perform exploratory analyses involving multiple testing.

Both daily-value and whole-signal metrics can be calculated from any type of multi-day recording that enables estimation of sleep and wake times (e.g. sleep diaries, actigraphy, PSG). The whole-signal metrics IS and SRI require an assessment of sleep/wake state or activity across all 24 h of the multi-day recording. In practice, and in our simulations, this means that whole-signal metrics both require knowledge of and are sensitive to the timing of naps and awakenings. Sleep diary formats that collect less complete information, such as diaries that only measure the onset and offset timing of the main sleep episode and do not collect timing of naps or awakenings, may therefore systematically differ in their estimation of IS and SRI, compared with diary formats that do collect this information. A similar issue may arise with actigraphy data if naps are ignored, as some scoring algorithms only allow one sleep episode per 24 h [39]. In contrast, StDev and CPD metrics require only knowledge of the onset and offset timing of the main sleep episode. Notably, we found that IS and SRI values were sensitive to the total amount of time awake per night, rather than the specific timing/frequency of awakenings. In the case of diary formats where the nightly duration but not the specific timing of awakenings is collected, it may therefore be valid to treat the awakenings as equally distributed across the nighttime sleep episode for calculation of IS or SRI.

Sensitivity to naps or awakenings The whole-signal metrics, IS and SRI, are the preferred metrics when there are multiple sleep episodes (e.g. naps) and/or sleep-wake patterns appear to be fragmented by awakenings. The daily-value metrics, StDev and CPD, are not designed to properly assess such patterns, as they make assumptions about the structure of sleep (e.g. one main sleep episode per day). The relative sensitivity of IS and SRI to these patterns differed. Compared to a consolidated sleep-wake

pattern (no naps or awakenings), the occurrence of naps lowered sleep regularity by up to ~10% of the scale range for both IS and SRI, whereas awakenings had a relatively greater impact on IS (up to ~50% for IS vs. ~15% for SRI). These findings indicate that IS and SRI may differentially associate with outcomes in samples that involve non-consolidated and/or highly fragmented sleep (e.g. shift workers, older adults).

Sensitivity to “all-nighters” We found a striking difference in the behavior of IS compared with other sleep regularity metrics when simulating all-nighters. By increasing the number of all-nighters, we found that values of StDev, CPD, and SRI ultimately returned to a “perfectly regular” score for constant wakefulness (no sleep at all), whereas IS yielded monotonically lower (“less regular”) values. This occurred in the calculation of IS because the column variance (i.e. variance within the same epoch across all days) decreased at a faster rate than the overall variance. While this scenario is admittedly unrealistic, it uncovered a fundamental property of IS. Beyond regularity, IS also measures the *rhythmicity* of a sleep–wake pattern: for a sleep–wake rhythm to be considered regular by IS, alternating sleep and wake states are required, whereas for other metrics, consistency is sufficient. Under more common scenarios, where all-nighters occur only occasionally (less than 10% of nights), this property has little practical significance, but it may be relevant to metric performance under extreme conditions such as rotating shiftwork with extended (>24 h) work shifts [40] or in early-chronotype night-shift workers [41].

Strengths and limitations

In this paper, we used simulated sleep–wake patterns to illustrate fundamental properties of the sleep regularity metrics we considered. Rather than basing our analysis on any particular dataset, we used simplified scenarios to precisely identify the effects of specific factors on measurements of sleep regularity. This approach is both a strength and a weakness. The strength of this approach is that it enables determination of the precise quantitative effects of each factor, without results depending on idiosyncrasies or size of any particular sample. The weakness of this approach is that our results may not generalize to all populations. Real datasets may include co-variation between different sources of variance in ways that we have not modeled and may include different statistical distributions from those we have modeled. For example, rotating shift workers would not be expected to follow independent, normally distributed sleep patterns [42]. The ranges we have simulated at times also exceed levels of variability reported in the literature or are even unrealistic (e.g. 100% all-nighters); we chose to include such cases to determine how the metrics behaved under extreme conditions (e.g. if a U-shaped relationship would indeed return to the starting value).

The five metrics selected here to quantify sleep regularity are not exhaustive. Other metrics and approaches have been used in this space, including the coefficient of variation (CoV) [43], square of successive differences (SSD) [44, 45], mean range of n-day moving window [46], statistical modeling [47, 48], and other custom measures [49]. Another popular metric, often computed alongside IS, is intradaily variability (IV), which is used to assess sleep fragmentation, for example, by naps and awakenings [13]. We did not include IV here because our aim was to

compare metrics that assess variability between days, rather than variability within days. Our findings demonstrate that fragmented sleep–wake patterns can be a source of variability between days. Studies in samples that involve sleep fragmentation are encouraged to include IV alongside sleep regularity metrics to help discriminate between within-day and between-day variability.

Conclusion

Our findings demonstrate that deciding which sleep regularity metric is most appropriate for a given study depends on a combination of the type of data gathered, the study parameters (length and sample size), and which aspects of sleep regularity are considered most pertinent to the research question. We have exemplified such a decision process in [Figure 7](#).

Type of data

Figure 7 (left panel): In cases where the whole epoch-by-epoch sleep–wake signal cannot be reconstructed (e.g. limited assessments of sleep timing), investigators will be restricted to using daily-value metrics for sleep regularity (StDev and CPD). In cases where the whole sleep–wake signal is available, investigators should prefer whole-signal metrics for sleep regularity (IS and SRI) if naps and/or awakenings are an important feature to capture.

Study parameters

Figure 7 (middle panel): In the case of small samples or short studies, investigators should carefully consider our findings relating to statistical power for each metric, as well as bias in some sleep regularity metrics for assessments of ≤ 7 days. In cases where multiple metrics are suitable based on the type of data and study parameters, consecutive metrics (CPD and SRI) are more likely to capture the biological impact of irregular schedules via the circadian system, since day-to-day misalignment is what ultimately underpins circadian disruption [50].

Research question

Figure 7 (right panel): Finally, for more exploratory analyses, which are designed to understand the mechanisms that link irregularity to specific health outcomes, it may be worthwhile using multiple sleep regularity metrics to provide complementary information; for example, by understanding the degree to which irregularity is driven by SJL (daily vs. weekly variation). Which metrics to combine depends on which aspects of irregular sleep are assumed to drive the outcome of interest. If an investigator wishes to determine the contribution of sleep fragmentation (e.g. in populations with frequent awakenings or napping, such as patients with insomnia or neurodegenerative diseases, or young children), combining a daily-value metric (which is insensitive to fragmentation) and a whole-signal metric (which is sensitive to fragmentation) is recommended. If an investigator wishes to determine the relative importance of variability in sleep duration vs. sleep timing (e.g. in populations where late sleep times conflict with social demands to get up early) we propose

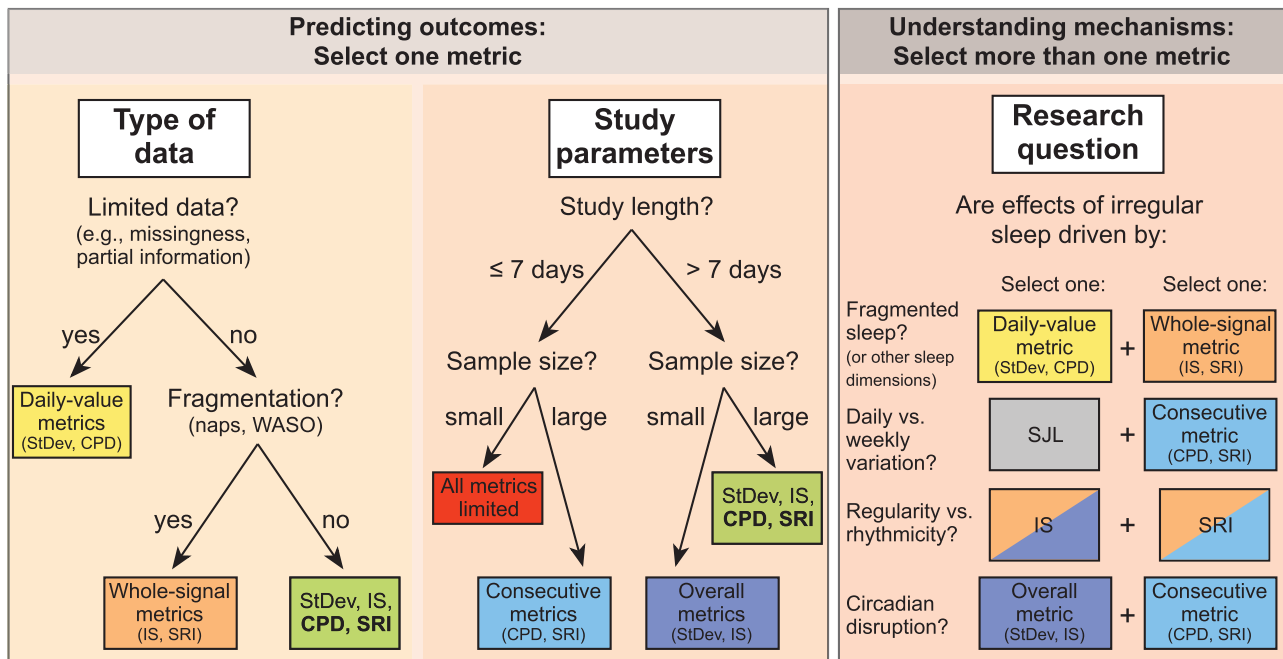


Figure 7. Choosing the right sleep regularity metric. Selecting the appropriate metric depends on the type of data, study parameters (length, sample size), and which aspects of sleep regularity are considered key to the research question. Decision trees are provided with respect to the type of data and study parameters, indicating the preferred metrics in each case. In cases where four metrics are all viable choices, the preferred metrics are bolded. With regards to the research question, investigators may wish to include more than one metric to test competing mechanisms, for example, including one daily-value metric and one whole-signal metric to test whether the effect of irregular sleep on an outcome is driven by sleep fragmentation by comparing their predictive values. StDev = intra-individual standard deviation; IS = interdaily stability; CPD = composite phase deviation; SRI = sleep regularity index; SJL = social jet lag; WASO = wake after sleep onset.

calculating a daily-value metric (e.g. CPD) separately for sleep timing and sleep duration. In populations with frequent all-nighters, such as college students and rotating shift workers, researchers may wish to compare IS and SRI, to determine the degree to which effects are driven by rhythmicity vs. regularity of sleep-wake patterns, given the metrics' differing responses to all-nighters. Finally, in cases where an investigator wishes to determine the importance of circadian disruption (e.g. in night shift workers), we suggest combining one overall metric with one consecutive metric, due to the different timescales they capture. Although we give examples of specific populations here, it is important to note that the choice of metrics to combine is ultimately driven by the research question, not the research population.

Sleep regularity metrics have added an important dimension to the measurement of sleep behaviors. Sleep regularity associates with a wide range of outcomes, often in cases where other dimensions of sleep behavior do not. This is possibly because irregular sleep can signify certain aspects of circadian disruption, which traditional sleep metrics do not capture. Sleep regularity can be interpreted as a proxy for variability in the timing of zeitgebers, such as light and meals. Our understanding of sleep regularity and its associations with health will therefore be further enriched by taking additional parallel measures of behaviors that co-vary with sleep schedules, including new ways of measuring variability in light exposure [51] or meal timing [18]. Analyses of these and other behaviors and timescales could be performed using methods similar to those we described here for sleep regularity. Each of these facets of behavior will ultimately help to map the mechanistic pathways from irregular schedules to poor health outcomes.

Funding

National Institutes of Health (NIH) (K24HL105664, R01GM105018, R01HL128538, P01AG009975). D.F. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – FI 2275/1-1, FI 2275/2-1. A.J.K.P. received a Turner Institute for Brain and Mental Health Strategic Research Grant.

Disclosure Statement

Financial Disclosure: E.B.K. has received travel reimbursement and/or meeting registration waiver from the DGSM, the Gordon Research Conference, the Santa Fe Institute, the Sleep Research Society, World Conference of Chronobiology and the National Sleep Foundation and Sanofi-Genzyme, has received grant review compensation from the Puerto Rico Trust, and consulted for the National Sleep Foundation. A.J.K.P. is an investigator on projects in the CRC for Alertness, Safety, and Productivity, and has received funding from Delos Inc. and Versalux Pty. Ltd. Non-financial Disclosure: none.

References

1. Bei B, Wiley JF, Trinder J, Manber R. Beyond the mean: a systematic review on the correlates of daily intraindividual variability of sleep/wake patterns. *Sleep Med Rev.* 2016;28:108–124. <https://doi.org/10.1016/j.smrv.2015.06.003>.
2. Makarem N, Zuraikat FM, Aggarwal B, Jelic S, St-Onge M-P. Variability in sleep patterns: an emerging risk factor for hypertension. *Curr Hypertens Rep.* 2020;22(2):19. <https://doi.org/10.1007/s11906-020-1025-9>.

3. Rodríguez-Colón SM, He F, Bixler EO, Fernandez-Mendoza J, Vgontzas AN, Calhoun S, et al. Sleep variability and cardiac autonomic modulation in adolescents – Penn State Child Cohort (PSCC) study. *Sleep Med.* 2015;16(1):67–72. <https://doi.org/10.1016/j.sleep.2014.10.007>.
4. Okun ML, Reynolds CF, Buysse DJ, Monk TH, Mazumdar S, Begley A, et al. Sleep variability, health-related practices, and inflammatory markers in a community dwelling sample of older adults. *Psychosom Med.* 2011;73(2):142–150. <https://doi.org/10.1097/PSY.0b013e3182020d08>.
5. Patel SR, Hayes AL, Blackwell T, Evans DS, Ancoli-Israel S, Wing YK, et al. The association between sleep patterns and obesity in older adults. *Int J Obes.* 2014;38(9):1159–1164. <https://doi.org/10.1038/ijo.2014.13>.
6. Chontong S, Saetung S, Reutrakul S. Higher sleep variability is associated with poorer glycaemic control in patients with type 1 diabetes. *J Sleep Res.* 2016;25(4):438–444. <https://doi.org/10.1111/jsr.12393>.
7. Spruyt K, Molfese DL, Gozal D. Sleep duration, sleep regularity, body weight, and metabolic homeostasis in school-aged children. *Pediatrics.* 2011;127(2):e345–e352. <https://doi.org/10.1542/peds.2010-0497>.
8. Doane LD, Gress-Smith JL, Breitenstein RS. Multi-method assessments of sleep over the transition to college and the associations with depression and anxiety symptoms. *J Youth Adolesc.* 2015;44(2):389–404. <https://doi.org/10.1007/s10964-014-0150-7>.
9. Bernert RA, Hom MA, Iwata NG, Joiner TE. Objectively assessed sleep variability as an acute warning sign of suicidal ideation in a longitudinal evaluation of young adults at high suicide risk. *J Clin Psychiatry.* 2017;78(6):e678–e687. <https://doi.org/10.4088/JCP.16m11193>.
10. Kuula L, Pesonen A-K, Heinonen K, Kajantie E, Eriksson JG, Andersson S, et al. Naturally occurring circadian rhythm and sleep duration are related to executive functions in early adulthood. *J Sleep Res.* 2018;27(1):113–119. <https://doi.org/10.1111/jsr.12581>.
11. McBean AL, Montgomery-Downs HE. Timing and variability of postpartum sleep in relation to daytime performance. *Physiol Behav.* 2013;122:134–139. <https://doi.org/10.1016/j.physbeh.2013.09.003>.
12. Vanderlind WM, Beevers CG, Sherman SM, Trujillo LT, McGeary JE, Matthews MD, et al. Sleep and sadness: exploring the relation among sleep, cognitive control, and depressive symptoms in young adults. *Sleep Med.* 2014;15(1):144–149. <https://doi.org/10.1016/j.sleep.2013.10.006>.
13. Witting W, Kwa IH, Eikelenboom P, Mirmiran M, Swaab DF. Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease. *Biol Psychiatry.* 1990;27(6):563–572.
14. Wittmann M, Dinich J, Meroz M, Roenneberg T. Social Jetlag: misalignment of biological and social time. *Chronobiol Int.* 2006;23(1–2):497–509. <https://doi.org/10.1080/07420520500545979>.
15. Fischer D, Vetter C, Roenneberg T. A novel method to visualize and quantify circadian misalignment. *Sci Rep.* 2016;6:38601. <https://doi.org/10.1038/srep38601>.
16. Phillips AJK, Clerx WM, O'Brien CS, Sano A, Barger LK, Picard RW, et al. Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. *Sci Rep.* 2017;7(1):3216. <https://doi.org/10.1038/s41598-017-03171-4>.
17. Fischer D, McHill AW, Sano A, Picard RW, Barger LK, Czeisler CA, et al. Irregular sleep and event schedules are associated with poorer self-reported well-being in US college students. *Sleep.* 2020;43(6):1–12. <https://doi.org/10.1093/sleep/zsz300>.
18. McHill AW, Hilditch CJ, Fischer D, Czeisler CA, Garaulet M, Scheer FAJL, et al. Stability of the timing of food intake at daily and monthly timescales in young adults. *Sci Rep.* 2020;10(1):20849. <https://doi.org/10.1038/s41598-020-77851-z>.
19. Imam A, Winnebeck EC, Buchholz N, Froguel P, Bonnefond A, Solimena M, et al. Circadian, sleep and caloric intake phenotyping in type 2 diabetes patients with rare melatonin receptor 2 mutations and controls: a pilot study. *Front Physiol.* 2020;11:564140. <https://doi.org/10.3389/fphys.2020.564140>.
20. Watson LA, McGlashan EM, Hosken IT, Anderson C, Phillips AJK, Cain SW. Sleep and circadian instability in delayed sleep-wake phase disorder. *J Clin Sleep Med.* 2020;16(9):1431–6. <https://doi.org/10.5664/jcsm.8516>.
21. Murray JM, Phillips AJK, Magee M, Sletten TL, Gordon C, Lovato N, et al. Sleep regularity is associated with sleep-wake and circadian timing, and mediates daytime function in delayed sleep-wake phase disorder. *Sleep Med.* 2019;58:93–101. <https://doi.org/10.1016/j.sleep.2019.03.009>.
22. Cohen S, Fulcher BD, Rajaratnam SMW, Conduit R, Sullivan JP, Hilaire MAS, et al. Sleep patterns predictive of daytime challenging behavior in individuals with low-functioning autism. *Autism Res.* 2018;11(2):391–403. <https://doi.org/10.1002/aur.1899>.
23. Sano A, Yu AZ, McHill AW, Phillips AJK, Taylor S, Jaques N, et al. Prediction of Happy-Sad mood from daily behaviors and previous sleep history. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2015:6796–6799. <https://doi.org/10.1109/EMBC.2015.7319954>.
24. Brooks AT, Raju S, Barb JJ, Kazmi N, Chakravorty S, Krumlauf M, et al. Sleep regularity index in patients with alcohol dependence: daytime napping and mood disorders as correlates of interest. *Int J Environ Res Public Health.* 2020;17(1):331–346. <https://doi.org/10.3390/ijerph17010331>.
25. Pye J, Phillips AJ, Cain SW, Montazerolghaem M, Mowszowski L, Duffy S, et al. Irregular sleep-wake patterns in older adults with current or remitted depression. *J Affect Disord.* 2021;281:431–437. <https://doi.org/10.1016/j.jad.2020.12.034>.
26. Mascaró L, Phillips AJK, Clark JW, Straus LD, Drummond SPA. Diurnal rhythm robustness in individuals with PTSD and insomnia and the association with sleep. *J Biol Rhythms.* 2021;36(2):185–195. <https://doi.org/10.1177/0748730420984563>.
27. Lunsford-Avery JR, Engelhard MM, Navar AM, Kollins SH. Validation of the sleep regularity index in older adults and associations with cardiometabolic risk. *Sci Rep.* 2018;8(1):14158. <https://doi.org/10.1038/s41598-018-32402-5>.
28. Roenneberg T, Pilz LK, Zerbini G, Winnebeck EC. Chronotype and social jetlag: A (self-) critical review. *Biology.* 2019;8(3):54. <https://doi.org/10.3390/biology8030054>.
29. Roenneberg T, Kuehnle T, Juda M, Kantermann T, Allebrandt K, Gordijn M, et al. Epidemiology of the human circadian clock. *Sleep Med Rev.* 2007;11(6):429–438. <https://doi.org/10.1016/j.smrv.2007.07.005>.
30. Roenneberg T, Allebrandt KV, Meroz M, Vetter C. Social jetlag and obesity. *Curr Biol.* 2012;22(10):939–943. <https://doi.org/10.1016/j.cub.2012.03.038>.
31. Dillon HR, Lichstein KL, Dautovich ND, Taylor DJ, Riedel BW, Bush AJ. Variability in self-reported normal sleep across the

- adult age span. *J Gerontol B Psychol Sci Soc Sci*. 2015;70(1):46–56. <https://doi.org/10.1093/geronb/gbu035>.
32. Petrov ME, Weng J, Reid KJ, Wang R, Ramos AR, Wallace DM, et al. Commuting and sleep: results from the hispanic community health study/study of Latinos Sueño Ancillary Study. *Am J Prev Med*. 2018;54(3):e49–e57. <https://doi.org/10.1016/j.amepre.2017.11.006>.
 33. Cha W. pwrAB: power analysis for AB testing. R package version 0.1.0. 2017.
 34. Zuraikat FM, Makarem N, Redline S, Aggarwal B, Jelic S, St-Onge M-P. Sleep regularity and cardiometabolic health: is variability in sleep patterns a risk factor for excess adiposity and glycemic dysregulation? *Curr Diab Rep*. 2020;20(8):38. <https://doi.org/10.1007/s11892-020-01324-w>.
 35. Curran PJ, Bauer DJ. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu Rev Psychol*. 2011;62:583–619. <https://doi.org/10.1146/annurev.psych.093008.100356>.
 36. Huang T, Mariani S, Redline S. Sleep irregularity and risk of cardiovascular events: the multi-ethnic study of atherosclerosis. *J Am Coll Cardiol*. 2020;75(9):991–999. <https://doi.org/10.1016/j.jacc.2019.12.054>.
 37. Savin KL, Patel SR, Clark TL, Bravin JI, Roesch SC, Sotres-Alvarez D, et al. Relationships of sleep duration, midpoint, and variability with physical activity in the HCHS/SOL Sueño ancillary study. *Behav Sleep Med*. 2020;18:1–12. <https://doi.org/10.1080/15402002.2020.1820335>.
 38. Shoji KD, Tighe CA, Dautovich ND, McCrae CS. Beyond mean values: Quantifying intraindividual variability in pre-sleep arousal and sleep in younger and older community-dwelling adults. *Sleep Sci*. 2015;8(1):24–30. <https://doi.org/10.1016/j.slsci.2015.02.005>.
 39. Hees VT van, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, et al. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. *PLoS One*. 2015;10(11):e0142533. <https://doi.org/10.1371/journal.pone.0142533>.
 40. Barger LK, Sullivan JP, Blackwell T, O'Brien CS, St Hilaire MA, Rahman SA, et al. Effects on resident work hours, sleep duration, and work experience in a randomized order safety trial evaluating resident-physician schedules (ROSTERS). *Sleep*. 2019;42(8):zsz110. <https://doi.org/10.1093/sleep/zsz110>.
 41. Vetter C, Fischer D, Matera JL, Roenneberg T. Aligning work and circadian time in shift workers improves sleep and reduces circadian disruption. *Curr Biol*. 2015;25(7):907–911. <https://doi.org/10.1016/j.cub.2015.01.064>.
 42. Boivin DB, Boudreau P. Impacts of shift work on sleep and circadian rhythms. *Pathol Biol*. 2014;62(5):292–301. <https://doi.org/10.1016/j.patbio.2014.08.001>.
 43. Breneman CB, Kline CE, West DS, Sui X, Porter RR, Bowyer KP, et al. The effect of moderate-intensity exercise on nightly variability in objectively measured sleep parameters among older women. *Behav Sleep Med*. 2019;17(4):459–469. <https://doi.org/10.1080/15402002.2017.1395337>.
 44. Chung K-F, Poon YPY-P, Ng T-K, Kan C-K. Correlates of sleep irregularity in schizophrenia. *Psychiatry Res*. 2018;270:705–714. <https://doi.org/10.1016/j.psychres.2018.10.064>.
 45. Burgess HJ, Park M, Wyatt JK, Rizvydeen M, Fogg LF. Sleep and circadian variability in people with delayed sleep-wake phase disorder versus healthy controls. *Sleep Med*. 2017;34:33–39. <https://doi.org/10.1016/j.sleep.2017.02.019>.
 46. Roane BM, Seifer R, Sharkey KM, Van Reen E, Bond TLY, Raffray T, et al. What role does sleep play in weight gain in the first semester of University? *Behav Sleep Med*. 2015;13(6):491–505. <https://doi.org/10.1080/15402002.2014.940109>.
 47. Ong Jason C., Hedeker Donald, Wyatt James K., Manber Rachel. Examining the variability of sleep patterns during treatment for chronic insomnia: application of a location-scale mixed model. *J Clin Sleep Med*. 2016;12(6):797–804. <https://doi.org/10.5664/jcsm.5872>.
 48. Bei B, Seeman TE, Carroll JE, Wiley JF. Sleep and physiological dysregulation: a closer look at sleep intraindividual variability. *Sleep*. 2017;40(9):zsz109. <https://doi.org/10.1093/sleep/zsz109>.
 49. Knutson KL, Rathouz PJ, Yan LL, Liu K, Lauderdale DS. Intraindividual daily and yearly variability in actigraphically recorded sleep measures: the CARDIA study. *Sleep*. 2007;30(6):793–796. <https://doi.org/10.1093/sleep/30.6.793>.
 50. Vetter C. Circadian disruption: What do we actually mean? *Eur J Neurosci*. 2018;51(1):531–550. <https://doi.org/10.1111/ejn.14255>.
 51. Cain SW, McGlashan EM, Vidafar P, Mustafovska J, Curran SPN, Wang X, et al. Evening home lighting adversely impacts the circadian system and sleep. *Sci Rep*. 2020;10(1):19110. <https://doi.org/10.1038/s41598-020-75622-4>.