

Capturing Real-World Habitual Sleep Patterns with a Novel User-centric Algorithm to Pre-Process Fitbit Data in the All of Us Research Program: Retrospective observational longitudinal study.

Hiral Master, Jeffrey Annis, Jack Ching, Karla Gleichauf, Lide Han, Peyton Coleman, Kelsie Full, Neil Zheng, Douglas Ruderfer, John Hernandez, Logan Schneider, Evan Brittain, Evan Brittain

Submitted to: Journal of Medical Internet Research
on: January 24, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 52

 Figures 53

 Figure 1..... 54

 Figure 2..... 55

 Figure 3..... 56

 Figure 4..... 57

 Figure 5..... 58

 Multimedia Appendixes 59

 Multimedia Appendix 1..... 60

Capturing Real-World Habitual Sleep Patterns with a Novel User-centric Algorithm to Pre-Process Fitbit Data in the All of Us Research Program: Retrospective observational longitudinal study.

Hiral Master^{1*} PhD, MPH, PT; Jeffrey Annis^{1*} PhD; Jack Ching^{2*} PhD; Karla Gleichauf² PhD; Lide Han¹ PhD; Peyton Coleman¹; Kelsie Full¹; Neil Zheng³; Douglas Ruderfer¹; John Hernandez²; Logan Schneider²; Evan Brittain¹; Evan Brittain¹

¹ Vanderbilt University Medical Center Nashville US

² Google (United States) Mountain View US

³ Yale University New Haven US

*these authors contributed equally

Corresponding Author:

Evan Brittain

Vanderbilt University Medical Center
Nashville
US

Abstract

Background: Commercial wearables like Fitbits quantify sleep metrics using fixed calendar times as the default measurement periods, which may not adequately account for individual variations in sleep patterns. To address this, experts in sleep medicine and wearables developed a user-centric algorithm that more accurately reflects actual sleep behaviors, aiming to improve wearable-derived sleep metrics.

Objective: The study aimed to describe the development of the new (user-centric) algorithm, and how it compares with the default (calendar-relative), and offers best practices for analyzing All of Us Fitbit sleep data on a cloud platform.

Methods: The default and new algorithms was implemented to pre-process and then compute sleep metrics related to schedule, duration, and disturbances using high-resolution Fitbit sleep data from 8,563 participants (median age 58.1 years, 72% female) in the All of Us Research Program (v7 Controlled Tier). Variation in typical sleep patterns was computed by taking the differences in the mean number of primary sleep logs classified by each algorithm. Linear mixed-effects models were used to compare differences in sleep metrics across quartiles of variation in typical sleep patterns.

Results: Out of 8,452,630 total sleep logs over a median of 4.2 years of Fitbit monitoring, 401,777 (5%) non-primary sleep logs identified by default algorithm were reclassified to primary sleep by the user-centric algorithm. Variation in typical sleep patterns ranged from -0.08 to 1. Among participants with the most variation in typical sleep patterns, the new algorithm identified more total sleep time (by 17.6 minutes; $P < 0.001$), more wake after sleep onset (by 13.9 minutes; $P < 0.001$), and lower sleep efficiency (by 2.0%; $P < 0.001$), on average. There were only modest differences in sleep stage metrics between the two algorithms.

Conclusions: The user-centric algorithm captures the natural variability in sleep schedules, offering an alternative way to pre-process and evaluate sleep metrics related to schedule, duration, and disturbances. R package is publicly available to facilitate the implementation of this algorithm for clinical and translational use.

(JMIR Preprints 24/01/2025:71718)

DOI: <https://doi.org/10.2196/preprints.71718>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to the public.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://www.jmir.org/preprint/71718>, the full text will be available to the public.

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/preprint/71718>, the full text will be available to the public.



Original Manuscript

Capturing Real-World Habitual Sleep Patterns with a Novel User-centric Algorithm to Pre-Process Fitbit Data in the *All of Us* Research Program: Retrospective observational longitudinal study.

Authors:

Hiral Master^{1,2*}, Jeffrey Annis^{1-3*}, Jack H. Ching^{4*}, Karla Gleichauf⁴, Lide Han^{2,5}, Peyton Coleman^{2,6}, Kelsie Full⁷, Neil Zheng^{8,9}, Douglas M. Ruderfer^{2,5,6,10}, John Hernandez⁴, Logan D. Schneider⁴, Evan L. Brittain¹⁻³

**These co-authors contributed equally.*

Affiliations:

¹ Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA

² Center for Digital Genomic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

³ Division of Cardiovascular Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

⁴ Google, Mountain View, CA, USA

⁵ Division of Genetic Medicine, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA

⁶ Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

⁷ Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

⁸ Yale School of Medicine, Yale University, New Haven, CT, USA

⁹ Brigham and Women's Hospital, Boston, MA, USA

¹⁰ Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

Corresponding author: Evan L. Brittain, MD, MSc - evan.brittain@vumc.org

Abstract:

Background: Commercial wearables like Fitbits quantify sleep metrics using fixed calendar times as the default measurement periods, which may not adequately account for individual variations in sleep patterns. To address this, experts in sleep medicine and wearables developed a user-centric algorithm that more accurately reflects actual sleep behaviors, aiming to improve wearable-derived sleep metrics.

Objective: The study aimed to describe the development of the new (user-centric) algorithm, and how it compares with the default (calendar-relative), and offers a practical guide for analyzing *All of Us* Fitbit sleep data on a cloud platform.

Methods: The default and new algorithms was implemented to pre-process and then compute sleep metrics related to schedule, duration, and disturbances using high-resolution Fitbit sleep data from 8,563 participants (median age 58.1 years, 72% female) in the *All of Us* Research Program (v7 Controlled Tier). Variation in typical sleep patterns was computed by taking the differences in the mean number of primary sleep logs classified by each algorithm. Linear mixed-effects models were used to compare differences in sleep metrics across quartiles of variation in typical sleep patterns.

Results: Out of 8,452,630 total sleep logs over a median of 4.2 years of Fitbit monitoring, 401,777 (5%) non-primary sleep logs identified by default algorithm were reclassified to primary sleep by the user-centric algorithm. Variation in typical sleep patterns ranged from -0.08 to 1. Among participants with the most variation in typical sleep patterns, the new algorithm identified more total sleep time (by 17.6 minutes; $P<0.001$), more wake after sleep onset (by 13.9 minutes; $P<0.001$), and lower sleep efficiency (by 2.0%; $P<0.001$), on average. There were only modest differences in sleep stage metrics between the two algorithms.

Conclusion: The user-centric algorithm captures the natural variability in sleep schedules, offering an alternative way to pre-process and evaluate sleep metrics related to schedule, duration, and disturbances. R package is publicly available to facilitate the implementation of this algorithm for clinical and translational use.

Keywords: sleep; Fitbit; algorithms; *All of Us* Research Program

Introduction

Consumer wearables such as Fitbits are increasingly used in research studies to investigate the impact of biomarkers and lifestyle behaviors on health outcomes, supporting applications from screening to health management. A leading example is the National Institutes of Health funded *All of Us* Research Program, which collects longitudinal Fitbit data with privacy-preserving linkage to electronic health records (EHR) and other data types[1,2]. Studies based on this dataset have documented the association of Fitbit-derived activity and sleep metrics with the onset of chronic diseases[3,4]. Growing interest in wearables research has also given rise to recent industry standards to document best practices for the validation and use of digital health endpoints for physical activity[5,6], heart rate,[7] and sleep patterns[8]. In parallel, researchers have also developed pre-processing methods for activity to account for biases and limitations inherent to wearable-measured data in real-world settings[9,10]. Recent generation Fitbit devices, which employ heart rate variability and limb movement, have reported promising performance to measure standard sleep metrics such as wake after sleep onset (WASO), and total sleep time (TST) and classify sleep versus wake episodes more accurately than clinical actigraphy, which only accounts for limb movement[11,12]. Additional research is still needed, however, to have a diverse toolkit to pre-process and validate wearable-derived sleep metrics[13] especially for populations underrepresented in datasets used for algorithm development such as the elderly and people with multiple chronic health conditions[14]. This limitation has prompted professional clinical associations to advise caution with their use in clinical practice[15].

Sleep metrics accessible to researchers via the Fitbit Web Application Programming Interface (API) currently rely on a default sleep algorithm, *isMainSleep*, that was primarily designed for general consumer use. As a default, this algorithm relies on fixed calendar date and calculates summary sleep metrics based on primary sleep logs (the set of logs that belong to a user's primary sleep period), classified as those with the longest duration ending on each calendar day. While this approach generally performs well, the *isMainSleep* algorithm does not fully characterize primary sleep periods that contain interruptions involving at least an hour of wakefulness. These interruptions result in the creation of multiple sleep logs, and only the longest log would be reported as the primary sleep period[16].

Burgess et al[13] have elucidated similar limitations and developed automated sleep data cleaning that requires Fitbit sleep data from third-party platform (Fitabase) and self-reported sleep diaries. However, in this study, in collaboration with clinical and technical experts in sleep medicine from industry and academia, we propose a fully data-driven algorithm that does not require self-reported sleep logs or manual scoring for pre-processing the Fitbit sleep data, allowing it to be more scalable for sleep studies across large populations. Specifically, we propose a user-centric algorithm based on an individual's typical sleep period (TSP), which accounts for variability in habitual sleep patterns and has the potential for reducing noise in sleep data and more accurately capturing sleep-wake metrics (shown in **Table 1**). This paper details the development of the TSP algorithm, compares it with the default calendar-relative *isMainSleep* algorithm, and offers a practical guide for analyzing *All of Us* Fitbit sleep data on a cloud platform. We also examine the participants characteristics whose sleep behaviors differ from typical sleep patterns and make the analytic pipelines publicly available, through an open-source R package, to foster implementation of Fitbit sleep data processing for clinical and research use.

Table 1: Definitions and computations of key sleep metrics

Metric	Conceptual Definition	Technical Implementation
--------	-----------------------	--------------------------

		Calendar-relative “isMainSleep” algorithm	User-centric algorithm	“TSP”
Sleep Schedule				
Primary Sleep Period	Period of time during which a typical person obtains the majority of their sleep	Based on the sleep log with the longest duration on each midnight-to-midnight calendar day	Based on the median bedtime and wake times across all relevant sleep logs centered around a user’s median mid-sleep point	
Bedtime (BT)	Start of the primary sleep period as algorithmically determined by Fitbit; may not directly reflect the intent to sleep	Start time of the first sleep log in the primary sleep period.		
Sleep Onset	Time of transition from wakefulness to sleep	Start time of the first sleep segment in the primary sleep period, with the first occurrence of a level that is not ‘wake’, ‘awake’, or ‘restless’.		
Sleep Offset	Time of the final awakening	End time of the final sleep segment in the primary sleep period, after which there are no more levels that are not ‘wake’, ‘awake’, or ‘restless’.		
Wake time (WT)	End of the primary sleep period; corresponds to an “out of bed” or “lights on” time	End time of the final sleep log in the primary sleep period.		
Time attempting to sleep (TATS)	Duration of the primary sleep period between a user’s bedtime and wake time for a given day. This includes all sleep logs as well as any intervening non-sleeping time between sleep logs that are not captured by a sleep log	Duration between BT and WT	Duration between BT of the first sleep log and WT of the final sleep log, including any intervening non-sleep log time.	
Total sleep period duration	Duration between a user’s sleep onset and sleep offset for a given day.	Sleep offset datetime - sleep onset datetime		
Midsleep point (MSP)	Midpoint time between sleep onset and sleep offset	$(\text{Sleep onset datetime} + \text{sleep offset datetime}) / 2$		
Sleep Duration (within Primary Sleep Period)				
Total sleep time (TST) in primary sleep period	Total time spent asleep during the primary sleep period	Total duration of all sleep segments, where the level of the sleep segment is not ‘awake’, ‘wake’, or ‘restless’		
Total wakefulness duration	Total time spent awake during the primary sleep period	Total duration of all sleep segments, where the level of the sleep segment is	Total duration of all sleep segments, where the level of the sleep segment is ‘awake’, ‘wake’, or	

		‘awake’, ‘wake’, or ‘restless’	‘restless’, in addition to any periods of ‘imputed wake’ between sleep logs.
Sleep efficiency percentage	Total time spent asleep as a proportion of the primary sleep period	(TST /TATS) *100.	
Sleep onset latency (SOL)	Time between bedtime and sleep onset	Sleep Onset - BT Note: Because sleep logs are algorithmically initiated, it is often 0 for automatically created sleep logs.	
Latency from wake to end of log	Time between the final awakening to wake time	WT - Sleep Offset	
Sleep Disturbances (within Primary Sleep Period)			
Wake after sleep onset (WASO)	Total time spent awake after sleep onset and before the final awakening. Note: only includes wake periods > 3 minutes.	Total duration of all ‘wake’, ‘awake’ and ‘restless’ sleep segments between sleep onset and sleep offset	Total duration of all ‘wake’, ‘awake’ and ‘restless’ sleep segments between sleep onset and sleep offset, as well as intervening non-sleep time between sleep logs
Number of awakenings (NWAK)	Total number of contiguous periods spent awake after sleep onset. Note: only includes wake periods > 3 minutes.	Count of all contiguous ‘wake’, ‘awake’ or ‘restless’ sleep segments between sleep onset and sleep offset	Count of all contiguous ‘wake’, ‘awake’ and ‘restless’ sleep segments between sleep onset and offset, as well as intervening non-sleep periods between sleep logs
Number of long awakenings	Total number of contiguous periods spent awake >=30 minutes after sleep onset, as a proxy measure of a meaningful awakening.	Count of all contiguous wake, awake and restless sleep segments >=30 minutes in duration between sleep onset and sleep offset	Count of all contiguous wake, awake and restless sleep segments between sleep onset and sleep offset, as well as any intervening non-sleep log periods between sleep logs >=30 minutes in duration
Length of longest wake	Duration of the longest contiguous period spent awake after sleep onset. Note: only includes wake periods > 3 minutes.	Maximum duration among all contiguous wake, awake and restless sleep segments between sleep onset and sleep offset	Maximum duration among contiguous wake, awake and restless sleep segments between sleep onset and sleep offset, as well as any intervening non-sleep log periods

			between sleep logs
Non-primary Sleep Period)			
Count of non-primary sleep period	Number of sleep logs outside of primary sleep period	Count of all sleep logs labeled “isMainSleep=false” that end on a calendar day	Count of all sleep logs labeled “TSP=false” that end prior to next BT
Total sleep time in non-primary sleep period (TST-NP)	Total duration of sleep across sleep logs outside of the primary sleep period	Total duration of all sleep logs labeled “isMainSleep=false” that end on a calendar day	Total duration of all sleep logs labeled “TSP=false” that end prior to next BT
Sleep Pattern and Stages			
Periods of wakefulness			
Awake	Total time spent awake (classic sleep) across all sleep segments.	Total duration of all sleep segments with level “Awake”	
Restless	Total time spent restless (classic sleep) across all sleep segments.	Total duration of all sleep segments with level “Restless”	
Wake	Total time spent in wake (staged sleep) across all sleep segments.	Total duration of all sleep segments with level “Wake”	
Imputed wake	Period of time spent awake between sleep logs	Not applicable	Total duration of intervening time between “TSP=true” sleep logs
Periods of sleep			
Asleep	Total time spent asleep (classic sleep) across all sleep segments.	Total duration of all sleep segments with level “Asleep”	
Light	Total time spent in light stage sleep across all sleep segments.	Total duration of all sleep segments with level “Light”	
Deep	Total time spent in deep stage sleep across all sleep segments.	Total duration of all sleep segments with level “Deep”	
REM	Total time spent in REM stage sleep across all sleep segments.	Total duration of all sleep segments with level “REM”	

Methods

Study design and population

Retrospective observational longitudinal study using secondary data analysis was employed to

implement the algorithms for preprocessing Fitbit sleep data. We used *All of Us* participants' deidentified data that were made available to researchers in the version 7 controlled tier (C2022Q4R13) on the *All of Us* Researcher Workbench - a secured cloud-based platform[17]. Detailed description of this program has been published elsewhere[2]. Data for this study was derived from a convenience sample consisting of participants who owned Fitbit device and were able to contribute their Fitbit data since the time they created Fitbit account under the "Bring Your Own Device" (BYOD) Program. Specifically, in this study, we included 8,563 participants who provided EHR data and had any Fitbit sleep data (described below) from 2009 to 2022.

Ethical Considerations

The protocol for the human participant research conducted was reviewed by the institutional review board of the All of Us Research Program (protocol 2021-02-TN-001). Data security and participant confidentiality were safeguarded through multiple measures: secure data storage systems were implemented, access to identifying information was restricted, and confidentiality protocols were mandated through contractual agreements. Access to the Researcher Workbench was limited to authorized personnel by whom mandatory training had been completed and whose institutions were covered by active data use agreements. Compensation of US \$25 was provided via a cash payment, gift card, or electronic voucher when biological specimens (blood, saliva, or urine) were collected at designated partner facilities. To protect participant privacy in accordance with *All of Us* Research Program's Data and Statistics Dissemination Policy, reporting of value that corresponds to a specific, small number of participants ($n < 20$) or any other values that would allow to deduce this information for sample < 20 using mathematical formulae were not presented.

In this study, only the authorized authors who completed All of Us Responsible Conduct of Research training accessed the deidentified data from the Researcher Workbench (a secured cloud-based platform). Since the authors were not directly involved with the participants, Institutional Review Board review was exempted.

All of Us Fitbit Sleep data

De-identified daily summaries of Fitbit sleep data as well as higher-resolution, sequence-level sleep data obtained via the Fitbit Web API were made available on the Researcher Workbench, through a multi-step data ingestion and curation process detailed elsewhere[1]. While Fitbit provides additional data on short wake periods (which represent physiological awakenings that are 3 minutes or less) the TSP algorithm was implemented using only the sequence-level data currently available on the *All of Us* Researcher Workbench, which captures sleep periods of at least 1 hour, and intervening wake periods during sleep greater than 3 minutes. The sequence-level sleep data additionally provides detailed breakdown of sleep periods either into accelerometry-derived classic (non-staged) sleep patterns (i.e., awake, restless, asleep) or sleep stages (i.e., wake, light, deep, and REM), the latter of which are available on sleep periods 3 hours or longer and leverage heart rate data. This Fitbit data was retrospective and longitudinal in nature, and includes all available data since the time participants created their Fitbit accounts.

Resolutions of Fitbit sleep data

For ease of exposition, we first define the different resolutions over which Fitbit sleep data can be analyzed. At the broadest level, sleep can be conceptualized as encompassing an entire sleep period. A single sleep period may contain one or more sleep logs, including any interruptions that may have occurred (which corresponds to intervening wakefulness of ≥ 1 hour between sleep logs).

Each sleep log consists of a consecutive sequence of sleep segments that are separated by less than one hour, and may be classified as primary sleep or non-primary sleep. Each constituent sleep segment is associated with a sleep level, and has a start timestamp and duration. The sleep level indicates whether the sleep segment represents a period of sleep ('asleep' from Fitbit's classic sleep pattern algorithm; 'light', 'deep', or 'REM' from Fitbit's sleep staging algorithm) or wakefulness ('awake' or 'restless' from Fitbit's classic sleep pattern algorithm; 'wake' from Fitbit's sleep staging algorithm).

Primary sleep period

In this paper, we introduce the term "primary sleep period" to refer to the period of time during which a typical person obtains the majority of their sleep. Accurate determination of the primary sleep period is particularly important for assessing related sleep-wake metrics shown in **Table 1**. Most canonical "regular" sleepers with a daytime schedule will have such primary sleep periods between 11PM and 7AM. However, some people have less clear-cut sleep patterns. For example, a night shift worker may aim to sleep immediately after work (e.g., 8AM-1PM), and then take naps between 5PM and 7PM. Conceptually, we would like to label the hours between 8AM-1PM as their primary sleep period for that day and any time outside of the primary sleep period as their non-primary sleep period. In this instance, the time between 5PM-7PM would be considered as total sleep time during the non-primary sleep period.

A second example involves participants with significant disruptions in their sleep (e.g., those who fall asleep from 10PM-11:30PM, are awake in bed until 1:30AM, and asleep again until 6:30AM). In such a case, we believe that it is reasonable to define their primary sleep period as being from 10PM to 6:30AM, even though it was interrupted by a prolonged 2-hour period of wakefulness. The underlying assumption for stitching these two disjointed sleep logs into one (10PM-6:30AM) sleep period under the user-centric algorithm is that if a sleep log is at least 1 hour in duration and either overlaps with the typical sleep period or ends within 1 hour of sleep occurring during the typical sleep period, it likely reflects an intent to sleep. In such cases, the sleep log should be stitched into the primary sleep period, even if there is a significant intervening wakefulness period (e.g., likely due to insomnia). Additionally, it may be appropriate to include this sleep in the primary sleep period, given the sleep health implications. To clarify, inadvertent dozing doesn't tend to happen for over 1 hour, particularly in "the zone of maximum circadian drive for wakefulness", i.e., right before typical bedtime. Therefore, it is reasonable to account for such sleep segments as part of the primary physiologic and/or behavioral sleep period, to capture unhealthy sleep habits and provide insights into holistic sleep architecture.

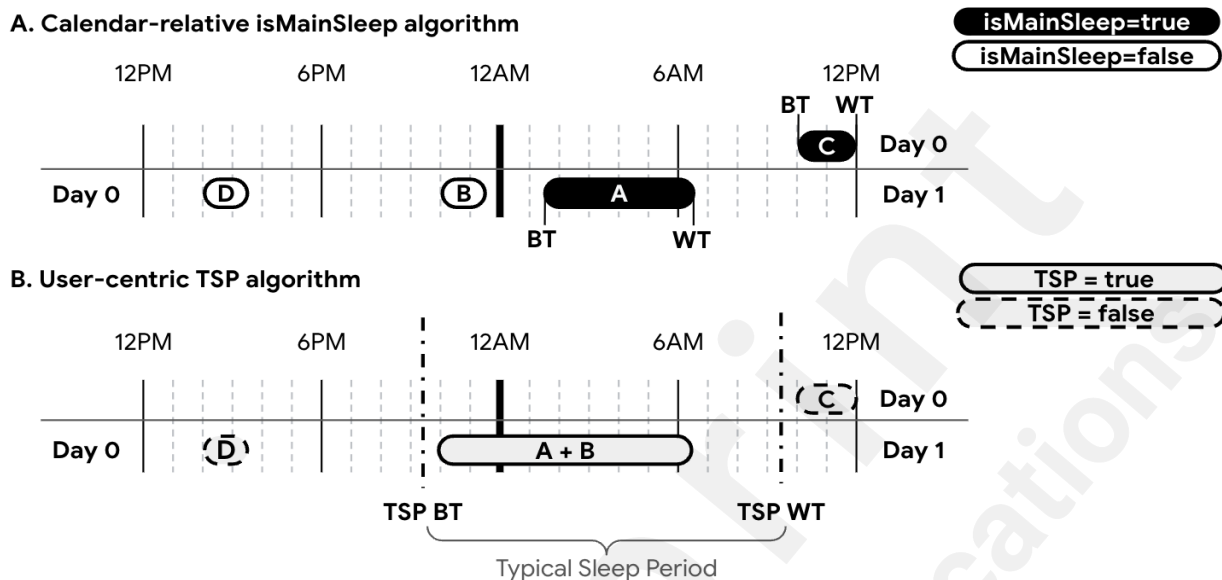
Primary sleep log classification under isMainSleep and TSP algorithms

An individual's sleep may be fragmented and distributed over time. In a conceptual example illustrated in **Figure 1**, there are four sleep logs spanning 2 calendar days, i.e. C (Day 0; 10am-12pm), D (Day 0; 2-3:30pm), B (Day 0; 10-11:30pm) and A (Day 1; 1:30-6:30 am). In order to accurately compute sleep metrics, it is imperative to capture all primary sleep logs that fall within the primary sleep period, and distinguish them from non-primary sleep logs.

As shown in **Figure 1A**, isMainSleep classifies sleep logs C and A as the primary sleep on days 0 and 1, respectively, since they are the sleep logs with the longest durations on each respective midnight-to-midnight calendar day. In contrast to isMainSleep, the TSP algorithm (technical description below) follows a user-centric approach and is agnostic to time-of-day (i.e. not dependent on a fixed calendar time window). TSP classifies both sleep logs A and B as primary sleep on Day 1

(Figure 1B). It also re-classifies sleep log C as non-primary sleep on Day 0 because it falls outside of the typical sleep period. Since sleep log D is neither the longest duration sleep log on Day 1, nor does it overlap with the typical sleep period, both the isMainSleep and TSP algorithms classify it as non-primary sleep on Day 1.

Figure 1A-B: Conceptual overview of how the calendar-relative “isMainSleep” and user-centric “TSP” algorithms handle high-resolution sleep data.



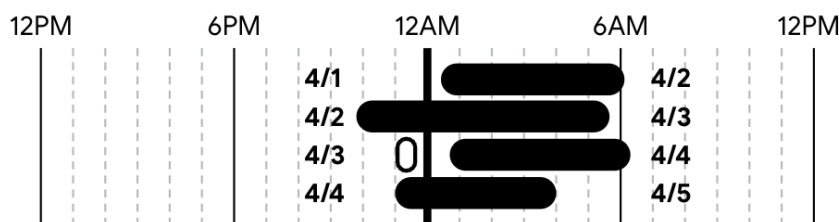
TSP BT and TSP WT are derived from the median bedtime and wake time across all relevant sleep logs around a user's median mid-sleep point (MSP) (not shown). TSP = typical sleep period, BT = bedtime, WT = wake time.

Real-world data example

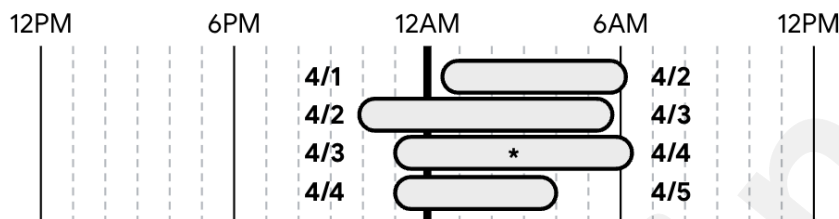
Figure 2 provides an illustrative example of the limitation of the isMainSleep algorithm based on actual Fitbit measured sleep data from one of the co-authors (LS) who has consented to share his data. In this example, the primary sleep period spanning 4/3 to 4/4 (April 3rd to April 4th) includes 2 sleep logs, each starting and ending on 4/3 and 4/4 respectively. The isMainSleep algorithm classifies the longer sleep log on 4/4 as primary sleep, but excludes the initial period of sleep that began on 4/3 (which is misclassified as a non-primary sleep on 4/3). Compared to the full primary sleep period, the isMainSleep algorithm incorrectly identifies a much later bedtime, shorter TATS, TST, WASO, fewer NWAK, and higher sleep efficiency percentage.

Figure 2A-B. Calendar-relative ‘isMainSleep’ vs.user-centric ‘TSP’ algorithm - an illustrative data example.

A. Calendar-relative isMainSleep algorithm



B. User-centric TSP algorithm



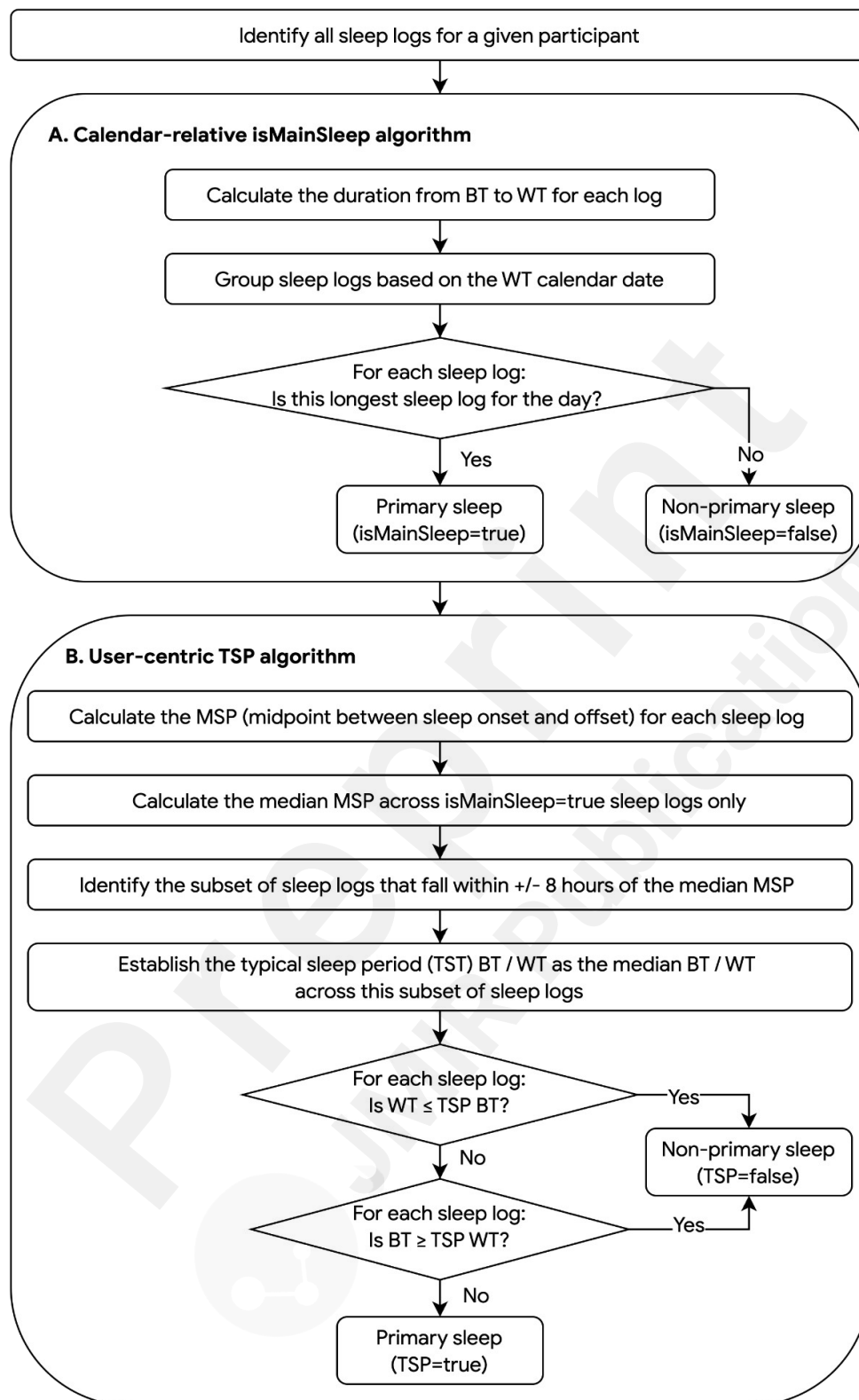
The data shown in this figure is sourced directly from the Fitbit Web API sleep record for one of the co-authors (LS) who has consented to share data for illustration purposes. Please note that non-primary sleep logs are not shown in the visualization as none were logged during this timeframe.

In contrast, the TSP algorithm classifies both sleep logs, and the intervening wake time between the logs as the primary sleep period which spans 4/3 to 4/4 (* in **Figure 2**). Note that the primary sleep periods ending on 4/2, 4/3, and 4/5 are correctly classified by the isMainSleep algorithm as they correspond to the longest duration sleep log ending on each calendar day so there is no change under TSP.

Technical implementation of both algorithms

The decision making process for classifying sleep logs using both algorithms are illustrated in Figure 3A-B. The calendar-relative isMainSleep algorithm identifies primary sleep as the longest sleep logs on each calendar day (as shown in Figure 3A). The user-centric TSP algorithm is agnostic of calendar date and aggregates logs based on a user's typical sleep period through a multi-step process. The algorithm first establishes the typical time range during which the user logs their longest sleep each day, and identifies primary sleep as those sleep logs that overlap with this range. (as shown in Figure 3b and described in detail in the method supplement).

Figure 3A-B: Decision making process for sleep log classification using calendar-relative and user-centric algorithms.



Variation from typical sleep patterns

To characterize the extent to which a participant's sleep logs vary from the typical sleep pattern assumed by the calendar-relative isMainSleep algorithm, we apply the TSP algorithm and calculate the difference in the number of TSP- versus isMainSleep-classified primary sleep logs per night for each participant (**Equation 1**). Depending on how individual sleep logs are re-classified by TSP,

there may be an increase, decrease, or no change to the number of primary sleep logs ascribed to a given night.

Equation 1: Measure of variation from typical sleep pattern for a given participant:

$$\text{Variation from typical sleep pattern} = \frac{1}{n} \sum_{i=1}^n (SL_i^{TSP} - SL_i^{isMainSleep})_0$$

for the i -th day in a monitoring period of n total days given the number of sleep logs SL

For primary sleep periods that include two or more sleep logs, this measure increases as disjointed logs are “stitched” by TSP to obtain the full primary sleep period (i.e., increase from 1 to 2 on Day 1 of Figure 1). For primary sleep periods that include a isMainSleep-classified primary sleep log that falls outside of the participant’s typical sleep period, this measure decreases as these logs are “cleaned” by TSP and re-classified as naps (i.e., decrease from 1 to 0 on Day 0 of Figure 1). For primary sleep periods when neither “stitching” nor “cleaning” is applied, there is no change. Therefore, the net change in the mean number of primary sleep logs per night reflects the extent to which a participant’s primary sleep tends to be interrupted (net increase under TSP), tends to fall outside of their typical sleep period (net decrease under TSP), or is otherwise consistent with typical sleep patterns (no change under TSP).

Demographic and clinical characteristics:

Demographic characteristics such as age, sex at birth, race/ethnicity, and education were self-reported by participants by completing the Basics survey at the time of enrollment. EHR data from several healthcare provider organizations who are funded and part of *All of Us* Program was harmonized using the Observational Medical Outcomes Partnership common data model[18,19]. This harmonized EHR data was used to derive body mass index (BMI) and incident insomnia cases, which were quantified using International Classification of Diseases (ICD) billing codes mapped to phecodes, as described elsewhere[20,21]. To determine incident insomnia cases, any diagnosis before or during the first 6 months of Fitbit sleep monitoring were not counted as “incident cases” in this study.

Analysis:

Descriptive statistics for participants’ characteristics were presented by median and interquartile range for continuous variables, and frequency for categorical variables. Mann–Whitney U and chi-squared tests for continuous and categorical variables, respectively, were used to compare participants characteristics included versus excluded in the analytical dataset for this study. The isMainSleep and TSP algorithms were applied using sequence-level Fitbit data on Researcher Workbench using R programming language[22]. The number of primary versus non-primary sleep logs based on these two algorithms were evaluated using a 2X2 matrix table.

A data-driven quartile approach was taken to empirically examine the differences in variation from typical sleep patterns across clinical and demographic characteristics and sleep onset /offset patterns. Further, sleep metrics (listed in **Table 1**) were quantified using both algorithms across these quartiles. To account for repeated measures within individuals while accounting for between participant differences in the impact of the algorithm, mixed-effects linear regression models were used to evaluate the main (i.e. quartiles and algorithms) and interaction effects. Specifically, these models were used to further examine how sleep metrics, such as time attempting to sleep (TATS), TST, WASO, number of awakenings (NWAK), total sleep time in non-primary sleep period (TST-NP) and

counts of non-primary sleep, total wakefulness duration, and sleep efficiency percentage vary between the isMainSleep and TSP algorithms. Post-hoc pairwise comparisons were conducted using the Wilcoxon signed rank test. To test the stability of the finding, differences in sleep metrics using both algorithms were evaluated by looking at participants who had a higher proportion of “cleaned” versus “stitched” logs.

Given the longitudinal nature of Fitbit sleep data, we additionally explored whether our findings were sensitive to the amount of sleep data provided. We conducted sensitivity analyses by excluding participants who had fewer than 4 hours of sleep data on greater than 30% of days. These thresholds were informed by prior mathematical models as the theoretical cutoff at which insufficient sleep would be unsustainable[23].

Results

Of 413,457 total participants in v7 Controlled tier on Researcher workbench, 8,563 participants (72% female and 88% white) provided any sleep and EHR data and were included in the analytical sample (Table 2). The median Fitbit sleep monitoring duration was 4.2 years, which included sleep data over weekends.

Table 2: Participant characteristics for those included and excluded from the analytical cohort.

		Included N=8563	Excluded* N=404894	Test Statistic
Descriptive Statistics		Median (IQR) or % (n /N)		
Age (years)		58.1 (43.1 - 69.1)	57.1 (41.1 - 69.1)	F1 413455=15.6, P<0.0011
Sex at birth				$\chi^2_{21}=382$, P<0.0012
	Female	72 (6002 /8341)	61 (243563 /396393)	
	Male	28 (2339 / 8341)	39 (152830 /396393)	
Race				$\chi^2_{22}=1655$, P<0.0012
	Black	6 (454 /8008)	24 (78050 /328988)	
	Other	6 (486 /8008)	9 (28848 /328988)	
	White	88 (7068 /8008)	68 (222090 /328988)	
Education				$\chi^2_{22}=2779$, P<0.0012
	College	71 (5955 /8332)	45 (176390 /391595)	
	No college	6 (509 /8332)	29 (112937 /391595)	
	Some college	22 (1868 /8332)	26 (102268 /391595)	
BMI (kg/m ²)		28.4 (24.7 - 33.5)	28.3 (24.4 - 33.6)	F1 331488=6.35, P=0.0121
Fitbit monitoring duration (years)		4.2 (2.2 - 6.3)	4.0 (1.9 - 6.2)	F1 14906=11.6, P<0.0011
*participants were excluded because of not having EHR and Fitbit sleep data. Please note that total participants for demographic characteristics such as sex at birth, race, ethnicity, education are different from the primary analytical sample size because participants who have missing responses such as “none indicated”, “skip” have been removed.				

Impact of algorithms on sleep logs and participant characteristics

Almost 93% of sleep logs shown in categories A and D indicate 'No change' using either algorithm (**Table 3**), suggesting these logs were consistently classified as being primary vs. non-primary sleep by both algorithms. However, 5% (n= 401,777) of sleep logs in category B were classified as non-primary sleep under isMainSleep, but re-classified as primary sleep under TSP after “stitching”. More than 2% (n= 180,547) of sleep logs in category C classified as primary sleep under isMainSleep actually fall outside of the typical sleep period, and were re-classified as non-primary sleep via “cleaning”. The mean change in primary sleep logs per night after applying TSP, which measures variation from typical sleep patterns, ranged from -0.076 to 1 (**Supplemental Figure1**). Negative values represent a relatively larger net “cleaning” effect, as TSP reduced the mean number of primary sleep logs per night among participants whose non-primary sleep were labeled as primary sleep under isMainSleep. The values between 0 and 0.05, represent participants for whom TSP had relatively modest net impact because their sleep logs were consistently classified under both algorithms. The values >0.05 represent a relatively larger net “stitching” effect, as TSP increased the mean number of primary sleep logs per night among participants whose sleep tends to be interrupted and were thus incompletely captured under isMainSleep.

Table 3: 2x2 matrix table of primary sleep logs* as classified by calendar-relative and user-centric algorithms.

Number of sleep logs*		Calendar-relative (isMainSleep) algorithm	
		Primary Sleep “isMainSleep=true”	Non-Primary Sleep “isMainSleep=false”
User-centric (TSP) algorithm	Primary Sleep “TSP=true”	A (No change) 84.38% (N=7,132,635)	B (Stitching) 4.75% (N=401,777)
	Non-Primary Sleep “TSP=false”	C (Cleaning) 2.14% (N=180,547)	D (No change) 8.73% (N=737,671)
*Note: The sum of logs in each category exceeds the total sleep logs (N=8,435,172) due to 17,458 sleep logs reporting more than one value for isMainSleep. This discrepancy arises from the simplified method used to reconstruct sleep logs, as the log ID from the Fitbit Web API was not available in the dataset at the time of analysis on the Researcher Workbench.			

Table 4 shows participant characteristics organized by variation in typical sleep patterns. Q1 represents the lowest quartile of variation from typical sleep patterns and Q4 the highest quartile, in which the participants' sleep logs have a higher proportion of “cleaned” and “stitched” logs, respectively. Participants in Q4 are older (61.1 ± 14.7 vs. 54.5 ± 15.5 years) and have a higher incidence of insomnia (19% vs 10%) compared to those in Q1 (**Table 4**).

Table 4: Participants characteristics across quartiles (Q1-Q4) of variation from typical sleep patterns

Characteristics	Q1	Q2	Q3	Q4
-----------------	----	----	----	----

		(N=2140)	(N=2141)	(N=2141)	(N=2141)
Change in number of primary sleep logs (TSP - isMainSleep) per night		[-0.08,0.01]	[0.01,0.02]	[0.02,0.06]	[0.06,1]
Descriptive Statistics		Median (IQR) or % (n /N)			
Age (years)		56.2 (41.2 - 67.2)	55.2 (41.2 - 67.2)	58.2 (44.2 - 69.2)	64.2 (50.2 - 73.2)
Sex at birth					
	Female	72 (1491/2078)	73 (1530/2098)	74 (1540/2088)	69 (1441 /2077)
	Male	28 (587/2078)	27 (568/2098)	26 (548/2088)	31 (636/2077)
Race					
	Black	4 (71/1985)	4 (85/2021)	7 (132/2006)	8 (166/1996)
	Other	6 (114/1985)	7 (148/2021)	6 (120/2006)	5 (104/1996)
	White	91 (1800/1985)	88 (1788/2021)	87 (1754/2006)	86 (1726/1996)
Education					
	College	75 (1566/2077)	75 (1581/2095)	69 (1448/2085)	66 (1360/2075)
	No college	5 (107/2077)	5 (105/2095)	7 (136/2085)	8 (161/2075)
	Some college	19 (404/2077)	20 (409/2095)	24 (501/2085)	27 (554/2075)
BMI, kg/m²		27.2 (23.7 - 31.5)	27.7 (24.3 - 32.5)	29.0 (25.2 - 33.9)	30.1 (26.0 - 35.7)
Fitbit monitoring duration (years)		3.4 (1.6 - 5.5)	4.7 (2.6 - 6.49)	4.7 (2.5 - 6.5)	4.0 (2.0 - 6.3)
Incident Insomnia diagnosis		10 (221/2107)	10 (220/2105)	14 (305/2104)	19 (389/2083)

Sleep metrics

Upon visual inspection of the heatmaps for sleep onset /offset (**Figure 4A-B**), the TSP algorithm effectively recategorizes all isMainSleep-classified primary sleep logs that fall outside of 12:00 AM to 12:00 PM across all quartiles, as noted by the dramatic reductions in short sleep logs occurring around midday (ie, likely non-primary sleep mislabeled by isMainSleep as primary sleep). Further, we observe considerable differences in median sleep onset /offset under both algorithms and greater variability in sleep onset and offset times among participants with greater variation from typical sleep patterns. Among Q4 participants, sleep onset and offset were shifted by 6 and 37 minutes, respectively, whereas the sleep onset and offset were shifted by 5 and 15 minutes respectively among Q1 participants.

Figure 4A-B: Sleep onset /offset by quartiles of variation from typical sleep patterns using calendar-relative (A) and user-centric (B) algorithms.

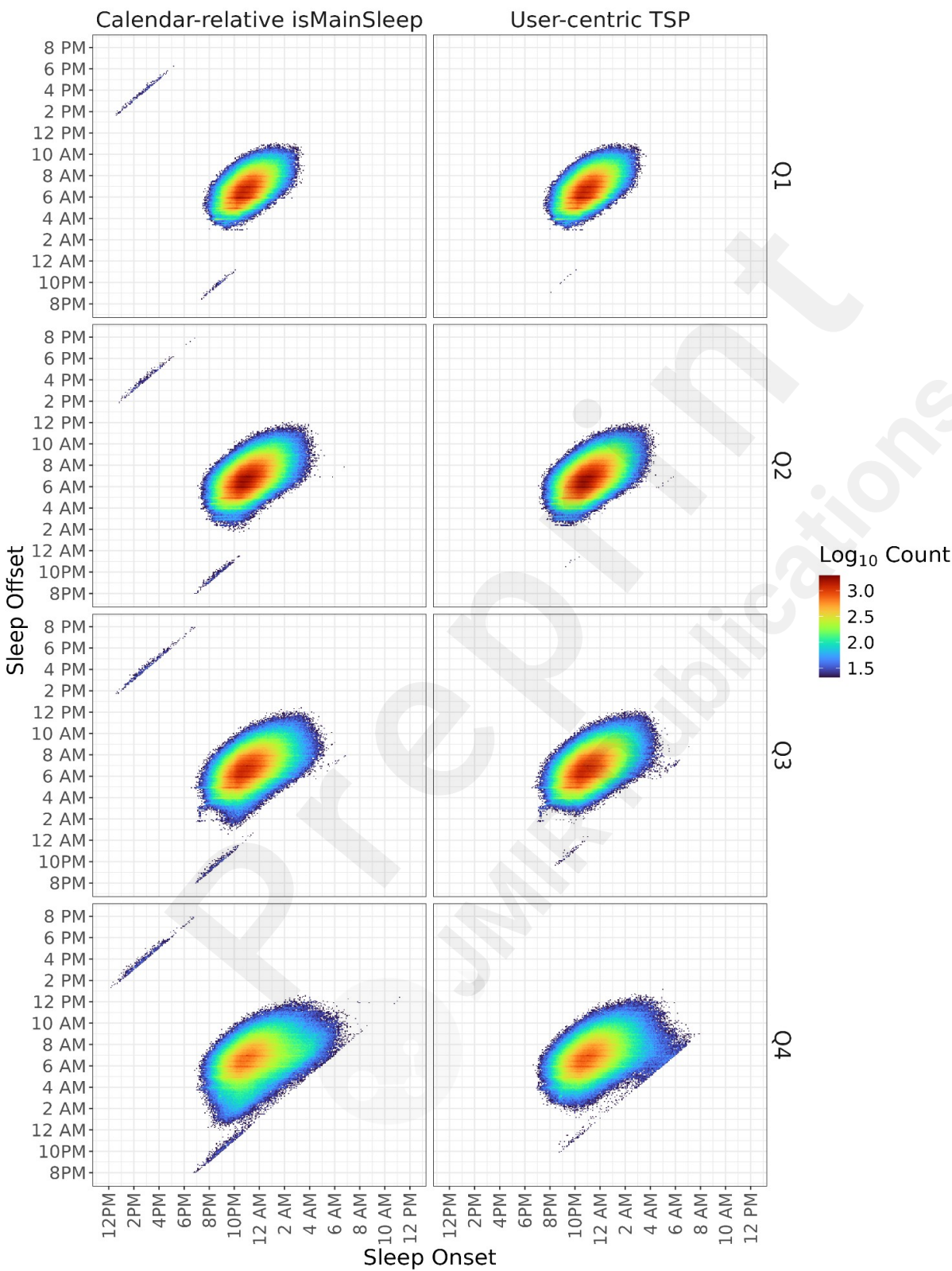
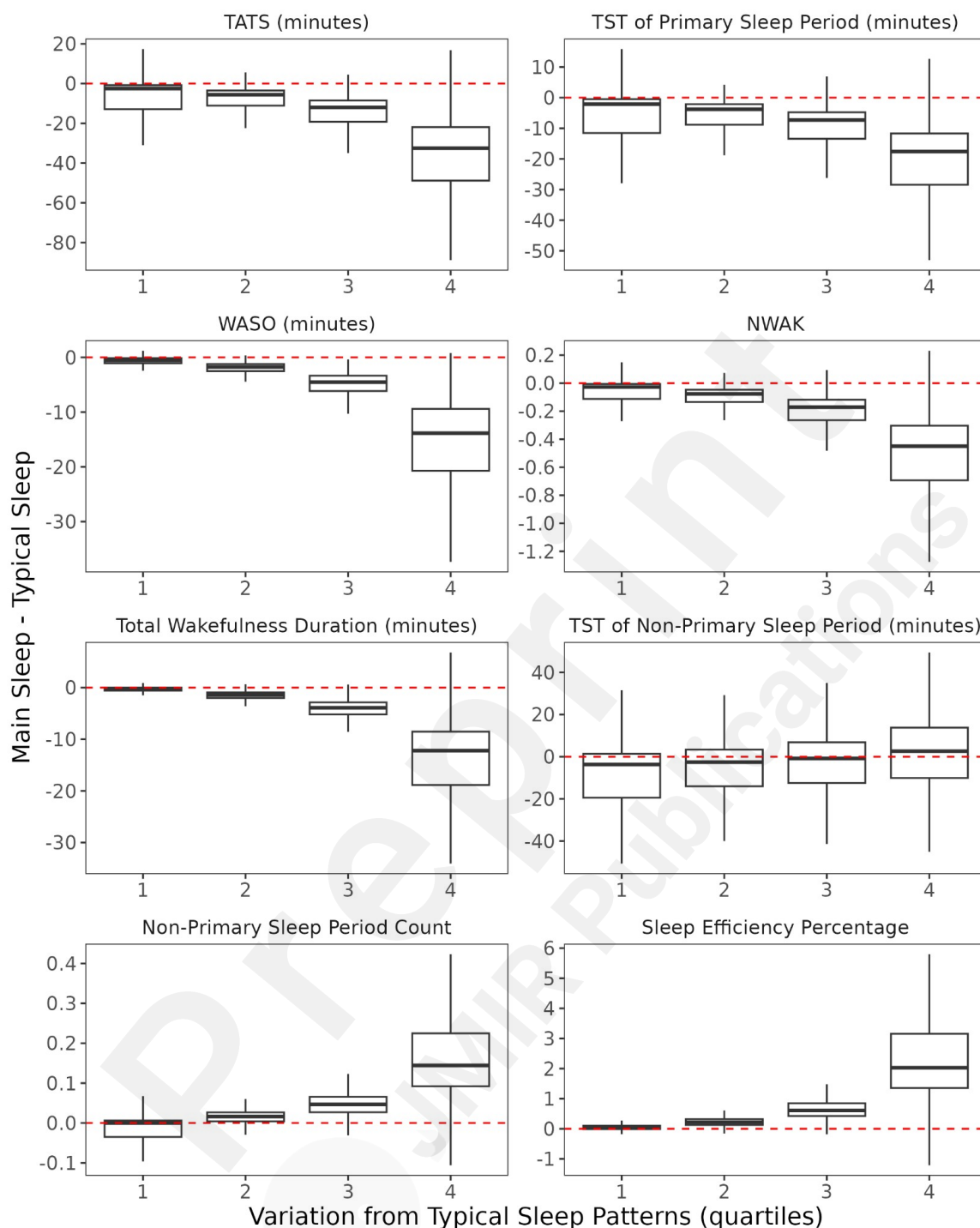


Figure 5 and Supplemental Table 1 shows significant interactions ($P<.001$) between the algorithm applied and all key sleep metrics across quartiles. For participants in Q4, the median (IQR) TATS

was 404.72 (350.4, 447.91) under isMainSleep, compared to 442.03 (399.17, 482.93) under TSP (Supplemental Table 1). This suggests that median TATS computed using isMainSleep is shorter by 32.5 minutes ([21.9, 48.8], $P < 0.001$) compared to TSP. Similarly, lower values were observed for TST by 17.6 median minutes ([11.7, 28.4], $P < 0.001$), WASO by 13.9 median minutes ([9.4, 20.7], $P < 0.001$), NWAK by 0.5 median counts ([0.3, 0.7], $P < 0.001$) and total wakefulness duration by 12.2 median minutes ([8.5, 18.9], $P < 0.001$) when computed using isMainSleep compared to TSP (Figure 5). On the other hand, among participants in Q4, isMainSleep-derived median non-primary sleep counts [0.31 (0.22, 0.46)] and sleep efficiency percentage [91.27 (89.66, 92.67)] were higher by 0.14 ([0.09, 0.22]), $P < 0.001$ and 2.0% ([1.35, 3.2]), $P < 0.001$, when compared to TSP-derived non-primary sleep counts [0.15 (0.07, 0.27)] and sleep efficiency percentage [88.96 (86.71, 90.51)], respectively. We observe similar trends for these metrics comparison between algorithms across other quartiles (Q1-Q3), however, the magnitude of differences were attenuated relative to Q4 (Supplemental Table 1).

Figure 5: Boxplots of paired differences between user-centric (TSP) and calendar-relative (isMainSleep) algorithms for each of the key hypothesized sleep metrics across the quartiles of variation from typical sleep patterns.



Box bounds, midline, and whiskers represent the IQR, median, and $1.5 \times \text{IQR}$, respectively. Values above the dotted red line indicate a higher value from the isMainSleep algorithm relative to the TSP algorithm; whereas, values below the red line indicate a lower value.

Differences in other sleep metrics shown in Table 1 are provided in Supplemental Table 1. Specifically, although there are statistically significant differences in sleep stage metrics (light, deep, REM) between the two algorithms, they are modest (<5 minutes). Similar findings in sleep metrics were observed when examining the differences using 3-level categories of variation from typical sleep patterns (Supplemental Table 2) and when repeating these analyses by excluding participants who had fewer than 4 hours of sleep data on greater than 30% of days (results not shown).

Discussion

Principal findings

We developed and implemented a user-centric algorithm that more appropriately classifies primary sleep logs based on users' habitual sleep patterns using the All of Us Fitbit sequence level sleep data. This TSP algorithm re-classified 5% of sleep logs previously classified as non-primary sleep as primary sleep, thus more accurately capturing all short sleep segments within the users' primary sleep period, particularly among individuals with more disrupted sleep patterns. The default calendar-relative isMainSleep algorithm resulted in lower values for TATS, TST, WASO, total wakefulness duration, and NWAK, and higher values for counts of non-primary sleep and sleep efficiency percentage compared to the user-centric TSP algorithm. Comparatively, the algorithm had an expectedly modest impact on sleep stage and sleep pattern metrics. These findings remained consistent across the empirically established quartiles of variation from typical sleep patterns and within the subgroup of participants who contributed above a minimum threshold of sleep data.

Interpretation of findings

While the algorithm did not affect 93% of sleep logs, we found that participants who had a higher preponderance of "stitched" logs, i.e., evidence suggesting disrupted sleep, also had higher incidence of insomnia. These findings suggest that while the two algorithms may perform similarly well in the general healthy population, mis-estimation of sleep metrics by the isMainSleep algorithm may be particularly pronounced among participants with more disrupted sleep patterns and physiology such as insomnia. The American Academy of Sleep Medicine estimates that 30% of adults have symptoms of insomnia[24]. Our prior study[3] underscored the potential value of long-term monitoring of sleep stages, duration, and regularity in identifying chronic disease risk, which aligns with broader research on the limitations of traditional sleep monitoring methods (e.g. 7-to-14-day actigraphic assessments requiring sleep diaries and manual annotations) and the benefits of advanced technologies to passively and reliably track sleep-wake data[25,26]. Given that there is a growing body of evidence supporting the notion that sleep regularity may be an even more important predictor of health outcomes and duration[27], getting a more accurate picture of the day over day patterns - beyond just a week or two - of sleep-wake behaviors and sleep disturbances is essential to drawing correct inference about the bidirectional relationship between sleep and health. Our method is particularly sensitive to correctly accounting for the most disturbed sleep patterns that tend to cause failures of algorithms derived from generally healthy sleepers.

The user-centric algorithm takes a significant step to advance epidemiologic and sleep-disorder specific sleep research relying upon Fitbit sleep data, by aligning the identification of participants' primary sleep periods with homeostatic and physiological sleep patterns, rather than arbitrary calendar times. For instance, through 'stitching' and 'cleaning' of sleep logs, the TSP algorithm does not meaningfully impact sleep staging-based metrics, but it impacts the calculation of sleep-wake metrics such as BT, WT, TST, TATS, WASO, counts of non-primary sleep, sleep efficiency percentage, etc. For context, the average difference in these metrics in the most-affected population constitutes about half of the minimum clinically important difference (MCID) suggested by studies assessing insomnia treatment response [28,29]. Consequently, this method enhances the precision of these day-over-day behavioral sleep metrics and provides more relevant insights into the behavioral and physiological factors influencing sleep health for most people some of the time, and for some people (likely those with chronic insomnia) much of the time. This refinement in pre-processing

sleep data from wearables is critical for advancing our understanding of the complex interplay between longitudinal measurements of sleep-wake patterns and various health conditions[27,30].

While found to be statistically significant, durations in staged sleep had only modest differences between two algorithms, which are unlikely to be clinically meaningful. This is likely because most sleep logs that are re-classified (from non-primary to primary sleep period logs) and “stitched” by the TSP algorithm are less than 3 hours and, therefore, do not have sleep staging information. As the TSP algorithm is focused on refining measures related to sleep-wake proportionality using existing sleep logs via data modeling, it does not address limitations that underlie the data derived from the Fitbit sleep staging algorithm. For example, sleep logs can consist of 1 or more sleep segments without sleep stages are “stitched” with staged sleep segments by TSP to form a primary sleep period. TSP does not update reported sleep stages for the entire sleep period since the raw signal data used to estimate sleep stages are not available. However, this only occurs relatively infrequently, affecting <0.2% (N=15,029) of the TSP-classified primary sleep logs (N=7,518,329) where both staged and unstaged sleep co-exist. To further enhance the value of the TSP algorithm, we provide a flag to identify all TSP hybrid stitched sleep logs (where staged and unstaged sleep segments can co-exist), offering a simple method for filtering out nights with potentially underestimated sleep architecture (i.e., those nights with staged and unstaged segments). This flag is important for sleep stage-based research because it identifies sleep logs that are fully captured and staged by the default isMainSleep algorithm.

We found clinically meaningful differences for metrics related to sleep disturbances and duration, i.e., the TSP algorithm achieved the intended result of addressing issues with sleep-wake proportionality, particularly among those individuals with the most disrupted sleep patterns. Therefore, relying solely on the isMainSleep algorithm to identify primary sleep logs could cause mis-aggregation of sleep logs and inaccuracy in calculating sleep and wake durations as well as day-over-day bed /wake time metrics. Both metrics have significant health implications[27,30]. As shown in our conceptual and data examples, the isMainSleep algorithm is more likely to mis-aggregate the sleep to the wrong day, and also mis-characterize primary sleep as a non-primary sleep (and vice versa), particularly when sleep is broken on either side of midnight. While the analysis and comparison conducted in this study were data-driven, future studies are needed comparing the performance of the TSP algorithm against polysomnography and/or actigraphy with sleep diaries, including participants with atypical sleep patterns or sleep disorders.

This study provides a practical guide for utilizing the user-centric sleep algorithm on a cloud-based platform and offers a user-friendly R package to streamline its implementation. Prior technical automation codes required self-reported sleep logs from participants[13], whereas R package developed in this study requires Fitbit sleep data from wearables only. While the validity of passive detection and reassignment of logs to the primary sleep period was explored in this manuscript, we hope that the open-source R package allows researchers to easily study whether the reassigned logs adequately capture sleep patterns that they are investigating, thereby promoting transparency and reproducibility, and making Fitbit Web API data more accessible for clinical and translational research. These contributions align with open science principles,[31,32] encouraging further refinement of sleep analysis methods and advancing sleep research. The algorithm’s applicability extends to all Fitbit devices, with potential to generalize the logic related to sleep-wake pattern identification to other data sources, such as manually annotated actigraphs and sleep data from other consumer wearable devices, with appropriate adaptation of the algorithm’s concept to the unique data types/representations from different sources. While external validation is needed, this open-source R package equips researchers with a tool informed by Fitbit developers and sleep medicine experts, fostering improved data interpretation and methodological advancements by addressing known

limitations in the current algorithms.

Limitations

While we have demonstrated the key advantages of the user-centric TSP algorithm in accounting for habitual sleep patterns, there remain a few important caveats. First, the absence of sleep data from gold standard measures like polysomnography prevents a direct evaluation of algorithm performance. Future research could address this by incorporating polysomnography or other validated sleep assessments to better validate the algorithms. Second, the user-centric TSP algorithm assumes that all periods of wakefulness imputed between “stitched” primary sleep logs are part of the primary sleep period. We are not able to distinguish between cases when the user was intentionally awake and engaged in purposeful activity or when the user was attempting to sleep during this imputed wake period. Therefore, TSP may introduce variations in activity levels due to differences in sleep and wake period breakdowns. Research studies designed to examine the 24-hour ActivityCycle model[33], i.e., inter-relatedness of different activity and sleep behaviors may need to perform sensitivity analyses to account for these variations and validate findings under diverse conditions. Third, TSP does not address inherent limitations to wearable data collection that may introduce errors in sleep logs, including issues such as low battery power, device removal during sleep, or poor heart rate variability data quality that prevents sleep staging. Given that the median monitoring time of Fitbit sleep data used in this study was 4.2 years, data outliers on a small proportion of nights are less likely to influence the overall understanding of habitual sleep behaviors over time. Lastly, the generalizability of our results may be limited by the inclusion of atypical sleep schedules, such as polyphasic sleep or rotating shift work, which our methods cannot adequately characterize. However, prior study leveraging All of Us Fitbit sleep data showed that <7% of weekdays had a sleep onset outside of traditional sleep hours of 8:00 p.m. to 2:00 a.m., suggesting that the probability of atypical sleep schedules is low in this cohort[3]. Future studies could explore alternative algorithms or additional data sources to better understand and address the effects of these variations from typical sleep patterns. That being said, we do want to highlight that this new algorithm’s approach to identifying the primary sleep period of the individual, based on their typical sleeping pattern rather than by the longest sleep log that ends on a given calendar day, is more likely to accurately capture and aggregate sleep logs that reflect an individual’s primary sleep period, even if it occurs at an atypical time of day. This requires that there is at least some regularity to that sleep schedule to be able to (1) leverage to longitudinal data in identifying what is “usual” for that individual and (2) identify logs that are sufficiently overlapping the identified TSP/primary sleep period window to be labelled accordingly. In the latter case, only the most extremely varying schedules (which would be very unhealthy and physiologically difficult to maintain for extended periods of time, from a circadian rhythm standpoint) are likely to have primary sleep sessions that have no overlap - thus not being able to be assigned to the TSP - with the “usual” primary sleep period of the individual. There have been active efforts by the *All of Us* Research Program to address diversity gaps in the Fitbit data collection through the Wearables Enhancing All of Us Research (WEAR) study, which provides devices to participants from underrepresented communities at no cost. This might provide future research opportunities to re-evaluate this algorithm in a socio-demographically diverse population.

Conclusion

In summary, the user-centric algorithm represents a significant advancement in the analysis of sleep-wake patterns by better aligning with individuals' habitual sleep behaviors, rather than relying on fixed calendar times. This algorithm captures the natural variability in sleep schedules, offering a more accurate tool for evaluating sleep metrics related to schedule, duration and disturbances. The

robustness of the user-centric algorithm enhances its utility in both clinical and translational research, where individualized and precise sleep assessments are essential. To support its widespread adoption, we have developed a comprehensive and user-friendly R package, enabling seamless integration of the TSP algorithm into existing research workflows. This tool holds considerable potential for advancing personalized sleep medicine and deepening our understanding of sleep-related health outcomes.



Abbreviations:

EHR: Electronic health records

WASO: Wake after sleep onset

TST: Total sleep time

TSP: Typical sleep period

API: Application Programming Interface

BMI: Body mass index

ICD: International Classification of Diseases

TATS: Time attempting to sleep

NWAK: Number of awakenings

TST-NP: Total sleep time in non-primary sleep period

BT: Bedtime

WT: Wake time

MSP: Midsleep point

Funding

This material is based on work that is partially funded by an unrestricted gift from Google. The study was supported by National Institutes of Health (NIH) grant nos. R21 HL172038 (E.L.B., D.M.R.), R61 /R33 HL158941 (E.L.B.) and R01 FD007627 (E.L.B.).

Data availability

To ensure privacy of participants, All of Us data used for this study are available to approved researchers following registration, completion of ethics training and attestation of a data use agreement through the All of Us Research Workbench platform, which can be accessed via <https://workbench.researchallofus.org/>. However, the real-world data from one of the co-authors (L.S.) used in the example shown in Supplemental Figure 1 is made available on the GitHub, link here: <https://github.com/annisjs/typical.sleep>.

Code availability

Code used for this study can be made immediately available to any approved researchers on the All of Us Research Workbench platform by contacting our study team. Additionally, the R package and the code with results shown in illustrative example (Figure 2) have also been made available on

GitHub, link here: <https://github.com/annisjs/typical.sleep>.

Acknowledgements

We thank *All of Us* participants for their contributions, without whom this research would not have been possible. We also thank the NIH's All of Us Research Program for making available the cohort examined in this study. The All of Us Research Program is supported by the NIH, Office of the Director: Regional Medical Centers (1 OT2 OD026549, 1 OT2 OD026554, 1 OT2 OD026557, 1 OT2 OD026556, 1 OT2 OD026550, 1 OT2 OD 026552, 1 OT2 OD026553, 1 OT2 OD026548, 1 OT2 OD026551, 1 OT2 OD026555; IAA: AOD21037, AOD22003, AOD16037, AOD21041); Federally Qualified Health Centers (HHSN 263201600085U); Data and Research Center (5 U2C OD023196); Biobank (1 U24 OD023121); The Participant Center (U24 OD023176); Participant Technology Systems Center (1 U24 OD023163); Communications and Engagement (3 OT2 OD023205, 3 OT2 OD023206); and Community Partners (1 OT2 OD025277, 3 OT2 OD025315, 1 OT2 OD025337, 1 OT2 OD025276).

Funding

This material is based on work that is partially funded by an unrestricted gift from Google. The study was supported by National Institutes of Health (NIH) grant nos. R21 HL172038 (E.L.B., D.M.R.), R61 /R33 HL158941 (E.L.B.) and R01 FD007627 (E.L.B.).

Competing Interests

D.M.R. has served on advisory boards for Illumina and Alkermes and has received research funds unrelated to this work from PTC Therapeutics. E.L.B. has received research funds unrelated to this work from United Therapeutics. J.H.C., K.G., J.H. and L.D.S. are employees of Google and own Alphabet stock as part of the standard compensation package. L.D.S. has been compensated for participation in speakers' bureaus and advisory boards by Eisai Pharmaceuticals, Jazz Pharmaceuticals, Avadel Pharmaceuticals and Harmony Biosciences, unrelated to this work. J.H. serves on the board of directors for ResMed and owns ResMed stock. Fitbit and Google were not involved in the collection, management and analysis of the data, nor in the decision to submit the manuscript for publication. Team members from Fitbit and Google provided input during the design and conduct of the study, code review using the data shown in illustrative example, participated in interpretation of results, and participated in preparation, review and approval of the manuscript. The All of Us Research Program was not involved in the design and conduct of the study; collection, management and analysis of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication. To ensure privacy of participants, data used for this study were accessed and available to approved researchers only following registration, completion of ethics training and attestation of a data use agreement through the All of Us Research Workbench platform, which can be accessed via <https://workbench.researchallofus.org/login>. The other authors declare no competing interests.

Inclusion & Ethics Statement: All collaborators who have contributed to study design, study implementation, data ownership, intellectual property and authorship of publications have been included as authors. The roles and responsibilities were agreed amongst collaborators ahead of the research. The research is locally relevant and has this been determined in collaboration with local partners. This research was not severely restricted or prohibited in the setting of the researchers. The findings of this research did not result in stigmatization, incrimination, discrimination or otherwise

personal risk to participants Local and regional research relevant to our study was taken into account in citations.

Author Contribution: Drs. Master, Annis and Brittain had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: Master, Annis, Ching, Hernandez, Schneider and Brittain. Acquisition, analysis, or interpretation of data: Master, Annis, Ching, Hernandez, Schneider and Brittain. Drafting of the manuscript: Master, Annis, Ching, Hernandez, Schneider and Brittain. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: Master, Annis, Ching, Hernandez, Schneider and Brittain. Obtained funding: Brittain. Administrative, technical, or material support: Master, Annis and Brittain. Supervision: Schneider, Brittain.

References:

1. Master H, Kouame A, Marginean K, Basford M, Harris P, Holko M. How Fitbit data are being made available to registered researchers in All of Us Research Program. *Pac Symp Biocomput.* 2023;28: 19–30. Available: <https://www.ncbi.nlm.nih.gov/pubmed/36540961>
2. Investigators AoURP. The “All of Us” Research Program. *N Engl J Med.* 2019;381: 668–676. doi:10.1056/NEJMSr1809937
3. Zheng NS, Annis J, Master H, Han L, Gleichauf K, Ching JH, et al. Sleep patterns and risk of chronic disease as measured by long-term monitoring with commercial wearable devices in the All of Us Research Program. *Nat Med.* 2024. doi:10.1038/s41591-024-03155-8
4. Master H, Annis J, Huang S, Beckman JA, Ratsimbazafy F, Marginean K, et al. Association of step counts over time with the risk of chronic disease in the All of Us Research Program. *Nat Med.* 2022;28: 2301–2308. doi:10.1038/s41591-022-02012-w
5. Consumer Technology Association. Physical activity monitoring for step counting (ANSI/CTA-2056-A). 2021. Available: <https://shop.cta.tech/products/physical-activity-monitoring-for-step-counting-ansi-cta-2056-a>
6. The Digital Health Measurement Collaborative Community (DATAcc) by the Digital Medicine Society (DiMe). Core Measures: Sleep Measurement System. In: DATAcc by DiMe [Internet]. 12 Apr 2024 [cited 24 Jul 2024]. Available: <https://datacc.dimesociety.org/core-measures-sleep/all-core-measures/>
7. Consumer Technology Association. Physical activity monitoring for heart rate - real world analysis (ANSI/CTA-2065.1). 2023. Available: <https://shop.cta.tech/products/physical-activity-monitoring-for-heart-rate>
8. Consumer Technology Association. Definitions and characteristics for wearable sleep monitors (ANSI/CTA/NSF-2052.1-A). 2022. Available: <https://shop.cta.tech/products/definitions-and-characteristics-for-wearable-sleep-monitors-ansi-cta-nsf-2052-1-a>
9. Lederer L, Breton A, Jeong H, Master H, Roghanizad AR, Dunn J. The Importance of Data Quality Control in Using Fitbit Device Data From the Research Program. *JMIR Mhealth Uhealth.* 2023;11: e45103. doi:10.2196/45103

10. Chan A, Chan D, Lee H, Ng CC, Yeo AHL. Reporting adherence, validity and physical activity measures of wearable activity trackers in medical research: A systematic review. *Int J Med Inform.* 2022;160: 104696. doi:10.1016/j.ijmedinf.2022.104696
11. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of Wristband Fitbit Models in Assessing Sleep: Systematic Review and Meta-Analysis. *J Med Internet Res.* 2019;21: e16273. doi:10.2196/16273
12. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Performance assessment of new-generation Fitbit technology in deriving sleep parameters and stages. *Chronobiol Int.* 2020;37: 47–59. doi:10.1080/07420528.2019.1682006
13. Burgess HJ, Rodgers AA, Rizvydeen M, Mongefranco G, Fayyaz Z, Fejer A, et al. Lessons learned on the road to improve sleep data extracted from a Fitbit device. *Sleep.* 2025 [cited 24 Jan 2025]. doi:10.1093/sleep/zsae290
14. Ong JL, Golkashani HA, Ghorbani S, Wong KF, Chee NI, Willoughby AR, et al. A data driven approach for choosing a wearable sleep tracker. *bioRxiv.* 2023. doi:10.1101/2023.10.12.23296981
15. Khosla S, Deak MC, Gault D, Goldstein CA, Hwang D, Kwon Y, et al. Consumer Sleep Technology: An American Academy of Sleep Medicine Position Statement. *J Clin Sleep Med.* 2018;14: 877–880. doi:10.5664/jcsm.7128
16. Get sleep log by date. [cited 22 Jul 2024]. Available: <https://dev.fitbit.com/build/reference/web-api/sleep/get-sleep-log-by-date/>
17. Mayo KR, Basford MA, Carroll RJ, Dillon M, Fullen H, Leung J, et al. The All of Us Data and Research Center: Creating a Secure, Scalable, and Sustainable Ecosystem for Biomedical Research. *Annu Rev Biomed Data Sci.* 2023;6: 443–464. doi:10.1146/annurev-biodatasci-122120-104825
18. Klann JG, Joss MAH, Embree K, Murphy SN. Data model harmonization for the All Of Us Research Program: Transforming i2b2 data into the OMOP common data model. *PLoS One.* 2019;14: e0212463. doi:10.1371/journal.pone.0212463
19. Ramirez AH, Sulieman L, Schlueter DJ, Halvorson A, Qian J, Ratsimbazafy F, et al. The All of Us Research Program: Data quality, utility, and diversity. *Patterns (N Y).* 2022;3: 100570. doi:10.1016/j.patter.2022.100570
20. Zheng NS, Feng Q, Kerchberger VE, Zhao J, Edwards TL, Cox NJ, et al. PheMap: a multi-resource knowledge base for high-throughput phenotyping within electronic health records. *J Am Med Inform Assoc.* 2020;27: 1675–1687. doi:10.1093/jamia/ocaa104
21. Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Inform.* 2019;7: e14325. doi:10.2196/14325
22. Website. Available: R Core Team (2021). R: A language and environment for statistical ## computing. R Foundation for Statistical Computing, Vienna, Austria. ## URL <https://www.R-project.org/>

23. McCauley P, Kalachev LV, Smith AD, Belenky G, Dinges DF, Van Dongen HPA. A new mathematical model for the homeostatic effects of sleep loss on neurobehavioral performance. *J Theor Biol.* 2009;256: 227–239. doi:10.1016/j.jtbi.2008.09.012
24. American Academy of Sleep Medicine (AASM). Insomnia. 2008 [cited 25 Sep 2024]. Available: <https://aasm.org/resources/factsheets/insomnia.pdf>
25. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable Sleep Technology in Clinical and Research Settings. *Med Sci Sports Exerc.* 2019;51: 1538–1557. doi:10.1249/MSS.0000000000001947
26. Robbins R, Seixas A, Masters LW, Chanko N, Diaby F, Vieira D, et al. Sleep tracking: A systematic review of the research using commercially available technology. *Curr Sleep Med Rep.* 2019;5: 156–163. doi:10.1007/s40675-019-00150-1
27. Windred DP, Burns AC, Lane JM, Saxena R, Rutter MK, Cain SW, et al. Sleep regularity is a stronger predictor of mortality risk than sleep duration: A prospective cohort study. *Sleep.* 2024;47. doi:10.1093/sleep/zsad253
28. Papaconstantinou E, Cancelliere C, Verville L, Wong JJ, Connell G, Yu H, et al. Effectiveness of non-pharmacological interventions on sleep characteristics among adults with musculoskeletal pain and a comorbid sleep problem: a systematic review. *Chiropr Man Therap.* 2021;29: 23. doi:10.1186/s12998-021-00381-6
29. Wang L, Pan Y, Ye C, Guo L, Luo S, Dai S, et al. A network meta-analysis of the long- and short-term efficacy of sleep medicines in adults and older adults. *Neurosci Biobehav Rev.* 2021;131: 489–496. doi:10.1016/j.neubiorev.2021.09.035
30. Saint-Maurice PF, Freeman JR, Russ D, Almeida JS, Shams-White MM, Patel S, et al. Associations between actigraphy-measured sleep duration, continuity, and timing with mortality in the UK Biobank. *Sleep.* 2024;47. doi:10.1093/sleep/zsad312
31. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, du Sert NP, et al. A manifesto for reproducible science. *Nat Hum Behav.* 2017;1: 0021. doi:10.1038/s41562-016-0021
32. Mazzotti DR, Haendel MA, McMurry JA, Smith CJ, Buysse DJ, Roenneberg T, et al. Sleep and circadian informatics data harmonization: a workshop report from the Sleep Research Society and Sleep Research Network. *Sleep.* 2022;45: zsac002. doi:10.1093/sleep/zsac002
33. Rosenberger ME, Fulton JE, Buman MP, Troiano RP, Grandner MA, Buchner DM, et al. The 24-Hour Activity Cycle: A New Paradigm for Physical Activity. *Med Sci Sports Exerc.* 2019;51: 454–464. doi:10.1249/MSS.0000000000001811
34. Consensus Conference Panel, Watson NF, Safwan Badr M, Belenky G, Bliwise DL, Buxton OM, et al. Recommended Amount of Sleep for a Healthy Adult: A Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society. *J Clin Sleep Med.* 2015 [cited 25 Sep 2024]. doi:10.5664/jcsm.4758
35. Hirshkowitz M, Whiton K, Albert SM, Alessi C, Bruni O, DonCarlos L, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. *Sleep Health.* 2015;1: 40–43. doi:10.1016/j.sleh.2014.12.010

Methods Supplement:

Technical implementation of the TSP algorithm:

The implementation steps outlined below are applicable to sleep data captured from all Fitbit devices.

Step 1: Construct Fitbit sleep logs from sequence-level sleep data

Currently, the Fitbit sequence level sleep data available on All of Us Researcher Workbench only includes the start date time and duration for each sleep segment. Therefore, to implement the user-centric TSP algorithm on All of Us Researcher Workbench, we first apply a stitching rule to the sequence level sleep data to reconstruct Fitbit sleep logs by aggregating consecutive sleep segments that are separated by less than one hour. While this simplified approach may not perfectly replicate the structure of algorithmically-constructed Fitbit sleep logs, the discrepancy is likely minimal as our post hoc QA (quality assurance) process identified data inconsistencies in only <0.2% of sleep logs (e.g. sleep logs reporting more than 1 value for “isMainSleep”). This step of re-constructing sleep logs can be omitted, however, if the sleep data is accessed directly from the Fitbit Web API and includes the logID data element. In this case, sleep logs are simply those sequence level data that share a common logID.

Step 2: Identify the typical sleep period

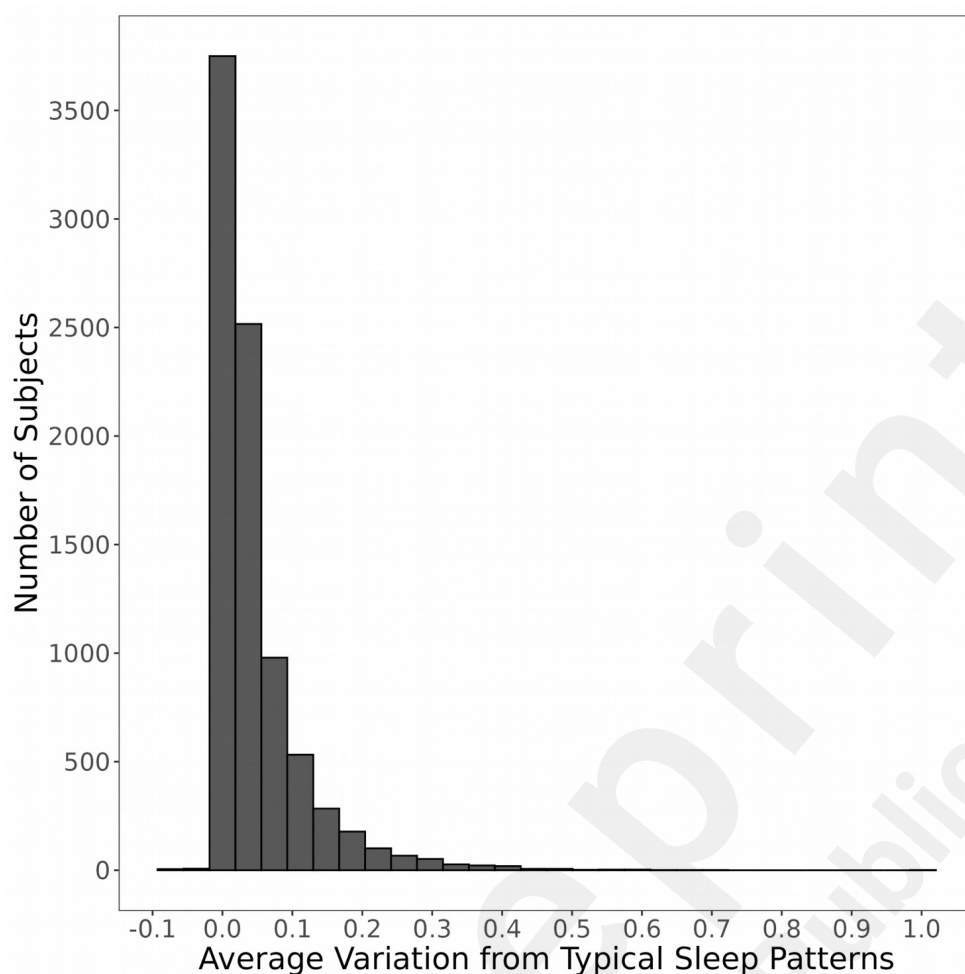
Next, we calculate the mid-sleep point (MSP) as the midpoint time between sleep onset and sleep offset for each isMainSleep-classified primary sleep log. We determine the median MSP across all available “isMainSleep=true” logs for each user, and also establish a time range of median mid-sleep point + /- 8 hours as the overlap period over which sleep logs are deemed relevant for establishing the user’s typical sleep period. Lastly, each user’s TSP bedtime and TSP wake time are calculated as the median bedtime and wake time across all relevant sleep logs, with bedtimes derived from the start of the first relevant sleep log in each sleep period and wake times derived from the end of the last relevant sleep log in each sleep period.

Step 3: Classify sleep logs based on overlap with the typical sleep period

All sleep logs that overlap with the typical sleep period, i.e. with bedtime (BT) and wake time (WT) between the TSP bedtime and wake time as defined above, are classified as primary sleep (“TSP=true”). All other sleep logs that do not meet the above criteria are classified as non-primary sleep (“TSP=false”), as these are most likely non-primary sleep outside of the typical sleep period.

Given how inconsistent sleep schedules may lead to incorrect BT /WT estimates, if the time between the TSP BT and WT is <5 hrs (i.e., indicative of habitually unhealthy short sleep durations), then the TSP BT and WT are set as the 7-hour window of sleep opportunity around the median MSP, based on the minimum recommended sleep opportunity for adults[34,35] (e.g., BT=MSP-3.5hrs and WT=MSP+3.5hrs). Though we acknowledge this as a simplifying assumption, the prevalence of this edge case was relatively low (<1% of sleep logs, n = 69,067), and contributed by <5% of the total analytical sample (n= 417). In addition to BT and WT, these TSP-classified primary sleep logs can now be used to calculate other metrics related to a user’s sleep schedule, sleep duration, and sleep disturbances.

Supplemental Figure 1: Distribution of change in the mean number of primary sleep logs per night as a measure of variation from typical sleep patterns.



Negative values represent participants whose non-primary sleep were mislabeled as primary sleep under isMainSleep. Values between 0 and 0.05 represent participants whose sleep logs were consistently classified under both algorithms. Values >0.05 represent participants whose sleep tends to be interrupted, and were previously incompletely captured under isMainSleep.

Supplemental Tables

Supplemental Table 1: Descriptive stats of sleep metrics by calendar-relative (isMainSleep) and user-centric (TSP) algorithms by quartiles for variation from typical sleep patterns								
Variable	Algorithm	Q1	Q2	Q3	Q4	P-value (A^)	P-value (B^^)	P-value (AxB^^)
Mean # logs / day (Range)	NA	[-0.0761,0.00735]	[0.00735,0.0235]	[0.0235,0.0594]	[0.0594,1]			
Number of participants	NA	2140	2141	2141	2141			
Number of logs	isMainSleep	1428784	2122928	2042406	1711392	NA	NA	NA
	TSP	1408141	2118034	2080945	1902153			
Person-days	isMainSleep	1427450	2132910	2063404	1756258	NA	NA	NA
	TSP	1420310	2123700	2051308	1738234			
Sleep schedule								
Bedtime (BT)	isMainSleep	22:50:12 [22:01:54,23:36:00]	23:06:18 [22:27:12,23:51:54]	23:40:36 [22:45:24,00:46:12]	23:13:36 [22:27:12,00:04:06]	<0.0001	0.9465	<0.0001
	TSP	22:54:06	23:09:00	23:33:48	23:15:12			

		[22:08:24,23:39:54]	[22:30:00,23:55:48]	[22:38:06,00:42:30]	[22:28:24,00:06:36]			
Sleep onset	isMainSleep	22:55:54 [22:07:24,23:41:12]	23:10:54 [22:32:12,23:56:54]	23:43:06 [22:49:36,00:49:00]	23:18:24 [22:31:18,00:08:30]	<0.0001	<0.0001	<0.0001
	TSP	23:00:06 [22:14:54,23:45:48]	23:13:36 [22:35:24,00:00:42]	23:37:42 [22:41:12,00:46:18]	23:19:24 [22:32:54,00:11:06]			
Sleep offset	isMainSleep	06:16:30 [04:46:00,06:59:24]	06:28:36 [05:47:00,07:06:48]	05:56:18 [04:48:00,06:45:36]	06:17:54 [05:23:12,06:59:30]	<0.0001	<0.0001	<0.0001
	TSP	06:31:36 [05:41:48,07:12:36]	06:44:30 [06:08:24,07:19:42]	06:33:42 [05:42:36,07:20:24]	06:39:18 [05:55:30,07:17:24]			
Wake time (WT)	isMainSleep	06:19:54 [04:49:24,07:03:12]	06:32:06 [05:49:36,07:10:06]	05:59:06 [04:50:24,06:48:30]	06:20:36 [05:25:42,07:02:12]	<0.0001	<0.0001	0.0001
	TSP	06:34:54 [05:44:24,07:15:36]	06:47:24 [06:10:54,07:22:48]	06:36:30 [05:44:18,07:22:12]	06:42:00 [05:58:12,07:20:48]			
Time attempting to sleep (TATS)	isMainSleep	455.29 (401.05, 490.54)	456.37 (421.53, 485.75)	441.44 (400.32, 476.44)	404.72 (350.4, 447.91)	<0.0001	0.435	<0.0001
	TSP	466.23	465.32	457.55	442.03			

		(425.13, 499.34)	(433.67, 494.25)	(420.89, 490.86)	(399.17, 482.93)			
Total sleep duration	isMainSleep	446.63 (391.27, 480.3)	447.18 (413.27, 476.94)	433.84 (393.46, 468.08)	398.74 (345.23, 441.66)	<0.0001	<0.0001	<0.0001
	TSP	457.05 (416.54, 489.53)	456.64 (425, 485.36)	449.63 (412.56, 482.44)	435.74 (393.1, 476.47)			
Midsleep point (MSP)	isMainSleep	02:32:42 [01:31:12,03:14:24]	02:48:36 [02:14:30,03:27:54]	02:49:12 [01:58:12,03:34:42]	02:47:12 [02:01:42,03:28:00]	<0.0001	<0.0001	<0.0001
	TSP	02:46:48 [02:04:36,03:25:00]	02:58:12 [02:25:42,03:36:36]	03:04:06 [02:20:18,03:51:48]	03:00:18 [02:18:30,03:40:24]			
Sleep duration								
TST of primary sleep period	isMainSleep	419.97 (363.65, 452.04)	418.46 (385.85, 446.26)	401.71 (363.19, 434.31)	365.73 (313.6, 406.82)	<0.0001	<0.0001	<0.0001
	TSP	429.42 (389.58, 459.34)	425.48 (394.92, 453)	413.39 (376.88, 445.1)	387.15 (342.55, 428.13)			
Daily TST	isMainSleep	408.84	411.5	394.45	357.78	<0.0001	<0.0001	<0.0001

		(345.53, 442.93)	(378.32, 440.16)	(355.59, 426.41)	(305.42, 398.66)			
	TSP	417.96 (375.21, 448.83)	419.32 (387.54, 445.49)	405.34 (369.6, 436.39)	379.1 (333.34, 419.58)			
Total wakefulness duration	isMainSleep	33.1 (25.13, 41.39)	35.34 (29.24, 42.38)	36.19 (29.72, 43.65)	34.79 (27.05, 42.68)	<0.0001	<0.0001	0.1601
	TSP	33.45 (25.5, 41.81)	36.9 (30.58, 43.98)	40.11 (33.67, 48.29)	49.01 (40.03, 59.93)			
Sleep efficiency percentage	isMainSleep	92.54 (90.96, 93.85)	92.21 (90.85, 93.32)	91.72 (90.31, 92.9)	91.27 (89.66, 92.67)	<0.0001	<0.0001	<0.0001
	TSP	92.48 (90.9, 93.8)	91.98 (90.59, 93.08)	91.03 (89.65, 92.22)	88.96 (86.71, 90.51)			
Sleep onset latency (SOL)	isMainSleep	4.73 (2.99, 6.73)	4.44 (3.11, 6.18)	3.88 (2.7, 5.52)	3.11 (2.05, 4.52)	<0.0001	<0.0001	<0.0001
	TSP	4.88 (3.09, 6.95)	4.51 (3.17, 6.24)	3.93 (2.73, 5.58)	3.13 (2.09, 4.56)			
Latency from wake to	isMainSleep	2.99 (1.46, 4.54)	3.13 (2.04, 4.48)	2.83 (1.83, 4.12)	2.41 (1.48, 3.66)	<0.0001	<0.0001	0.1601

end of log	TSP	3.03 (1.5, 4.59)	3.1 (2.02, 4.45)	2.73 (1.77, 3.99)	2.28 (1.36, 3.51)			
Sleep Disturbances								
Wake after sleep onset (WASO)	isMainSleep	28.73 (22.07, 36.3)	31.25 (25.88, 37.53)	32.99 (26.95, 40.02)	33.06 (26.42, 40.91)	<0.0001	<0.0001	<0.0001
	TSP	29.96 (23.23, 37.9)	33.34 (27.94, 39.91)	37.73 (31.91, 45.21)	48.6 (40.28, 59.68)			
# of awakenings	isMainSleep	4.3 (3.25, 6)	4.9 (3.82, 6.57)	5.07 (3.98, 6.83)	4.81 (3.72, 6.41)	<0.0001	<0.0001	<0.0001
	TSP	4.42 (3.36, 6.24)	5.02 (3.95, 6.71)	5.3 (4.21, 7.11)	5.36 (4.21, 7.08)			
# of long awakenings	isMainSleep	0.03 (0.01, 0.06)	0.04 (0.02, 0.07)	0.05 (0.03, 0.08)	0.06 (0.03, 0.11)	<0.0001	<0.0001	<0.0001
	TSP	0.03 (0.01, 0.06)	0.05 (0.04, 0.08)	0.09 (0.06, 0.12)	0.18 (0.13, 0.26)			
Length of longest wake	isMainSleep	11.8 (9.28, 14.09)	12.56 (10.66, 14.43)	13.05 (10.98, 15.19)	13.58 (11.07, 16.09)	<0.0001	<0.0001	<0.0001
	TSP	12.45 (9.97, 14.93)	14.13 (11.65, 16.61)	16.91 (14.43, 19.39)	25.35 (22.87, 27.83)			

		14.71)	(12.07, 16.03)	(14.58, 19.35)	(21.08, 31.95)			
Sleep pattern and stages								
Wake	isMainSleep	36.71 (31.43, 43.15)	38.06 (32.9, 44.07)	39.32 (34.09, 45.33)	38.78 (33, 45.48)	<0.0001	<0.0001	<0.0001
	TSP	36.79 (31.48, 43.1)	38.15 (33.06, 44.21)	39.56 (34.32, 45.6)	39.27 (33.4, 46.15)			
Awake	isMainSleep	5.02 (3.38, 8.04)	5.8 (4.21, 7.95)	6.19 (4.37, 8.6)	6 (4.14, 8.68)	0.0036	0.5199	<0.0001
	TSP	5.2 (3.5, 8.36)	5.72 (4.15, 7.86)	5.94 (4.2, 8.33)	5.62 (3.75, 8.21)			
Restless	isMainSleep	16.66 (10, 26.41)	20.21 (13.94, 28.63)	21 (14, 30.14)	19.2 (12.19, 29.45)	<0.0001	0.5711	<0.0001
	TSP	17.9 (10.75, 27.47)	19.98 (13.82, 29.01)	20.38 (13.39, 29.87)	18.4 (11.19, 28.8)			
Imputed Wake	isMainSleep	NA	NA	NA	NA	<0.0001	NA	NA
	TSP	82.5 (65.5,	90.2 (74.53,	95.74 (81.8,	103.82			

		103.04)	106.29)	109.66)	(90.72, 119.98)			
Asleep	isMainSleep	350.6 (205.15, 428.46)	374.9 (280.11, 424.63)	346.77 (246.34, 407.17)	283.63 (184.52, 368.02)	<0.0001	0.3895	<0.0001
	TSP	377.01 (235.66, 439.15)	373.5 (269.37, 426.06)	340.99 (223.14, 406.85)	265.68 (162.42, 359.97)			
Light	isMainSleep	273.6 (249.73, 300.41)	273.28 (250.16, 298.37)	269.38 (244.31, 294.84)	256.5 (228.3, 284.74)	<0.0001	<0.0001	<0.0001
	TSP	274.4 (250.42, 300.88)	274.1 (251.22, 298.98)	271.16 (245.81, 296.45)	260.03 (232.42, 289.15)			
Light as % of TST	isMainSleep	63.77 (60.07, 68.46)	64.15 (60.6, 68.42)	65.3 (61.46, 69.52)	67.63 (63.03, 71.95)	<0.0001	<0.0001	<0.0001
	TSP	63.58 (59.96, 68.16)	63.94 (60.35, 68.2)	64.74 (60.9, 68.92)	65.48 (61.4, 69.89)			
Deep	isMainSleep	65.73 (54, 75.9)	64.22 (53.87, 74.37)	60.54 (50.78, 70.05)	53.12 (43.19, 63.54)	<0.0001	<0.0001	<0.0001
	TSP	66.23 (54.68, 76.19)	64.38 (54.01, 74.7)	60.81 (51.11,	53.8 (43.79,			

				70.49)	64.46)			
Deep as % of TST	isMainSleep	15.42 (12.99, 17.51)	15.11 (12.84, 17.37)	14.75 (12.47, 16.92)	13.84 (11.7, 16.28)	<0.0001	<0.0001	<0.0001
	TSP	15.45 (13.01, 17.52)	15.06 (12.84, 17.35)	14.66 (12.36, 16.8)	13.46 (11.3, 15.9)			
REM	isMainSleep	89.6 (74.7, 102.82)	88.11 (74.94, 101.05)	83.64 (70.24, 96.81)	73.85 (59.78, 87.84)	<0.0001	<0.0001	<0.0001
	TSP	89.97 (75.39, 103.19)	88.43 (75.35, 101.31)	84.17 (70.69, 97.16)	74.74 (60.71, 88.83)			
REM as % of TST	isMainSleep	20.82 (17.85, 23.19)	20.7 (17.88, 23.04)	20.13 (17.37, 22.41)	18.98 (16.26, 21.32)	<0.0001	<0.0001	<0.0001
	TSP	20.87 (17.93, 23.23)	20.64 (17.86, 23.03)	20.02 (17.27, 22.25)	18.43 (15.7, 20.92)			
Non-primary sleep period								
Count per	isMainSleep	0.04 (0.02,	0.08 (0.05,	0.15 (0.1,	0.31 (0.22,	<0.0001	<0.0001	<0.0001

day		0.09)	0.14)	0.22)	0.46)			
	TSP	0.06 (0.02, 0.15)	0.07 (0.03, 0.15)	0.1 (0.05, 0.2)	0.15 (0.07, 0.27)			
TST-NP	isMainSleep	90.34 (78.04, 105.45)	95.95 (84.27, 108.73)	100.51 (87.87, 112.86)	105.54 (91.02, 121.49)	<0.0001	<0.0001	<0.0001
	TSP	97.5 (82, 120.85)	100.02 (86.29, 119.12)	100.96 (86.93, 120.86)	101.93 (86.16, 123.33)			

Note that the total number of logs across quartiles may not equal the total number of logs given the simplified approach of definition that was used to identify sleep logs and in these analyses we only consider nights where a non-waking level is present.

A[^] - Refers to main effect of quartile

B^{^^} - Refers to main effect of algorithm

AxB^{^^^} - Refers to interaction effect of quartile and algorithm

Supplemental Table 2: Descriptive stats of sleep metrics by calendar-relative (isMainSleep) and user-centric (TSP) algorithms by 3-level categories of variation from typical sleep patterns

Variable	Algorith m	Group 1 (Cleaning)	Group 2 (No change)	Group 3 (most stitched logs)	P-value (A [^])	P-value (B ^{^^})	P-value (AxB ^{^^})
Mean # logs / day (Range)	NA	[-0.0076-0]	[0-0.0594]	>= 0.0594			

Number of participants	NA	59	6361	2143			
Number of logs	isMainSleep	54720	5538791	1711999	NA	NA	NA
	TSP	52627	5553851	1902795			
Person-days	isMainSleep	50453	5572696	1756873	NA	NA	NA
	TSP	50192	5544514	1738846			
Sleep schedule							
Bedtime (BT)	isMainSleep	23:40:30 [22:45:18,00:46:06]	22:25:06 [21:45:30,23:24:06]	23:03:54 [22:20:06,23:51:54]	<0.0001	0.0313	<0.0001
	TSP	23:33:48 [22:38:06,00:42:24]	22:25:48 [21:49:42,23:30:18]	23:06:36 [22:22:54,23:55:12]			
Sleep onset	isMainSleep	23:43:06 [22:49:30,00:48:48]	22:32:36 [21:50:18,23:29:12]	23:08:42 [22:25:12,23:56:00]	0.0254	<0.0001	<0.0001
	TSP	23:37:36 [22:41:12,00:46:12]	22:33:48 [21:54:30,23:34:00]	23:11:18 [22:28:18,23:59:48]			
Sleep offset	isMainSleep	05:56:18	06:24:48	06:22:12	<0.0001	<0.0001	<0.0001

	leep	[04:48:24,06:45:36]	[05:30:00,06:53:18]	[05:26:00,07:02:06]			
	TSP	06:33:36 [05:42:36,07:20:24]	06:34:00 [05:54:36,07:12:24]	06:39:00 [05:56:24,07:16:36]			
Wake time (WT)	isMainSleep	05:59:06 [04:50:24,06:48:30]	06:30:12 [05:35:24,06:56:48]	06:25:12 [05:28:42,07:05:18]	<0.0001	<0.0001	0.0003
	TSP	06:36:24 [05:44:18,07:22:12]	06:36:24 [05:58:00,07:14:24]	06:42:06 [05:58:42,07:20:00]			
Time attempting to sleep (TATS)	isMainSleep	486.19 (460.34, 518.9)	450.91 (408.83, 484.5)	404.8 (350.48, 447.9)	<0.0001	0.0474	<0.0001
	TSP	486.11 (462.57, 519.2)	462.81 (426.54, 494.75)	442.03 (399.17, 482.91)			
Total sleep duration	isMainSleep	475.42 (453.05, 509.34)	442.8 (400.92, 475.29)	398.88 (345.41, 441.54)	0.0113	<0.0001	<0.0001
	TSP	477.37 (454.67, 508.84)	454.52 (418.68, 485.34)	435.74 (393.25, 476.36)			
Midsleep point	isMainSleep	02:49:06	02:28:30	02:43:18	<0.0001	<0.0001	<0.0001

(MSP)	eep	[01:58:24,03:34:36]	[01:50:00,03:07:24]	[02:00:30,03:23:24]			
	TSP	03:04:00 [02:20:12,03:51:48]	02:37:24 [01:56:36,03:16:30]	02:54:48 [02:16:42,03:34:48]			
Sleep duration							
TST of primary sleep period	isMainSleep	434.61 (410.86, 469.47)	413.1 (371.34, 444.81)	365.77 (313.61, 406.78)	<0.0001	<0.0001	<0.0001
	TSP	439.42 (416.96, 468.34)	422.51 (387.25, 452.56)	387.31 (342.55, 428.08)			
Daily TST	isMainSleep	428.53 (404.18, 466.19)	404.9 (361.51, 437.34)	357.78 (305.46, 398.65)	<0.0001	<0.0001	<0.0001
	TSP	432.15 (408.59, 466.21)	413.89 (377.79, 444)	379.1 (333.36, 419.53)			
Total wakefulness duration	isMainSleep	39.74 (33.89, 47.36)	34.87 (28.18, 42.45)	34.79 (27.03, 42.7)	<0.0001	0.2681	0.9342
	TSP	41.61 (33.65, 52.22)	36.93 (29.85, 44.7)	49.01 (40.02, 59.92)			
Sleep efficiency	isMainSleep	91.12 (89.15,	92.16 (90.7,	91.27 (89.66,	<0.0001	<0.0001	<0.0001

percentage	eep	92.54)	93.37)	92.67)			
	TSP	91.28 (89.48, 92.99)	91.82 (90.32, 93.07)	88.96 (86.72, 90.51)			
Sleep onset latency (SOL)	isMainSleep	5.38 (3.78, 6.76)	4.31 (2.9, 6.16)	3.11 (2.05, 4.52)	<0.0001	<0.0001	<0.0001
	TSP	5.42 (3.81, 6.76)	4.39 (2.94, 6.26)	3.13 (2.08, 4.55)			
Latency from wake to end of log	isMainSleep	3.38 (2.62, 4.78)	2.97 (1.81, 4.41)	2.41 (1.47, 3.67)	<0.0001	0.2681	0.9342
	TSP	3.21 (2.64, 4.58)	2.94 (1.78, 4.35)	2.28 (1.36, 3.51)			
Sleep Disturbances							
Wake after sleep onset (WASO)	isMainSleep	34.38 (29.88, 42.81)	30.95 (24.95, 38.18)	33.06 (26.41, 40.91)	<0.0001	<0.0001	<0.0001
	TSP	37.05 (30.4, 48.62)	33.85 (27.42, 41.28)	48.59 (40.27, 59.67)			
# of awakenings	isMainSleep	4.83 (3.61, 6.67)	4.77 (3.7, 6.49)	4.81 (3.72, 6.41)	<0.0001	<0.0001	<0.0001
	TSP	4.83 (3.67, 6.91)	4.94 (3.85, 6.71)	5.36 (4.2, 7.07)			

# of long awakenings	isMainSleep	0.05 (0.04, 0.07)	0.04 (0.02, 0.07)	0.06 (0.03, 0.11)	<0.0001	<0.0001	<0.0001
	TSP	0.06 (0.04, 0.13)	0.06 (0.03, 0.1)	0.18 (0.13, 0.26)			
Length of longest wake	isMainSleep	13.72 (11.74, 15.87)	12.5 (10.36, 14.56)	13.58 (11.07, 16.09)	<0.0001	<0.0001	<0.0001
	TSP	14.65 (12.42, 17.92)	14.4 (11.86, 17.11)	25.35 (21.08, 31.95)			
Sleep pattern and stages							
Wake	isMainSleep	40.36 (34.95, 44.46)	38.08 (32.72, 44.25)	38.78 (32.97, 45.49)	0.0003	<0.0001	<0.0001
	TSP	40.11 (34.47, 44.18)	38.23 (32.88, 44.37)	39.27 (33.39, 46.16)			
Awake	isMainSleep	5.64 (4.27, 7.56)	5.72 (4, 8.18)	6 (4.14, 8.68)	0.9297	0.0071	<0.0001
	TSP	5.5 (4.1, 8)	5.67 (4, 8.13)	5.62 (3.75, 8.21)			
Restless	isMainSleep	21.64 (13.03, 35.19)	19.34 (12.73, 28.61)	19.2 (12.19, 29.44)	<0.0001	<0.0001	<0.0001
	TSP	23.01 (11.6, 32.87)	19.47 (12.89, 29)	18.4 (11.19, 28.77)			

Imputed Wake	isMainSleep				<0.0001	<0.0001	<0.0001
	TSP	39.27 (24.03, 57.12)	91.5 (74.91, 107.64)	103.79 (90.73, 119.94)			
Asleep	isMainSleep	317.63 (199.24, 390.04)	359.52 (245.87, 420.24)	283.63 (184.56, 368.08)	<0.0001	<0.0001	<0.0001
	TSP	289.2 (164.14, 371.02)	364.53 (243.9, 424.41)	265.68 (162.55, 359.92)			
Light	TSP	281.82 (254.77, 305.49)	271.89 (247.95, 297.61)	256.5 (228.33, 284.78)	<0.0001	<0.0001	<0.0001
	isMainSleep	282.54 (253.77, 306.21)	272.96 (249.05, 298.42)	260.03 (232.44, 289.13)			
Light as % of TST	TSP	64.17 (57.89, 67.42)	64.43 (60.77, 68.81)	67.63 (63.03, 71.95)	0.0334	0.22	<0.0001
	isMainSleep	63.63 (57.86, 67.29)	64.15 (60.47, 68.48)	65.48 (61.41, 69.89)			
Deep	TSP	65.35 (57.38, 75.45)	63.14 (52.74, 73.58)	53.12 (43.22, 63.56)	<0.0001	NA	NA
	isMainSleep	66.11 (58.9, 75.32)	63.51 (53.03, 73.99)	53.8 (43.8, 63.8)			

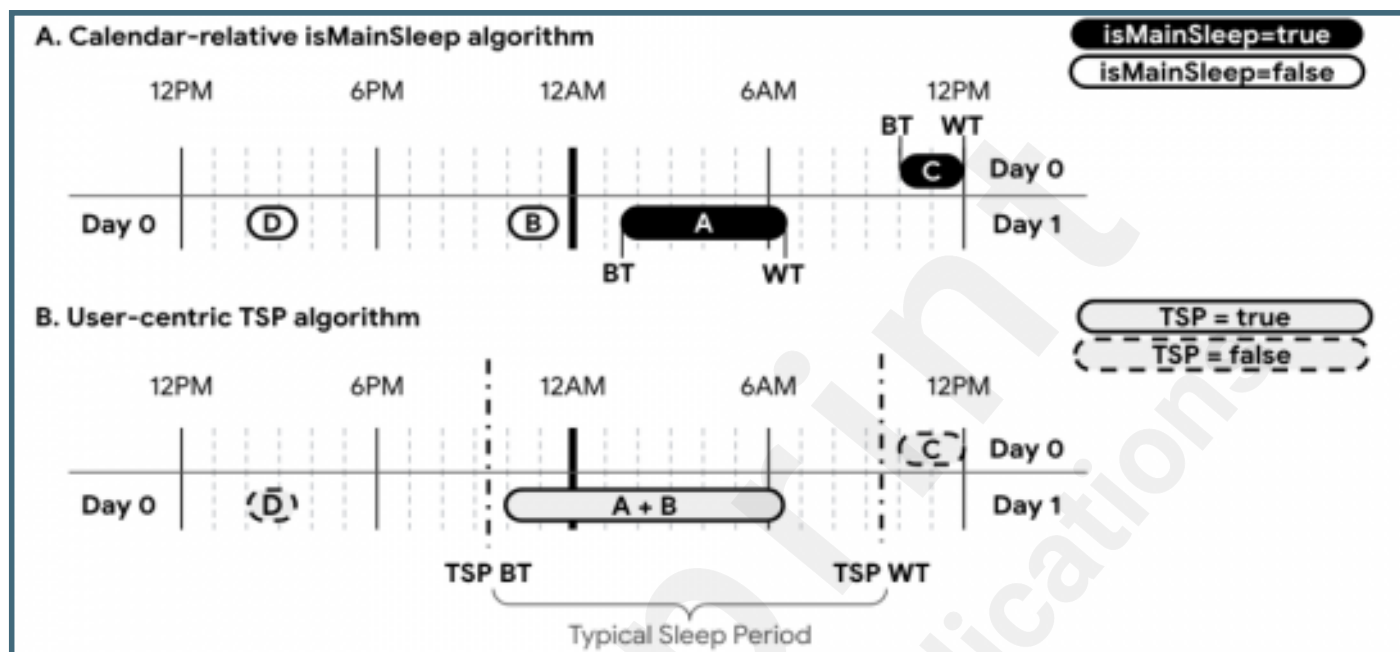
	eep	75.5)	73.93)	64.46)			
Deep as % of TST	TSP	14.82 (13.13, 16.2)	15.07 (12.74, 17.3)	13.84 (11.69, 16.28)	<0.0001	<0.0001	<0.0001
	isMainSleep	14.76 (13.13, 16.25)	15.02 (12.71, 17.26)	13.46 (11.3, 15.91)			
REM	TSP	89.36 (72.94, 100.61)	86.96 (72.86, 99.97)	73.85 (59.79, 87.85)	<0.0001	0.0002	<0.0001
	isMainSleep	89.41 (74.59, 100.55)	87.36 (73.37, 100.34)	74.74 (60.72, 88.87)			
REM as % of TST	TSP	19.43 (16.57, 22.33)	20.55 (17.71, 22.91)	18.98 (16.27, 21.32)	<0.0001	0.0002	<0.0001
	isMainSleep	19.6 (16.55, 22.38)	20.5 (17.68, 22.86)	18.43 (15.71, 20.93)			
Non-primary sleep period							
Count per day	isMainSleep	0.11 (0.06, 0.22)	0.09 (0.05, 0.16)	0.31 (0.22, 0.46)	<0.0001	<0.0001	<0.0001
	TSP	0.12 (0.05, 0.29)	0.08 (0.03, 0.17)	0.15 (0.07, 0.27)			
TST-NP	isMainSleep	89.98 (77.4, 101.55)	95.99 (83.56, 109.88)	105.44 (91.01, 121.48)	<0.0001	<0.0001	<0.0001

	TSP	97.83 (75.01, 112.58)	99.74 (85.5, 120.22)	101.9 (86.15, 123.28)			
Note that the total number of logs across 3-level group may not equal the total number of logs given the simplified approach of definition that was used to identify sleep logs and in these analyses we only consider nights where a non-waking level is present.							
A^	-	Refers	to	main	effect	of	quartile
B^^	-	Refers	to	main	effect	of	algorithm
AxB^^^ - Refers to interaction effect of quartile and algorithm							

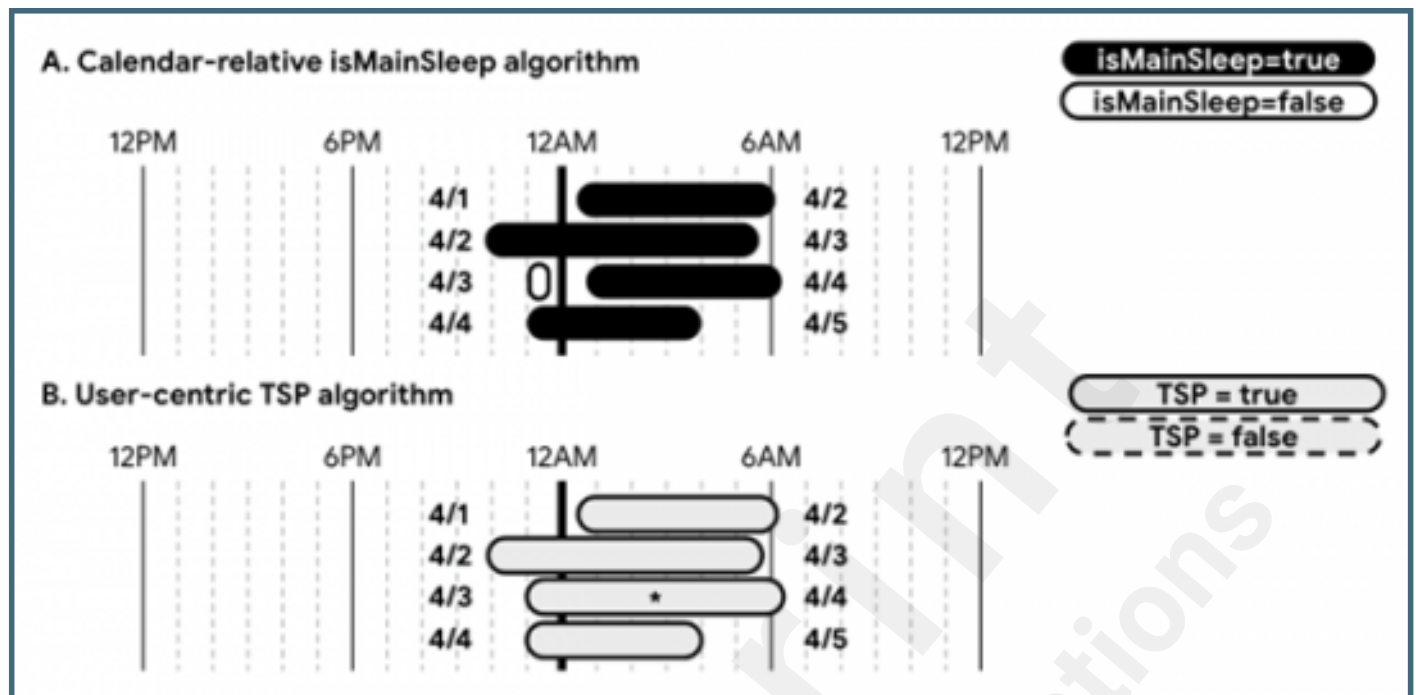
Supplementary Files

Figures

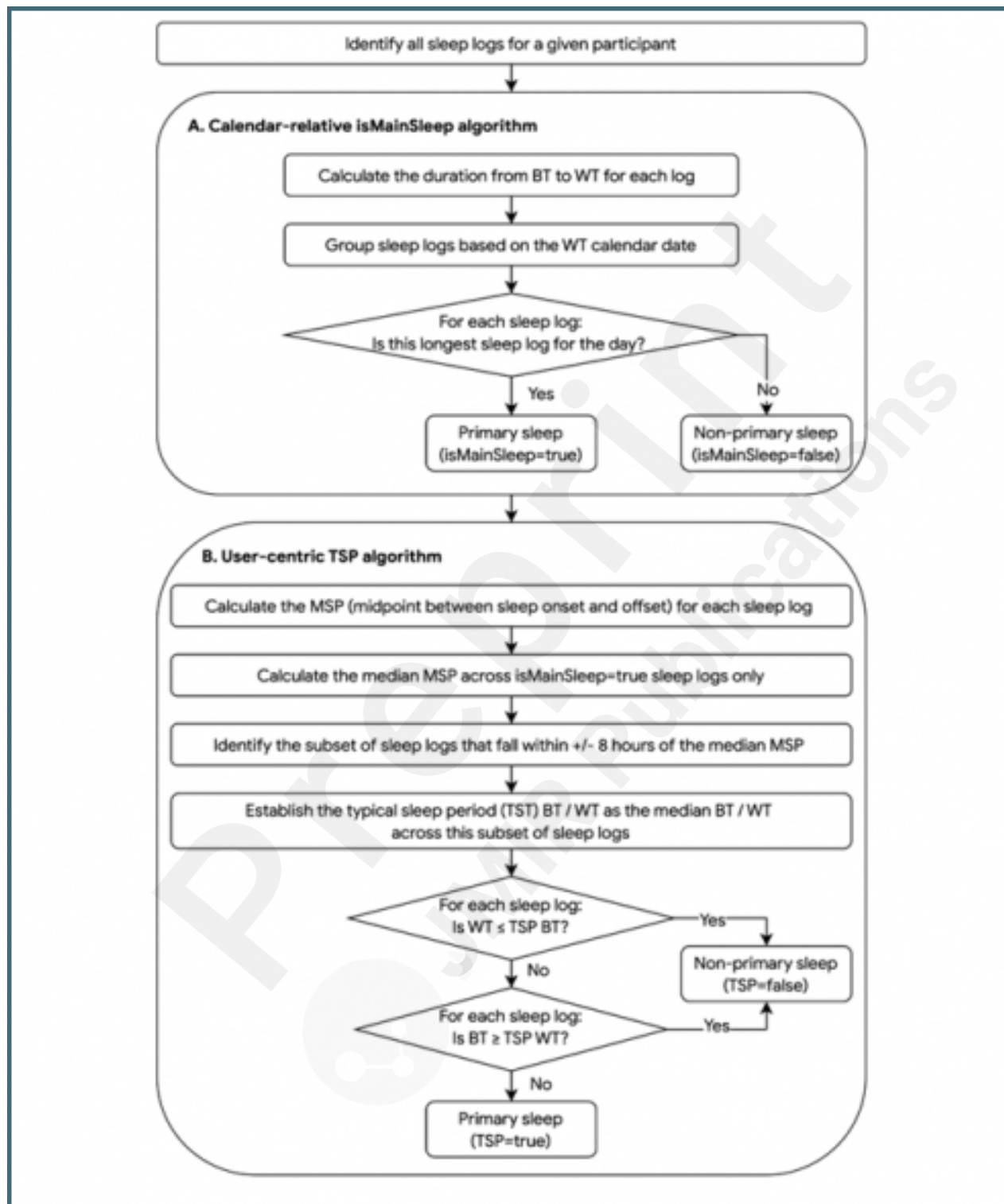
Conceptual overview of how the calendar-relative “isMainSleep” and user-centric “TSP” algorithms handle high-resolution sleep data. TSP BT and TSP WT are derived from the median bedtime and wake time across all relevant sleep logs around a user’s median mid-sleep point (MSP) (not shown). TSP = typical sleep period, BT = bedtime, WT = wake time.



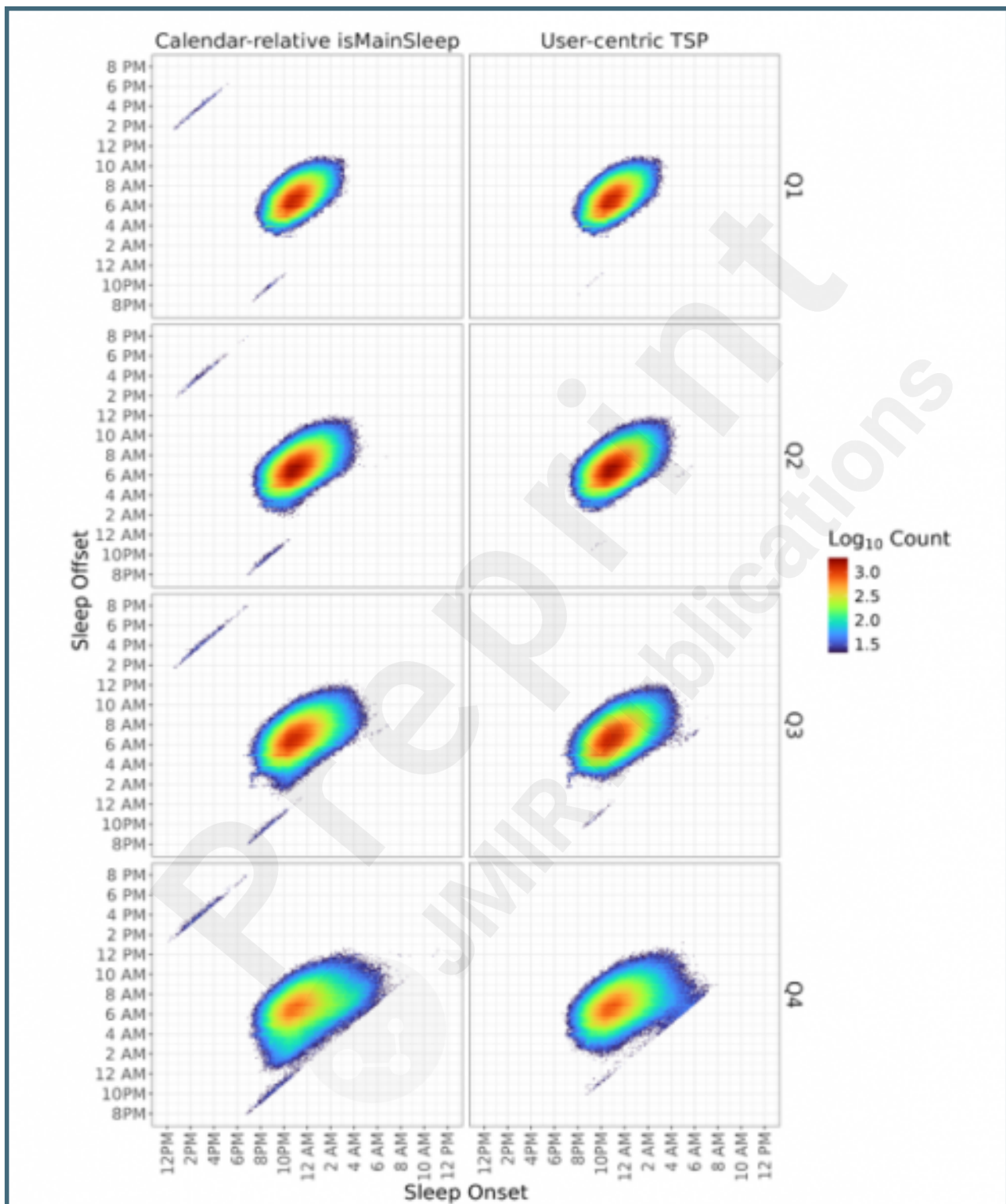
Calendar-relative 'isMainSleep' vs.user-centric 'TSP' algorithm - an illustrative data example.



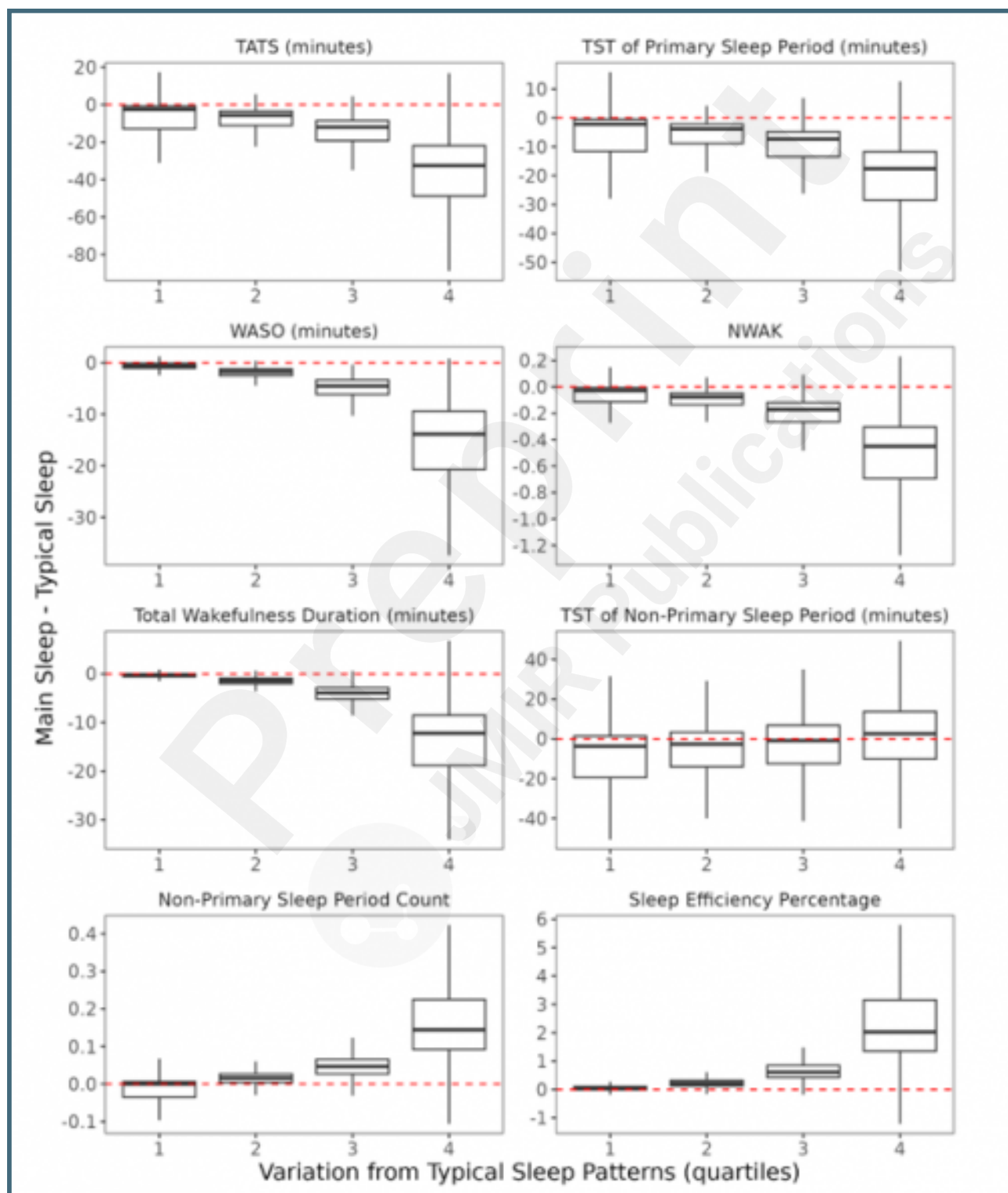
A-B: Decision making process for sleep log classification using calendar-relative and user-centric algorithms.



Sleep onset offset by quartiles of variation from typical sleep patterns using calendar-relative and user-centric algorithms.



Boxplots of paired differences between user-centric (TSP) and calendar-relative (isMainSleep) algorithms for each of the key hypothesized sleep metrics across the quartiles of variation from typical sleep patterns. Box bounds, midline, and whiskers represent the IQR, median, and $1.5 \times \text{IQR}$, respectively. Values above the dotted red line indicate a higher value from the isMainSleep algorithm relative to the TSP algorithm; whereas, values below the red line indicate a lower value.



Multimedia Appendixes

Distribution of change in the mean number of primary sleep logs per night as a measure of variation from typical sleep patterns. Negative values represent participants whose non-primary sleep were mislabeled as primary sleep under isMainSleep. Values between 0 and 0.05 represent participants whose sleep logs were consistently classified under both algorithms. Values >0.05 represent participants whose sleep tends to be interrupted, and were previously incompletely captured under isMainSleep.

URL: <http://asset.jmir.pub/assets/10fa4434e444578e49d079e94c9fffd5.png>

