World Sleep Society Recommendations for the Use of Wearable Consumer Health Trackers That Monitor Sleep

Michael WL. Chee, Mathias Baumert, Hannah Scott, Nicola Cellini, Cathy Goldstein, Kelly Baron, Syed A. Imtiaz, Thomas Penzel, Clete A. Kushida, on behalf of the World Sleep Society Sleep Tracker Task Force

PII: S1389-9457(25)00173-X

DOI: https://doi.org/10.1016/j.sleep.2025.106506

Reference: SLEEP 106506

To appear in: *Sleep Medicine*

Received Date: 28 March 2025

Accepted Date: 4 April 2025

Please cite this article as: Chee MW, Baumert M, Scott H, Cellini N, Goldstein C, Baron K, Imtiaz SA, Penzel T, Kushida CA, on behalf of the World Sleep Society Sleep Tracker Task Force, World Sleep Society Recommendations for the Use of Wearable Consumer Health Trackers That Monitor Sleep, *Sleep Medicine*, https://doi.org/10.1016/j.sleep.2025.106506.

World Sleep Society Recommendations for the Use of Wearable Consumer Health Trackers That Monitor Sleep

CONSUMERS

- A sufficiently good-quality CHT can provide valuable sleep and health information
- Focus on trends and patterns, not individual 'scores'; Use scores to motivate, not compete
- Sleep data is less reliable if sleep is significantly delayed, very short or fragmented

CLINICIANS

- Distinguish between 'fundamental' and 'exploratory' metrics
- Appreciate growing uses of CHT as well as their limitations in support of clinical diagnosis and monitoring
- Stay in touch with evolving best practices for usage

RESEARCHERS

- Harness CHT to broaden, deepen and diversify sleep norms and their relationship to health and disease
- Work with manufacturers to curate quality standards and a set of unified measures for different sleep health applications

MANUFACTURERS

- Implement common definitions of fundamental sleep measures to allow incorporation of sleep data into health records
- Support industry-wide quality standards for different applications
- Continue investing in sleep-themed features

Michael WL Chee, Mathias Baumert, Hannah Scott, Nicola Cellini, Cathy Goldstein, Kelly Baron, Syed Anas Imtiaz, Thomas Penzel, Clete Kushida

WORLD SLEEP SOCIETY
Advancing Sleep Health Worldwide

World Sleep Society Recommendations for the Use of Wearable Consumer Health Trackers That Monitor Sleep

Michael WL Chee[1], Mathias Baumert[2], Hannah Scott[3], Nicola Cellini[4,5], Cathy Goldstein[6], Kelly Baron[7], Syed A Imtiaz[8], Thomas Penzel[9], Clete A Kushida[10] on behalf of the World Sleep Society Sleep Tracker Task Force

1. Centre for Sleep and Cognition, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
2. Discipline of Biomedical Engineering, School of Electrical and Mechanical Engineering, The University of Adelaide, Adelaide, Australia
3. Flinders Health and Medical Research Institute: Sleep Health, College of Medicine & Public Health, Flinders University, Adelaide, Australia
4. Department of General Psychology, University of Padua, Padua, Italy
5. Human Inspired Technologies Research Center, University of Padua, Padua, Italy
6. University of Michigan Sleep Disorders Center, University of Michigan Health, Ann Arbor, Michigan, United States
7. Department of Family and Preventive Medicine, University of Utah, Salt Lake City, Utah, United States
8. Wearable Technologies Lab, Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom
9. Interdisciplinary Sleep Medicine Center, Charite Universitätsmedizin Berlin, Berlin, Germany
10. Division of Sleep Medicine, Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, California, United States

Word count: 15695 (with table excluding figure captions)
Number of Tables: 1
Number of Figures: 5

Address correspondence to: Michael.chee@nus.edu.sg

Keywords: Wearable sleep monitoring, Consumer sleep trackers, Sleep health assessment, Sleep Disorders, Electronic Health Records, Digital sleep health, Actigraphy, Sleep assessment, Heart Rate Variability, Sleep algorithms

Abstract

Wearable consumer health trackers (CHTs) are increasingly used for sleep monitoring, yet their utility remains debated within the sleep community. To navigate these perspectives, we propose pragmatic, actionable recommendations for users, clinicians, researchers, and manufacturers to support CHT usage and development. We provide an overview of the evolution of multi-sensor CHTs, detailing common sensors and sleep-relevant metrics. We advocate for standardized 'fundamental sleep measures' across manufacturers, distinguishing these from proprietary exploratory metrics with future potential. We outline best practices for using CHT-derived sleep data in healthy individuals while addressing current device limitations. Additionally, we explore their role in evaluating and managing individuals at risk for or diagnosed with insomnia, sleep apnea, or circadian rhythm sleep-wake disorders. Guidance is provided on device selection to align with their intended use and on conducting and interpreting performance evaluation studies. Collaboration with manufacturers is needed to balance feature comprehensiveness with clinical utility and usability. Finally, we examine challenges integrating heterogeneous sleep data into clinical health records and discuss medical device certification for specific wearable CHT features. By addressing these issues, our recommendations aim to inform the usage of CHTs in the global community and to begin bridging the gap between consumer technology and clinical application, maximizing the potential of CHTs to enhance both personal and community sleep health.

Abbreviations

AASM	American Academy of Sleep Medicine
AHI	Apnea-Hypopnea Index
ANSI	American National Standards Institute
API	Application Programming Interface
CBT-I	Cognitive Behavioral Therapy for Insomnia
CE	Conformité Européenne (European Conformity)
CHT	Consumer Health Tracker
CI	Confidence Interval
CRSWD	Circadian Rhythm Sleep-Wake Disorder
CTA	Consumer Technology Association
EBE	Epoch-by-epoch analysis
EDA	Electrodermal Activity
EHR	Electronic Health Record
FDA	Food and Drug Administration
FSM	Fundamental Sleep Measure
HRV	Heart Rate Variability
IS	Interday Stability
LOA	Limits of Agreement
MET	Metabolic Equivalent of Task
NN	Normal-to-Normal
NPMA	National Medical Products Administration
NREM	Non Rapid Eye Movement
NTC	Negative Temperature Coefficient
NSF	National Sleep Foundation
OSA	Obstructive Sleep Apnea
PABAK	Prevalence-Adjusted and Bias-Adjusted Kappa
PDMA	Pharmaceuticals and Medical Devices Agency
PPG	Photoplethysmography

PSG	Polysomnography
PSQI	Pittsburgh Sleep Quality Index
REM	Rapid Eye Movement
RMSSD	Root Mean Square of Successive Differences
RHR	Resting Heart Rate
RR	Respiratory Rate
SaMD	Software as a Medical Device
SDB	Sleep Disordered Breathing
SDNN	Standard Deviation of Normal-to-Normal
SHI	Sleep Health Index
SL	Sleep Latency
SAD	Standardized Absolute Difference
sMAPE	Symmetric Mean Absolute Percentage Error
SRI	Sleep Regularity Index
SWS	Slow Wave Sleep
TATS	Time Attempting to Sleep
TGA	Therapeutic Goods Administration
TST	Total Sleep Time
UKCA	United Kingdom Conformity Assessment
WASO	Wake After Sleep Onset

1.0. INTRODUCTION

1.1 Brief history of consumer health trackers and underlying technology

Wearable consumer health trackers (CHTs) entered the consumer market around 2011, with Jawbone, Fitbit, and Garmin being early pioneers. The first two included sleep tracking capabilities shortly after launch and are thus most widely credited as the first consumer sleep trackers using motion and heart rate signals for sleep detection. Despite its early success, Jawbone declared bankruptcy in 2017. Undeterred, numerous new manufacturers have entered the industry, focusing on different market segments. As many CHTs have incorporated sleep as a significant element of their health offerings, sleep tracking has grown in popularity even though most CHTs are purchased for fitness tracking [1].

The development of small, inexpensive sensors to capture various physiological signals has fueled the growth of the CHT market. Fitness and sleep tracking was initially based primarily on accelerometry, a technology used since the early 1970s [2]. However, it was not until the late 1980s that sleep researchers began to realize their potential [3, 4]. The pioneering work of Cole and Kripke for sleep tracking in adults [5] and Sadeh in adolescents [6] brought us 'actigraphy' as we know it today. Their legacy algorithms for sleep-wake discrimination remain influential even in today's machine learning era [7, 8]. Single-axis piezoelectric accelerometers have since been replaced by miniaturized triaxial sensors. The addition of rotational and gravitational effects measurement has resulted in up to '9-axis' motion assessment. The resolution of these devices has also improved with time, enabling new applications. For example, Apple's Series 7 watch yielded raw accelerometry data that outperformed a 'research grade' actigraph [9] when analyzed to infer pulse rate during sleep.

Photoplethysmography (PPG) also had a long gestation period, arising from technical issues that were solved in the early 1980s. This paved the way for pulse oximeters to become standard medical equipment[10]. Miniaturized PPG sensors were incorporated into wearables in the 2010s, with exercise heart rate being the primary application. PPG allows for continuous pulse rate monitoring, although this is subject to limitations discussed later. Along similar lines, temperature and skin conductance sensors have been added. Multisensor integration of physiological data has enabled some 'consumer devices' to outperform far more expensive 'research-grade' actigraphy [11, 12].

The rapid technological development of CHTs is targeted towards fulfilling consumer needs and manufacturer's business objectives, which may differ from medical and research goals [13]. Additionally, industry's capacity to develop and disseminate health-related applications at scale, provide a sophisticated user experience, and offer cutting edge data management far exceeds the capabilities of academics. These steady improvements will continue to benefit from a >10% annual compounded growth in sales of health trackers [1].

Recommendation:

Sleep medicine practitioners and researchers should keep abreast of technological advances and embrace the opportunities offered by consumer health trackers while remaining cognizant of their limitations. Sleep professionals should leverage the public interest and curiosity about sleep that these devices have engendered and contribute constructively to the growth of their clinical or research utilization.

1.2 Motivations for the recommendations

Growing personal adoption of health trackers (in contrast to devices given to study participants or patients for a short monitoring period) has tremendous potential to benefit human health and wellbeing by affording a more granular and comprehensive characterization of sleep over time periods longer than are practical with traditional actigraphy devices

Current recommendations regarding 'adequate sleep duration' [14, 15] are very heavily weighted on findings obtained from questionnaires about 'time in bed' rather than time asleep. These recommendations do not consider factors such as weekday-weekend sleep variations or sex, vocational, cultural, and sociodemographic influences. Furthermore, most of the world's scientific publications about sleep originate from North America and Europe, representing only 15% of the global population, limiting the generalizability of these recommendations [16]. Although awareness of the multidimensional nature of sleep is acknowledged [17], norms for sleep regularity, efficiency, and continuity across the lifespan and in different cultures remain to be established [18]. Health trackers and their associated apps could potentially collect relevant data to provide demographic-tailored and periodically updated recommendations.

Fueling growing adoption, novel methods for using physiological data obtained from health trackers are emerging. For example, temperature tracking to predict a woman's fertile period [19-21] and early detection of febrile illnesses [22] are by-products of incorporating temperature sensors into CHTs. Cardiovascular applications like the detection of atrial fibrillation [23], continuous blood pressure measurement [24], and monitoring of arterial stiffness [25, 26] have similarly arisen.

Consumer enthusiast blogs, tech-device reviewers, and social media influencers guide public perceptions about sleep and sleep measurement tools. These opinions are more widely disseminated and may influence purchasing decisions and product development more than clinicians or sleep scientists.

Collaboration with industry to establish standardized definitions for sleep-related variables and co-create benchmarks for measurement accuracy for different applications and contexts (e.g. persons with mostly normal nocturnal sleep, shift workers, persons with difficulty with sleep initiation or maintenance, persons with sleep disorders) is crucial. Relationships between scientists, clinicians, and

industry are needed to encourage manufacturers to prioritize the development of improved sleep-related applications, which will prompt the investment of necessary resources to accomplish this.

1.3 Purpose of the recommendations

Within the last seven years, several reviews and position papers [27-31] have addressed the promise and pitfalls of using CHTs in clinical and/or research settings. They have provided guidance on standardizing sleep measurements [32, 33] and their evaluation [34, 35]. The present recommendations update and complement these documents, expanding on practical points akin to the guide for actigraphy utilization for clinical and research purposes published over a decade ago [4].

In the present day, these recommendations aim to provide:

1. A user-oriented guide to the technology and fundamental measures relevant to sleep.
2. Actionable[1] information for healthcare professionals, providing an approach for advising their patients and/or clients.
3. Actionable information for end-users seeking to use these devices to understand their sleep patterns and how their sleep compares with their peers, keeping in mind the current limitations of CHT.
4. Preliminary guidance on how to use these trackers in individuals concerned about, at risk of, or diagnosed with common sleep disorders.

For the near-term future, we seek to provide a roadmap for:

1. Using standardized definitions for 'fundamental sleep measurements', delineating these from 'exploratory' measurements that lack definitions, gold-standard correlates, or normative data.
2. Harnessing, storing, and harmonizing wearable sleep health information for use by healthcare professionals or trained sleep health advisors to improve sleep health.

To provide suggestions for manufacturers to:

1. Adopt a standardized lexicon for 'fundamental sleep measures (FSM)' that all trackers will report, allowing for comparison of sleep parameters across different devices.
2. Differentiate FSM from exploratory sleep parameters and device features that do not yet have clear physiological or clinical significance.
3. Collaborate with sleep professionals to co-create and adhere to a set of measurement precision standards for FSM to enable generalizability across devices. Gold-standard measurement methods (i.e., PSG for sleep) should be used as a reference.

[1] advice, or recommendations that are clear, practical, and useful for taking specific actions.

2.0. SCOPE OF GUIDELINE

2.1. Audience: This document is primarily for knowledgeable end-users, healthcare professionals, researchers, and developers or manufacturers interested in the present and future uses of health trackers in sleep and allied health applications.

2.2. Focus on consumer wearables with health tracking: We focus only on electronic devices worn on the body, such as a watch, wristband, ring, and armband, able to continuously monitor and record health-related data as this is the most commonly purchased class of devices. While subject to standard consumer electronic safety regulations, these devices are not typically classified as medically regulated devices (though some manufacturers have obtained clearance for specific software features like atrial fibrillation detection). CHT are typically used by individuals to gain insights into their health and wellness, although they can also be integrated into broader health management programs.

Sleep nearables are an emerging consumer device category but are not covered here given the heterogeneity of sensor types and smaller representation in the marketplace and peer-reviewed research.

2.3. Focus on usage by adults: This document will focus on CHT use by adults from 20-65 years of age, the population with most empirical evidence to support the recommendations and the demographic that most users come from [1].

3.0. OVERVIEW OF WEARABLE SENSORS AND SLEEP MEASURES

3.1 Wearable sensors: What they measure, why, and things to note

This section provides background information on how CHT embedded sensors relate to sleep measurement. A comprehensive list of sensors [36], as well as additional technical details, such as sampling rate, filtering, quality control, and integration, are important in determining the signal quality and, thereafter, the inferences made. However, they are out of scope for this publication and are discussed elsewhere [28, 29, 37, 38].

Each sensor collects raw data that typically is further processed within the device to intermediate variables (Fig.1). Proprietary firmware, which is regularly updated, is responsible for this transformation. Reduced data is transferred to a smartphone application and, in many cases, to a cloud-based server where algorithms analyze multisensor data and output sleep, activity, and other health-related parameters. The specifics of these steps differ across manufacturers.

The transformation of sensor data to sleep measures is executed primarily by proprietary machine learning algorithms trained on data from nocturnal recordings of sleep in mostly healthy volunteers [39, 40] or clinical convenience samples. Prior to classifying sleep stages, proprietary algorithms attempt to determine sleep onset. As physiology or its mapping to sleep stages will differ according to age, body habitus, and health condition (including sleep disorders), this transformation seems to benefit from context-specific tuning. To date, models have also not been trained on naps and non-nocturnal sleeping habits.

Practical Notes:

Sleep detection and staging using motion and autonomic signals requires the user to have relatively normal limb mobility, perfusion, and autonomic function. The latter may be altered in persons with sleep disorders [41, 42] or other medical conditions. Whenever these basic assumptions are violated, or when there are extended periods of sitting or lying still when awake, these devices' outputs will be less accurate.

Most users will rely on manufacturer-provided outputs. Researchers or enthusiasts with technical expertise may want to access raw or minimally processed sensor data to perform specific analyses that are unavailable by default [9, 43]. However, these types of analyses are not frequently pursued by consumers or clinicians [44].

—Fig. 1 Around here

3.1.1 Accelerometry

Accelerometry is used to quantify body movement, which is reduced when we fall asleep [3, 4], further reduced in deeper NREM sleep, and absent during REM in healthy persons. Therefore, an accelerometer is a fundamental sensor for sleep detection. Most wearables use at least a single triaxial sensor to measure movement in three axes (rotation is often added, with some using the term ‘6-axis’). Accelerometry data is sampled many times a second, resulting in massive data volumes. Acceleration data is then processed with proprietary algorithms to yield intermediate variables like activity counts, steps, time spent at different intensities, and energy expenditure (e.g., calories or metabolic equivalents, METS).

In many instances (even with ‘research grade instruments’), only these reduced variables, e.g. ‘activity counts’ are provided for researcher use as ‘minimally processed data’.

The intermediate outputs (e.g. ‘activity counts’) of accelerometers can differ between manufacturers. Such variation likely explains consumer reports of step count output differences between devices.

Accelerometer data is informative of many parameters beyond sleep/wake state that are relevant to health. For example, body position and movement patterns of particular physical activities (walking, swimming, etc.), movement timing and intensity, gait metrics (symmetry, stride length, balance), and estimated energy expenditure. Recently, high-resolution accelerometry data has been used to infer pulse rate and respiration that can be fed into sleep staging algorithms. Additionally, one manufacturer has obtained FDA clearance to determine the risk of moderate to severe OSA using accelerometry data alone [45].

Practical Notes:

There are two important challenges to generalizing sleep detection using motion alone: 1. People who lie very still while awake. 2. Age (and other) differences in motion while sleeping.

A bed-partner potentially contributes to otherwise unexplained detected motion (and wakefulness).

Accessing 'raw' accelerometry data is often motivated by the desire to use existing open-source algorithms to provide sleep assessments for specific contexts. For example, age (children move more when asleep), sleep disorders, e.g. insomnia (patients may lie still while awake), obesity, etc. However, as stated earlier, most users will only utilize data processed into understandable endpoints.

3.1.2 Photoplethysmography (PPG)

Heart rate (HR) slowing and increases in heart rate variability (HRV) accompany the onset of sleep and shift during different stages of sleep. PPG sensors are ubiquitous in CHTs and allow for the measurement of pulse rate (PR) and pulse rate variability (PRV) as a proxy for HR and HRV [46], which contributes to sleep staging [39, 40].

PPG uses light reflection or transmission (or both) to capture changes in blood volume during systole and diastole. The intervals between consecutive peaks in the PPG signal are used to measure pulse rate and pulse rate variability to infer HR and HRV. PPG can be obtained at different wavelengths via photodiodes. The wavelength affects skin penetration depth and interaction with hemoglobin. Green and infrared (IR) light are used to infer HRV and respiration rate (RR), while red and IR light are used for oxygen saturation measurement. The quality of PPG signals can vary significantly across devices [47].

The PPG waveform is processed directly in the device's firmware. Raw PPG signal is not typically accessible to the user and may not even be stored on the device. Proprietary signal processing details are very important for inference, for example, how movement and skipped heartbeats are handled.

Signals may be combined to counteract artifacts and for more sophisticated applications assessing the PPG pulse waveform to estimate, e.g. arterial stiffness and blood pressure.

Sensors at the finger can achieve higher signal quality than at the wrist [48], although this comes at the expense of dealing with more motion artifacts during wakefulness. HRV derived from PPG, combined with movement data, are the signals most commonly used for inferring sleep stages. Algorithms are based on differences in heart rate and breathing regularity between REM and NREM sleep but with deep learning methods, the specifics of which parts of the signal are used are not transparent [49, 50].

Practical notes:

PPG signal is highly susceptible to artifacts (e.g., wrist/arm movements) and, typically, can only be considered reliable under conditions of proper skin contact with no external change in pressure applied to the tissue and limited or no movement (such as when one is asleep). This is especially important for HRV analysis or applications using the pulse waveform [51]. Proper wear of the device for sufficient time is essential for acquiring high-quality data.

HRV analysis requires a continuous (uninterrupted) high-quality data stream over several minutes. If pulse rate data is missing or invalid, HRV measurement is compromised [51]. Quality metrics (invisible to the consumer) are generated to detect and censor problematic signal segments to address this.

Importantly, signal processing cannot effectively recover poor-quality data associated with poor skin contact.

PPG signals can be degraded if the user has a condition that affects peripheral arteries. Even in the absence of disease, age, obesity/sarcopenia can affect the signal.

Skin tone can affect pulse oximetry. It is less clear how it affects other uses of the PPG signal [52]. Some manufacturers are actively working to address these issues.

3.1.3 Temperature

Distal skin temperature rises prior to sleep onset, followed by a drop in core body temperature, which can be used to detect sleep onset [53]. Distal and core body temperature changes often move in opposite directions. Thermoregulation is also less efficient during REM sleep than NREM sleep [54]. Distal skin temperature changes can thus assist with sleep onset detection and sleep staging. Temperature monitoring has other uses, such as fertility cycle tracking [19-21] or early detection of febrile infections.

Thermal sensing is most commonly based on thermistors, which register a reduction in resistance in response to an increase in temperature.

Practical Note:

Wearables usually report the temperature as relative changes with respect to a person's baseline data. Spurious readings not due to illness could result from sleeping in a new environment or an unusually cold or warm night in the habitual sleeping environment or hand position relative to one's body or blanket.

3.1.4 Electrodermal activity (EDA) sensors

These sensors measure changes in skin conductance, e.g. due to sweat secretion [55], and are not widely adopted presently. Sweating occurs to cool the core (thermoregulatory function) when one enters deep sleep, ceases during REM, and is low during light sleep. A parallel set of changes in fluctuation (frequency) of conductance (electrodermal activity; EDA) follows this pattern, with deep sleep showing the highest frequency and REM the lowest [56, 57]. Although there is a correlation between distal skin temperature change and EDA, several studies indicate that the two modalities provide independent information.

3.2 Fundamental Sleep Measurements: Tips on Usage and Caveats

This section provides definitions of fundamental sleep-related measures used and notes related to their measurement, as well as recommendations on using the data.

—- Fig. 2 Around here

3.2.1 Wearable vs. PSG measured sleep

Non-EEG wearables measure movement and autonomic signals that are modulated by sleep. They do not measure electrical signals related to thalamocortical activity on which much of our current knowledge about sleep is based.

However, the scale at which they can be deployed, their use over multiple nights in the field, and the multimodal information they gather, which can be integrated (e.g. light, phone usage, sound, ambient temperature, air quality), suggests that over time, they will yield rich information about sleep and the sleeping environment. Outside of sleep, they already collect useful health measures (e.g. HRV, physical activity), including those that can affect sleep.

Once a considerable knowledge base emerges, the characterization of sleep may shift to one based on non-EEG physiological signals or a hybrid of EEG and non-EEG physiological signals. Insofar as these can be synthesized to map to relevant health and wellbeing outcomes, a radical transformation of how we envision 'sleep measurement' may take place.

Recommendation

***Manufacturers*:** Fundamental sleep measures (FSM, see later) should be demarcated from more exploratory or proprietary measures. For professional users, manufacturers should provide information warranting the precision of their measurement in the aforesaid FSM. Characteristics of the dataset used to evaluate device performance in estimating FSM (i.e. the size of the test population, and included age, sex, comorbidities) would also be useful to provide. Such information could readily be posted and updated on the manufacturer's website

3.2.2 Bedtime and waketime

Bedtime, or TIB start time, refers to the time a person goes to bed. In laboratory studies, this designates when a person intends to initiate sleep and corresponds to the 'lights off' timing (Fig.2).

Waketime, or TIB end time, refers to the time a person gets out of bed in a laboratory study, corresponding to the 'lights on' timing.

In real life, many people enter bed prior to the time they intend to sleep and then engage in activities that involve minimal movement. Increasing the use of electronic devices in bed is one such activity. ANSI/CTA/NSF-2052.1-A [33] has defined a variable: the Time an individual is Attempting to Sleep (TATS Start Time), but this has not been used by manufacturers. Instead, 'bedtime' continues to be used while 'TATS Start Time' is implied. This is the most problematic aspect of sleep assessment in CHT sleep assessment and has ripple effects on the determination of sleep latency, sleep efficiency and has implications when using these devices in insomnia detection and treatment [58, 59].

A 'Research actigraph' usually comes with an event marker to signal an attempt to sleep. Researchers and clinicians often have individuals keep a sleep diary along with actigraphy recording. Outside the lab, accurate bedtime recording depends on the diligence and recall of participants who do not change their minds about their original intention to sleep.

As many people wake up to an alarm, wake time determination is often less problematic than determining bedtime/TATS Start Time. However, some individuals try to fall asleep after initial awakening before finally getting out of bed. Others awaken, either lying still or quietly using e-devices without moving significantly before finally increasing their activity and getting out of bed. This has led to the suggestion to determine TATS End Time and to specify Total TATS duration as a replacement for TIB to specify the total duration that an individual is in bed and is attempting to sleep, excluding time in bed not intending to sleep.

Automated methods to infer the intention to sleep vary in performance across devices, and as such, this measure is not currently standardized. Consumer devices often allow manual adjustment of bedtime and waketime.

Recommendations:

Users should be aware of differences in the accuracy of bedtime detection across manufacturers. This measure may be unreliable for persons with problems initiating sleep, who awaken earlier than desired, or who remain in bed engaged in quiet activity prior to or after the intended sleep period.

Users and practitioners should note that measurement errors notwithstanding, trend information over weeks could still help inform about sleep timing and encourage discussions on preferred vs. realized sleep timing. This can address bedtime planning issues, avoiding bedtime procrastination, and sleep hygiene.

Manufacturers should consider installing an indication system to improve the performance of this parameter. This could include a mechanical button, a signal through the app, ambient light detection, or sound detection. They should consider adding a journaling/tagging feature to enhance documentation of nights with prolonged sleep latency or mornings where participants spend significant time in bed post awakening.

3.2.3 Time in bed (TIB).

TIB is defined as the time between bedtime and waketime. It is affected by the same limitations as their constituent measures, bedtime, and waketime (Fig.2).

Recommendations:

All: TIB should be only referred to and used when bedtime and wake-up time are self-reported or manually signaled by the users. TIB should be distinguished from 'sleep period'.

Where available, TIB information can inform about the gap between entering bed and intending to sleep related to e-device use, for example.

Manufacturers: TATS Total Duration, should replace TIB eventually by combining multisensor data, including using those on the phone to determine it as automatedly as possible.

3.2.4 Sleep period, Sleep onset time, Sleep offset time

Sleep period is the interval between the initial sleep onset and the last sleep-to-wake transition, as determined by the device (Fig.2). Sleep onset time is the commencement of the first epoch of detected sleep. Sleep offset time is the time of the end of the last epoch of detected sleep in the main sleep period.

Sleep period is *not* equivalent to TIB. The latter is a self-report construct of the time period when sleep is intended, which is difficult to ascertain automatically. As such, the sleep period is more likely to be accurately determined.

Devices often require a minimum threshold of continuous sleep (often >3h) to designate a sleep bout. As such, a person with multiple shorter but highly fragmented sleep separated by long (>~1hr) wake periods may not have a meaningful TST value for that night's sleep. Instead, the record might show multiple sleep periods that night.

Recommendations:

All: For CHT, sleep period is preferred to TIB as a measure as it is objectively defined. Accordingly, sleep onset and sleep offset time should be used in dashboards as these are devices that accurately measure sleep.

Clinicians: Sleep midpoint is the time halfway between sleep onset and offset times. Appraisal of sleep midpoint that includes days with and without external occupational, educational, and family responsibilities can provide insight into chronotype, presence of social jet lag, and may suggest a circadian rhythm sleep wake disorder in the context of symptoms.

Manufacturers should consider including 'sleep period' in device readouts.

3.2.5 Total sleep time (TST).

The sum of all epochs classified as sleep within a TIB or Sleep Period (Fig.2).

TST has the most consistently agreed upon definition, greatest 'accuracy', and least variation in accuracy across different devices.

Devices often have a minimum threshold of continuous sleep (often >3h) to have this measure recorded. Persons with long periods of wakefulness (~> 30-60 minutes) after sleep onset may have records that do not summate all sleep times.

Recommendations:

Users should be aware that the accuracy of TST measurement is affected by the quality of the manufacturer's algorithm. They are most accurate for healthy working-age adults. State of physical health, conditions that affect mobility, obesity, and age (either children or old adults) can affect TST assessment.

Users should note that on nights with sleep lower than 80-85%, the accuracy of TST may be compromised [12].

Users and clinicians encountering a night(s) with apparently no sleep should exclude non-wear or incorrect wear and inspect the record for highly fragmented sleep.

Manufacturers should work to improve the sleep measurement in shift workers/persons who have a 'main sleep period' broken up or to allow manual editing of sleep periods in such situations. Non-wear and out of battery periods should be clearly labeled to avoid ambiguity.

3.2.6 Naps

These are periods of sleep outside the main (usually) nocturnal sleep period. Voluntary and involuntary sleep episodes are important to distinguish.

Naps are increasingly being built in as features in wearables, but their consistent detection and separation from periods of quiet wakefulness following physically active periods remain a technical challenge, and further research is needed to improve these. In view of this, some manufacturers allow users to edit sleep periods detected as naps.

Currently, the shortest known disclosed interval for nap detection ranges from 1 Minute (Garmin), 15 minutes (Oura) to 1 hour (Fitbit). This information is placed here for indicative purposes only and can be accessed through the manufacturer's webpages that are periodically updated.

Recommendations:

Users and clinicians: Nap detection varies in performance across devices [60]. Improvements in nap detection (e.g. incorporation of ambient light and body position as additional channels) are needed before this feature can be reliably used. Automated nap detection should be used in conjunction with journaling/logging features where available to delineate voluntary and involuntary napping. These cannot be automatically discerned.

Involuntary sleep during non-nocturnal or habitual periods can have negative health connotations [61], making it essential to distinguish these from deliberate/planned naps.

Manufacturers should provide the option to manually enter/label naps.

3.2.7 Sleep latency (SL)

This is the interval from the subjectively reported time when a person starts trying to fall asleep to the first epoch of sleep (sleep onset; Fig.2).

As this represents the combination of a subjective (TATS) and an objective measure, with the former being estimated, SL is challenging to assess accurately.

Over multiple nights, combined with the user's journaling or logging efforts, automated SL values may be useful in assessing excessively long or short sleep latency.

Recommendation:

All: Of the sleep variables listed here, SL is probably the one with the greatest difference between more and less accurate devices. Only relative values should be used for assessing trends (see notes on Bedtime assessment).

3.2.8 Wake after sleep onset (WASO)

We recommend defining this as the total wake time *within* the sleep period (Fig.2) which differs from the way AASM currently defines WASO.

Note this definition excludes the time a person may be spending in bed after the last recorded sleep but is trying to fall back to sleep but cannot. This time may be important to determine in persons with insomnia, but it cannot be automatically distinguished from one where someone awakens and lies relatively quietly in bed without the intention to sleep.

No PSG equivalent exists for 'brief arousal/awakening' from sleep. These contribute to an 'awakening' count, which could be a measure of sleep continuity, but no agreed upon norms exist.

The temporal correlation of WASO with PSG is dependent on how well periods of motionless wake are detected, balanced against excessive sensitivity to motion which will result in lower estimation of lower sleep time.

The optimal threshold for movement detection differs with age. Adolescents and children move more during sleep, leading to underestimating sleep duration. Body movement and change in sleeping position decline with age in relatively healthy people. In general, women move less than men, and obese people might move more during sleep.

Recommendation:

All: Age-related differences in movement patterns lead to a tendency to overestimate WASO in young and underestimate it in older adults. As older adults move less in general and the prevalence of insomnia rises with age, the lack of sensitivity of CHT to periods of motionless wake must be recognized, affecting the accuracy of sleep/wake detection in opposite directions in young and older adults.

3.2.9 Sleep Efficiency (SE)

SE is the ratio of TST to the total time with an intention to sleep: SE = TST / (SL + WASO +TST).

Recommendations:

Users and clinicians: The comparability of SE across devices depends on how accurately SL and WASO are determined. As a guide, on nights with SE lower than 80-85%, the accuracy of sleep measures is likely compromised (see above).

Users: Clinical advice should be sought if SE lower than 80-85% is persistent (e.g., >3 nights/week over several weeks).

3.2.10 Time spent in 'light', 'deep', REM sleep.

'Light sleep' is usually equivalent to polysomnography (PSG) N1 + N2 sleep, while 'deep sleep' is the proxy measure of PSG N3 sleep.

The ability of CHT derived sleep staging to approximate PSG is affected by signal quality and the data transformation algorithm.

In many devices, only sleep/wake discrimination is provided when the PPG signal is lost due to poor skin contact, excessive artifact from other reasons, or not acquired due to power conservation (low battery).

Recommendations:

Users and clinicians: Specialists disagree about using these indicators to assess trends in research settings. However, they are reasonably reliable in healthy participants at time intervals of several weeks. However, caution should be exercised when interpreting these data for other use cases.

3.2.11 Sleep Regularity

There are several parametric and non-parametric methods to calculate the regularity of sleep and wake, but three are most widely accepted:

(1) Standard deviation of the onset, offset, or midpoint of the main sleep period over a period (usually one week); additionally, the standard deviation of sleep duration over a similar interval may be computed.

(2) Interday stability is the normalized 24-hour value of a periodogram of movement with values from 0 to 1 and 1 being perfect alignment across days:

$$\mathrm{IS} = \frac{t \sum_{h=1}^{p} (\overline{x_h} - \bar{x})^2}{p \sum_{i=1}^{t} (x_i - \bar{x})^2}$$

where t: the total number of data points, p: the total number of data points per day, x_h: hourly means, x: the mean of all data points, xi: is individual data point from the activity time series.

(3) Sleep Regularity Index: the average likelihood (probability) of an individual being in the same state (wake or sleep) at any two time points 24 hours apart:

$$SRI = \mathrm{N}\sum_{t=1}^{N} |X_t - X_{t-24}| \times 100$$

where X_t is the binary sleep-wake state at time t (1 if asleep, 0 if awake); X_{t-24} is the binary sleep-wake state 24 hours prior to t. N is the total number of hours considered for the calculation.

Recommendations:

Clinicians: As no broadly accepted 'gold standard' for these derived measures currently exists, no quantitative clinical recommendation is made. Sleep regularity can be assessed over a week or a month; robust estimations require more than the limited number of days used in some existing studies[62].

Manufacturers should state what they measure if the data field 'regularity' is provided in their dashboard. Even without a formal indicator of sleep regularity, an intuitive representation of this construct may be obtained from viewing a series of bar graphs showing sleep timing over successive nights over weeks.

3.2.12 Sleep satisfaction

Self-perception of the overall quality of the previous night's sleep.

There is no standard for assessing sleep quality. Many researchers mix objective and subjective measures. We recommend the use of 'sleep satisfaction' to denote the latter.

EEG measures of N3 and REM do not clearly map onto sleep satisfaction [63]. Although N3 is associated with multiple neurocognitive benefits, waking up from N3 early in the night does not correspond to feeling 'rested'. In contrast, waking up from REM, which occurs more in the second half of the night, has a higher probability of this result [64].

Recommendations:

Users and clinicians: Some manufacturers have already provided tagging/journaling. This feature should be encouraged to facilitate the discovery of which personal behaviors and environmental exposures result in better subjectively and objectively measured sleep.

Manufacturers: Because simply logging in to an app is necessary to synchronize data, providing an option to tag / journal / log sleep satisfaction will be useful for long-term assessment of how a combination of wearable-derived physiological data might inform this.

3.2.13 Hypnogram.

This overview of sleep staging over the sleep period is commonly provided by wearables. It is most often based on 30s resolution output derived from manufacturer-specific integration of multimodal data (motion, HRV temperature‡, circadian factor‡, EDA‡).

‡when used

This data forms the basis for epoch-by-epoch comparisons with PSG. The displayed data is typically at a lower time resolution than obtainable from the device API, i.e. it is significantly smoothed.

Recommendations:

Users and clinicians: Be aware that short or more frequent state transitions may be present, some captured by the device but not displayed. This applies to wake periods during sleep as well. Short napping periods during daytime are often not detected [60]. Percentages of sleep stages displayed may also differ considerably from sleep stages recorded using PSG.

Manufacturers: Consider appending a 'quality' or 'confidence in data quality' metric to indicate to the user/clinician the reliability of the sleep assessment on a given night. They should indicate the temporal resolution of the data plotted and how much smoothing is applied in their displays (which can differ between smartphone and API outputs).

3.2.14 Manufacturer 'Sleep Score'

These are proprietary measures that provide an overall grading of a night of sleep in a single number.

This summary metric follows the motivation underlying the Pittsburgh Sleep Quality Index (PSQI) [65], the Sleep Hygiene Index and other questionnaire-based, aggregate sleep metrics. These attempts to reduce the multiple dimensions of sleep to a single summary measure.

While not known to be 'scientifically validated' at present, these non-standardized measures often use the same variables: sleep timing, duration, continuity, and efficiency that clinicians and scientists measure. Thus, conceptually, using a proprietary 'sleep score' is no different from widely used 'sleep quality indices' such as the PSQI or Sleep Health Index [66], that seek to reduce the multidimensional construct of sleep to a single measure. Large manufacturers could potentially better calibrate such sleep scores with a user base of millions than what has been achieved with existing indices.

Recommendations:

Users and clinicians: While absolute scores may currently not be comparable across manufacturers and represent a gross simplification of the multidimensional nature of sleep, they are more likely to engage the user than a long list of sleep variables. If the constituent measurements listed above have been properly evaluated with sufficient numbers of participants of age, sex distribution, and other

characteristics that represent the target market, and if the output measures meet acceptable standards, such measures should be considered useful for advising participants on score trends. Using such trend data may get the user to reflect on what affects their scores and how to make positive lifestyle changes.

All: This feature could be considerably strengthened if the calculation methods were transparent and standardized across products. Long term health outcome data could be integrated into sleep score interpretation, for instance, if consistently achieving a "sleep score" of 60 is linked to a higher risk of developing hypertension or metabolic syndrome than a score of 70, manufacturers could create disease risk models tailored to their devices. These models could then be evaluated and compared for predictive accuracy. At a minimum, this approach could serve as a foundation for users and clinicians to receive health advice independent of the specific device. While each manufacturer employs proprietary sensors, algorithms, and outputs, the health outcomes these scores aim to predict are, or can be, standardized (e.g. see AASM/SRS Discussion [67].)

3.2.15 Resting Heart Rate (RHR)

RHR is measured in the absence of physical activity. Historically it has been measured by palpation or using ECG. PPG-based methods allow for convenient, continuous monitoring, including during sleep. Technical considerations in PPG-based measurement of RHR and HRV were dealt with in an earlier section.

Measuring RHR during sleep significantly reduces the likelihood of encountering movement artifacts and standardizes posture. If the wearable is not worn during sleep, a motionless period during the day is used to assess this variable.

RHR is a well-established indicator of overall cardiovascular health. In general, a high RHR at a single time point predicts poorer cardiovascular health and higher mortality [68]. A change of ≥ 5 bpm in RHR over 5 years is a significant marker of poorer cardiac outcomes [69]. However, substantial interindividual variation exists, and it is recommended that trend data be evaluated since multiple factors can affect baseline RHR.

Women have a higher RHR than men. Chronic effects include cardiovascular fitness, cardiac function, obesity, smoking, diabetes mellitus, and medication intake. Medium-term (days) effects include seasonality, menstrual cycle, and a temporary illness. Psychosocial stress and physical fatigue can affect RHR at different time scales. Delayed bedtime [70] and alcohol intake [71] elevate HR; medications that affect the heart, in particular beta-blockers or stimulants, can affect RHR. A variety of non-cardiac illnesses can affect RHR, e.g. thyrotoxicosis.

In addition to day-to-day, week, and month-level changes in RHR, nocturnal temporal features can be informative. A delayed nadir may indicate circadian misalignments, such as those caused by shift work,

jet lag, or delayed sleep-wake phase disorder (DSWPD). Lack of a clear nadir may suggest poor sleep quality, insomnia, or a condition that affects autonomic balance, such as sleep apnea.

Recommendations:

Users and clinicians: For reliable RHR measurement, the device has to be consistently appropriately worn. RHR varies significantly across individuals, and the factors listed above that influence it within an individual need to be considered for meaningful interpretation. Simplistically, changes of > 5 bpm from day to day or over a week are not artifactual and merit attention.

Manufacturers: Specify whether RHR is measured when a person is asleep or during quiet wakefulness outside the sleep period. They should indicate measurement fidelity in healthy persons over a defined time interval to allow better between-device comparisons.

3.2.16 Heart rate variability (HRV)

HRV, in the case of wearable, refers to the variation of the interbeat interval measured over several (typically 5 minute) intervals [48]. Broadly, a higher HRV is favorable, and a lower HRV is associated with adverse health consequences. HRV varies greatly between individuals and is strongly age-dependent. Many factors cloud the simple interpretation of single readings [72].

Various metrics with different interpretations exist to quantify HRV [73]. Wearables most commonly report SDNN, the standard deviation of NN intervals (the time interval between successive quality assured pulse wave peaks, excluding arrhythmias, ectopic beats or artifactual waves) or the root-mean-square of (censored) beat-to-beat differences in interbeat intervals (RMSSD).

Longitudinal tracking of HRV has two broad uses: determination of optimal intensity of physical training and cardiovascular risk assessment.

Recommendations:

Users and clinicians: nocturnal HRV measurements are preferred as they are less susceptible to artifacts. For accurate measurements, the device has to be worn snugly to the skin and in a manufacturer-recommended orientation.

Frequent gaps in HRV measurement during rest or sleep, once improper wear is excluded, can indicate arrhythmia, frequent ectopics, and significant fluctuation of PPG amplitude (which has negative cardiovascular connotations); see also below.

As with RHR, interindividual variation and many intrinsic and external factors can affect HRV measurements. Trend analysis with personalized consideration of these values is necessary to interpret them appropriately.

Manufacturers:

State when data for HRV is collected (day, night, both), how HRV is calculated, and the time window used for analysis. Minimally, they should provide a quality metric for the HRV measure for the day as well as general accuracy metrics of their interbeat interval assessment algorithm using ECG as ground truth.

Off-wrist detection is currently used to protect users from fraudulent commercial use of the wearable for making purchases. This can be added to a log so that clinical / research users can determine wear time for health tracking.

3.2.17 Respiratory Rate (RR)

RR refers to the number of breaths taken per minute. Traditionally, RR is measured manually or using specialized clinical equipment, such as capnography or respiratory inductance plethysmography (RIP), which tracks the movement of the chest and abdomen. Wearables typically derive RR by tracking the PPG waveform amplitude modulation or the variations in heart rate during inhalation and exhalation. Hence, the technical limitations of PPG measurement discussed earlier also apply here. Additionally, some wearables employ accelerometry to capture subtle chest movements at the wrist although the accuracy may vary depending on the body/hand position.

RR is a vital indicator of respiratory and overall health [74, 75]. Deviations from typical RR patterns can signify respiratory illness, stress, or sleep disorders. While individual baseline RR varies, longitudinal tracking provides valuable insights, such as identifying early signs of illness or assessing recovery trends following an acute condition.

Recommendations:

Users and clinicians: As with HRV, nocturnal RR measurements are preferred, as movement artifacts are less likely to affect them. The device should be worn snugly against the skin and positioned per the manufacturer's guidelines to help improve accuracy. Consistent deviations in RR or frequent gaps in data collection may indicate poor device fit, improper use, or potential health concerns such as irregular breathing patterns, sleep-disordered breathing, or significant motion artifacts. Further investigation is warranted in such cases.

***Manufacturers*:** Specify how RR is calculated (e.g., from HRV via PPG, accelerometry, or a combination of methods), the measurement time interval (e.g. using 30 s windows) and whether the data is collected during sleep, daytime rest, or both. They should also offer users a signal quality indicator to help identify periods of low data reliability and consider including trend analysis tools within wearable interfaces to help monitor longitudinal changes in RR. Further, the measurement accuracy of RR outside the normal range, i.e. clinically relevant situations, should be stated.

3.2.18 Blood Oxygen Saturation

Blood oxygen saturation represents the percentage of hemoglobin in the blood saturated with oxygen. In CHT, oxygen saturation from the periphery (as opposed to an artery, hence the term SpO2) is estimated by analyzing the differential absorption of red and infrared light reflected into the PPG sensor.

SpO2 is an important indicator of respiratory and cardiovascular health. Normal SpO2 levels typically range from 95% to 100% at sea level. Persistent readings below 90% indicate hypoxemia. Short-term fluctuations may result from positional changes, transient hypoventilation, or sensor artifacts. It is important to note that SpO2 levels can be influenced by factors such as altitude [76], skin tone [77], peripheral perfusion [78], and device placement. Persons with darker skin tones might have lower blood oxygen saturation than indicated.

Recommendations:

Users and clinicians: For reliable SpO2 measurement, the device should be worn consistently, snugly, and according to the manufacturer's guidelines. Measurements during sleep are preferred to reduce motion artifacts. Significant deviations from baseline SpO2 levels, particularly persistent readings below 90% or frequent nocturnal desaturation events, warrant medical evaluation. However, given low sampling rate, normal values are not confirmatory for the absence of sleep disordered breathing. Users should also be aware of environmental factors, such as high altitude when interpreting SpO2 values.

Manufacturers:

Specify how the measurement is performed (e.g., continuously throughout sleep or in intervals), and the accuracy.

3.2.19 Breathing disturbance

CHTs often provide a qualitative or semi-quantitative assessment or numerical scores to assess breathing disturbance during sleep, which may indicate the risk, presence, and severity of sleep apnea. While quantitative breathing disturbance assessment via CHT may be potentially useful in the future, current definitions and measurement principles for assessing breathing disturbance vary broadly, and clinical validation is limited.

Recommendations:

Users and clinicians: If CHTs consistently detect frequent breathing disturbance and patients report related symptoms a sleep apnea test is warranted (see section 4.5). A notification of high sleep apnea risk from an FDA cleared application associated with a CHT warrants medical attention.

Manufacturers: Clearly define the breathing disturbance metric and validate against clinically used markers such as the AHI.

4.0. PRACTICAL RECOMMENDATIONS FOR USING SLEEP TRACKER DATA

4.1. Use *in Healthy Adults*

Healthy individuals, for the purpose of this section, are defined as persons with no known chronic health conditions, including sleep disorders. The devices may be used by consumers to support long-term sleep health monitoring for self-discovery, performance enhancement, or preventive health.

Recommendations for users and clinicians:

1. Physicians and users should note that current US National Sleep Foundation guidelines [14] about sleep duration are derived from a mixture of 'time in bed' and 'total sleep time' data, often self-reported, with greater weightage placed on the former [15]. Systematic work on adolescents with actigraphy indicates that objectively determined sleep duration is substantially lower than questionnaire-assessed measures [79]. Arising from this, there should be a ~5-20% discount in measured TST compared to these recommendations.
2. For other measures this panel deemed 'fundamental,' there is less clear guidance on what is appropriate. This is an active area in which the systematic collection of good quality wearable sleep tracker data can contribute to formulating. We provide a table for consideration that is a placeholder for current thinking.
3. Physicians and users should be aware that these devices have mostly been trained on healthy adult data [39, 40, 80], but see Olsen [81], within a limited age range. The measurement fidelity as well as values of inferred sleep measurements, are affected by age, obesity, physical and mental health, poor sleep, medications, and substance use.
4. Normative data on other sleep characteristics is not well-established but is evolving. For example, we encourage people to maintain regular sleep patterns, but quantitative standards are unavailable. Discussions on this topic are found elsewhere [18].
5. Enthusiastic users should be discouraged from using 'sleep scores' to compare with their or others' sleep.
6. For individuals where sleep tracking contributes to sleep anxiety (i.e., orthosomnia[82]), consider taking a break from sleep tracking or discontinuing. Education about the validity and use of sleep tracking metrics may reduce sleep tracking anxiety.
7. Individuals should be encouraged to review 1-week averages of sleep measures rather than single nights and educated on the acute effects of sleep loss vs. chronic exposure (e.g., acute

sleep loss affects performance/safety, but health risks are linked to chronic, not acute exposure).

8. The effects of alcohol or substance use on elevated nocturnal heart rate and sleep should be discussed to encourage healthier behavior (e.g., controlling alcohol intake, reducing late-night eating, and avoiding recreational drug use).
9. Weekend-weekday differences in sleep duration/quality should be discussed as it has health implications. Although an increase from 'medium' to 'long' sleep duration on the weekend is unfavorable [83], short weekday-short weekend sleep duration appears to be associated with poorer outcomes than a modest weekend-weekday difference [83-85].
10. Consistently late sleep timing is emerging as a risk indicator for both physical [86] and mental [87] illness.
11. Concern over minutes or percentages of SWS/deep sleep or REM sleep must be tempered. Users should be reminded that sleep staging values differ in quality across devices and are meant to be indicative only and are not authoritative. While reduced SWS/deep sleep in older adults is regarded as a risk factor for late-life cognitive decline and risk of dementia [88], data is inconsistent [89, 90], and there is presently no robust evidence that artificially increasing SWS/Deep Sleep alters outcomes. REM sleep is considered important for emotional wellbeing, but apart from reducing or eliminating intake of medications that suppress REM sleep when feasible, there is no evidence that remedying REM sleep 'insufficiency' has therapeutic benefit. Notably, excessive REM sleep may be present in persons with mood disorders [91]. REM may be suppressed in individuals taking medications (e.g. selective serotonin reuptake inhibitors/SSRIs).
12. If measured, significant changes in oxygen saturation should be followed up with a health professional. Device quality and wear need to be checked to reduce false alarms. Persons with darker skin tones might have lower blood oxygen saturation than indicated. Readings may be inaccurate if sampling is not sufficiently frequent.
13. Detected atrial fibrillation (where this feature is included and certified) and/or persistent gaps in heart rate data should be followed up with a health professional.
14. Integration of sleep and physical activity timing can be the basis for discussing how the user can plan adequate time for exercise [92].
15. Naps are presently not consistently detected in these devices. The minimum threshold for nap detection is presently device-dependent. Some manufacturers do not warrant nap detection. Napping is subject to cultural biases. Naps are not divided into voluntary and involuntary sleep episodes. This is important to note.
16. Following travel-related sleep disruption, shift work, or polyphasic sleep, the partitioning of sleep into 'main', 'subsidiary' or 'nap' sleep periods is manufacturer dependent. Device output is less reliable and should be dealt with on a case-by-case basis. Caution should be

taken in interpreting sleep measurements in this context, and corroboration with other measures, such as self-report, is useful. Traveling across time zones can interfere with the assignment of correct timing to sleep and wake periods.

17. Challenges to mental health and wellbeing may be present when a sudden shift in previously consistent sleep behavior is observed.
18. Wherever possible, sleep data should be integrated with other contextually relevant health information, like the influencers of sleep and the consequences of different sleep patterns.
19. There is tremendous potential to characterize and norm longitudinal sleep measurement with unprecedented levels of diversity and resolution if these are collected from sufficiently accurate devices worldwide [93].

4.2. *Sleep Tracking in Individuals Presenting with Sleep Symptoms*

A considerable proportion of the global community of CHT users are individuals who report chronic sleep disturbances or are at considerable risk of developing sleep disorders. This population tends to use CHTs to seek answers about their sleep difficulties.

These individuals may benefit from long-term sleep health monitoring similar to that provided by CHTs to healthy individuals. By providing long-term sleep trends, users may be able to identify potentially adverse sleep symptoms. Some wearable devices provide accurate information about breathing during sleep. These devices can be used as screening tools for sleep-disordered breathing. This usage mode can help people proactively seek sleep disorder treatment (see later). Sleep researchers are beginning to use CHTs as an additional screen for sleep disorders in research participants, following the decades-old practice of using research-grade actigraphy. Additionally, users may also use these devices to identify disordered sleeping timing, such as unusually delayed or advanced sleep patterns that occur with circadian rhythm sleep-wake disorders. However, it is important to note that according to the AASM guidelines, sleep trackers should not be used in isolation to diagnose sleep disorders [30].

The longitudinal use of CHTs also supports the early identification of disordered sleep, and, in selected cases, may provide the opportunity to monitor treatment outcomes. While there are appropriate concerns about the potential for ‘false positives’ or ‘false negatives’ from such use cases and the unnecessary worry that this may generate, the net benefit of widespread identification of individuals at high risk for sleep disorders is likely positive.

CHTs may also facilitate education about the importance and impacts of sleep health. For example, individuals may better understand the influence of sleep disordered breathing by viewing cardiovascular and other physiological signals (e.g. oxygen saturation) during sleep on untreated nights compared to nights with treatment. Some consumer apps also flag when a poor night of sleep has occurred, such as

sleeping for <6hr. This can encourage users to take proactive measures to mitigate any short-term risks (e.g., driving while sleep restricted) and seek to rectify their sleep disorder symptoms if they persist.

4.3. *Sleep Tracking in Individuals with Chronic Sleep Disorders*

Sleep tracking may be combined with existing diagnostic and therapeutic tools to engage and support patients with confirmed sleep disorders. However, they cannot be used for diagnostic or therapeutic purposes when deployed in isolation. The usefulness of CHT will differ depending on the specific condition. There is inadequate evidence to support clear recommendations in some groups, for example, those with hypersomnolence or periodic limb movement disorder.

4.4 Insomnia

While self-reporting is central to insomnia assessment [94], CHT can provide informative, objective data to support clinical management, possibly replacing research actigraphy devices that have been used for decades and using them in combination with sleep diaries. CHT-derived sleep timing and regularity may help exclude a circadian rhythm sleep-wake disorder. Further, insomnia is often comorbid with OSA in about 30 to 40% of insomnia cases [95]. High-quality consumer trackers, whose performance has been scientifically evaluated, can be useful for screening for sleep-disordered breathing in such cases.

While many persons may benefit from greater awareness of their sleep patterns and how these affect next-day function and/or may be a believer in the cultural phenomenon of self-tracking with technology (i.e., quantified self), some may develop unhealthy fixations (‘orthosomnia’ [82]) about attaining particular sleep 'benchmarks' they believe to be essential for their wellbeing. Anxieties about health data are not unique to sleep - for example, anxiety about not meeting a weight, blood-sugar control, or resting heart rate metric. Such anxieties should be recognized, and discontinuation of CHT may be recommended.

CHT can facilitate the administration of new treatment techniques. For example, Intensive Sleep Retraining, which was once confined to the laboratory. It involves allowing patients to fall asleep before being woken up after only 1-2 minutes of sleep, repeatedly for up to 24 hours [96]. Deprived of recuperative sleep, sleep pressure increases, and patients fall asleep more rapidly across successive sleep attempts. This approach helps recondition insomnia patients to fall asleep more quickly after the procedure, the efficacy of which is demonstrated in clinical trials [96, 97]. A consumer tracker has enabled the routine administration of this efficacious yet otherwise impractical technique outside of the laboratory environment, demonstrating the utility of consumer technologies in providing novel treatments for insomnia.

A few small studies have considered how CHTs may support the implementation of established treatments, such as cognitive behavioral therapy for insomnia (CBT-I). In a small study in which some patients had worn a CHT combined with CBT for insomnia, patients who wore trackers had higher adherence, potentially via higher user engagement [98]. Some studies have tested whether wearable devices as feedback can be therapeutic, particularly for people with significant sleep-wake state discrepancy (i.e., sleep misperception), and findings have not supported their use for this purpose [99].

Recommendations:

Clinicians: If insomnia is suspected, supplementary information should be gathered with a sleep diary (e.g., the NIH National Heart, Lung, and Blood Institute Sleep Diary https://www.nhlbi.nih.gov/resources/sleep-diary) because sleep tracking is generally less accurate in people with insomnia than in healthy adults due to more prolonged periods of quiet wakefulness in bed. Accordingly, inferred sleep latency and brief wake periods may also be less accurate than for healthy sleepers.

With the above caveats in place, CHTs may help identify objective evidence of sleep disturbances (e.g. prolonged WASO), particularly among populations with difficulty reporting their sleep problems (e.g. patients who have difficulty keeping sleep logs, such as adolescents and individuals with memory deficits).

If there is evidence of frequent sleep disturbance on CHT (e.g., at least 3 times per week for at least 3 months in the case of chronic insomnia[100]), it may be useful to initiate a discussion to identify whether the patient has self-reported sleep complaints and would benefit from insomnia treatment (e.g., CBT-I). Patients may have objective sleep disturbance but, in some cases, do not complain about their sleep or desire treatment.

Sleep tracking can assist in engaging and motivating patients to adhere to treatment and some devices are useful for treatment administration.

Diagnosis of insomnia disorder does not require objective evidence of sleep disturbance on CHT. It may be helpful to discuss objective results with the patient, particularly in regard to educating patients on typical/expected sleep patterns for their age and gender. Patients with insomnia often have unrealistic expectations about sleep (e.g., sleeping 8 hours each night, not waking up in the night), and discussing their data may provide an opportunity to adjust maladaptive beliefs about their sleep.

Some patients will report markedly different sleep compared to the data from CHT (e.g., sleep-wake state discrepancy or sleep misperception). There is no standard treatment protocol for managing this phenomenon and no current evidence to support using devices in this situation.

Users and clinicians: CHT results can be useful to identify poor sleep health (e.g., irregular sleep, delayed sleep timing), and some devices can screen for sleep disordered breathing and circadian rhythm comorbidities, which should be corroborated with gold-standard testing.

4.5 Sleep Apnea

CHTs have the potential to assess sleep apnea. Here, we use the term "sleep apnea" instead of obstructive sleep apnea (OSA) since it is unknown whether CHTs can distinguish between the types of sleep apnea events (i.e., obstructive, central, mixed).

CHT embedded sensors record signals that may be useful in identifying sleep apnea. Given the proprietary nature of algorithms used to analyze physiological signals collected by CHT sensors, the inputs to CHT sleep apnea classifiers remain conjectural. Many devices feature pulse oximetry, which enables the detection of episodic oxygen desaturation (intermittent hypoxia), a hallmark of sleep apnea associated with adverse cardiovascular outcomes [101, 102]. However, to detect intermittent hypoxia associated with sleep apnea, a higher resolution (sampling every 1-2 seconds) is needed than what is provided by most devices (sampling every few minutes). In addition to oximetry, cardiac and respiratory parameters acquired by PPG as well as accelerometry, may be used to detect sleep apnea[103].

As of the end of 2024, the Apple Watch and Samsung Galaxy Watch have obtained FDA clearance for the identification of individuals at risk for moderate to severe sleep apnea who have not yet been diagnosed. Withings received EU approval for this function in early 2025. Apple has disclosed that its sleep apnea detection algorithm uses patterns in triaxial acceleration data to detect the presence of apneic events[45]. As of writing, there are limited reports of CHT performance for sleep apnea detection. No CHT device is FDA cleared for the diagnosis or longitudinal monitoring of OSA.

However, CHT will likely have a role in managing sleep apnea patients as multi-night ambulatory assessments that they can perform conveniently and at scale. As there is significant night-to-night variability in apnea severity [104] CHT could provide a better indication of disease severity than the current practice of single-night PSG studies. CHTs' ability to track a broad range of health-related metrics, such as physical activity and sleep patterns, which may interact with sleep apnea severity, may also engage patients in their personal evaluation and management of sleep disordered breathing[105].

Well-designed performance evaluation studies are needed to assess the ability of CHT to detect and accurately quantify the severity of sleep apnea. Such studies should evaluate the rate and sources of error, artifact, and data loss. Randomized controlled trials that implement CHTs in the evaluation and management of sleep apnea are needed to determine if incorporating CHTs into the sleep apnea care model improves relevant clinical outcomes such as time to diagnosis, cost efficiency, symptom control, quality of life, and sequelae.

Recommendations:

Users: When CHT that are FDA cleared for screening of moderate-to-severe sleep apnea reveal that individuals are at high risk, these individuals should seek evaluation for sleep apnea by a medical professional.

Clinicians: Individuals identified by CHT as high risk for sleep apnea should be evaluated with an FDA cleared home sleep apnea test or in-laboratory sleep study (polysomnography) at the healthcare provider's direction. A low risk use case for CHT in sleep apnea management is to leverage CHT as a patient engagement tool to promote adherence to sleep apnea treatment.

A systematic comparison of sleep apnea sensing modalities, signal processing methods and classification algorithms would facilitate the selection of the best approaches to incorporate for improved OSA detection and assessment.

4.6 Circadian rhythm sleep-wake disorders

A cornerstone of the evaluation and management of circadian rhythm sleep-wake disorders (CRSWD) is the estimation of sleep timing with the use of FDA-cleared actigraphy. In fact, the International Classification of Sleep Disorders, Third Edition, Text Revision recommends that actigraphy monitoring be used, whenever possible, for 7-14 days to demonstrate an advance, delay, irregular, or free-running pattern of sleep-wake timing to assist with the diagnosis of advanced sleep-wake phase disorder, delayed sleep-wake phase disorder, irregular sleep-wake rhythm disorder, and non-24-hour sleep-wake rhythm disorder respectively [106]. Additionally, a systematic review of actigraphy research by the AASM [107] identified that actigraphy yields significantly distinct information from sleep logs and can estimate sleep parameters with sufficient accuracy compared to PSG, leading to the recommendation that actigraphy should be used in the assessment of patients with circadian rhythm sleep-wake disorder. Extrapolating these recommendations, sleep timing derived from wearable health tracker sleep measurements could be used as supportive data to diagnose CRSWD.

While the most immediate use of wearable sleep tracking is to visualize abnormal patterns of sleep-wake timing in suspected CRSWD, mathematical modeling of 24-hour motion signals may be useful to quantify circadian properties of the rest-activity rhythm. Traditionally, cosinor analysis has been applied to actigraphy to estimate acrophase, mesor, period, and amplitude of the rest-activity rhythm [108]; however, this analysis may not be ideally suited to patterns that change over time, and a growing field of research using both physiologically-based [109] and non-parametric data-driven methods [110] has proposed algorithms to measure circadian rhythmicity. Similar to processing acceleration data from traditional actigraphy with these algorithms, acceleration data from off-the-shelf wearables can be

analyzed to predict dim light melatonin onset (the gold standard objective marker of the biological clock) within one hour of error in healthy individuals [111].

The multi-sensor properties of wearable health trackers allow for additional streams of data beyond motion (e.g. heart rate, temperature, and ambient light) to act as inputs to circadian estimation models. Already, CHT-acquired heart rate and temperature have been used to estimate the circadian phase in healthy participants [112, 113]. Given the difficulty in diagnosing CRSWD and the dependency of treatment timing on the objective circadian phase [114], wearable-derived estimates of the central circadian clock could transform clinical sleep medicine. This future capability is uniquely plausible with 24/7 data collection from wearable devices [113].

Therefore, wearable health trackers can augment clinical history in the diagnosis of CRSWD by allowing for visualization of the timing of sleep-wake patterns as well as providing estimates of the central circadian clock, although the latter requires further investigation in patients with CRSWD. Additionally, patient-centered outcomes should be assessed to determine the utility of incorporating such measures into clinical evaluation and CRSWD management. For example, research questions might address whether integrating wearable data reduces the time to diagnosis of CRSWD or improves response to treatment.

Recommendations:

Clinicians: Visually inspect and use sleep measured by CHT as an adjunct to the clinical history for diagnosing and managing CRSWD in the same way traditional actigraphy has been employed.

Further investigation of circadian rhythms estimates from wearable health trackers to enable use in clinical applications.

5.0. PERFORMANCE EVALUATION OF SLEEP MEASUREMENT

5.1 Framework for Performance Evaluation of CHT for Sleep Measurement

In 2021, Meghini and colleagues proposed a standardized framework for evaluating the performance of new sleep-tracking devices against a reference device, most commonly PSG, but sometimes involving a 'research actigraph'[35]. Performance evaluation typically occurs in the setting of an overnight PSG, either for investigation of the device(s) of interest or a clinical convenience sample of individuals undergoing PSG for diagnostic purposes while concurrently wearing a CHT.

Prior to evaluating performance, all device recordings must be temporally synchronized as closely as possible on an epoch-to-epoch basis, as device epoch lengths usually differ from the standard 30s PSG

epoch length. Alongside, as PSG differentiates NREM N1 and N2 sleep while these are collapsed into 'light' sleep by wearable algorithms, PSG epoch labels have to be adjusted.

Performance evaluation involves three sets of analyses. Discrepancy Analysis is performed to quantify the differences in duration between the device-derived sleep measures and PSG values. For this, individual epochs of each stage of sleep, both before (wake before sleep onset) and after sleep onset (wake after sleep onset, light sleep, deep sleep, and REM), are concatenated to assess individual-level and group-level differences in sleep stage classification.

Bland-Altman plots (Fig. 3A,3B) visually depict these differences and biases. Biases can be systematic - where there is consistent under or overestimation of a sleep parameter by the device as well as proportional, where the magnitude of the bias is affected by the true value of the measure. (e.g. when underestimation of TST is of a higher magnitude on nights/persons with shorter TST)

Epoch-by-epoch comparisons, the third analysis, provide a more granular assessment of the sleep tracker's accuracy (Fig. 3C) by comparing the agreement between the tracker and the reference method at the level of individual 30s epochs, e.g. via conventional confusion matrices and classification performance metrics such sensitivity, specificity and accuracy (Fig. 3D, Table 1).

For sleep-wake classification, specificity (which evaluates correctly classified awake epochs; alternate name: 'sensitivity to wake') is a more discriminatory measure than sensitivity or accuracy. This is because sensitivity (to sleep) and accuracy are typically very high (~90%) in healthy individuals who spend most of the allotted time in bed asleep [115].

Sleep stage classification performance may also be quantified by accuracy or preferably using Cohen's kappa statistic, which accounts for the possibility that agreement occurs by chance. The use of prevalence-adjusted bias-adjusted kappa (PABAK) may be beneficial given the imbalance of classes of sleep stages and wakefulness throughout the night is great.

— Fig. 3 Around here

Besides these standard analytical approaches, additional analyses can be used in a performance evaluation. For instance, equivalence testing analyses can assess whether the parameters derived from the PSG and the CHT are similar enough for practical and clinical purposes [43, 116]. The symmetric mean absolute percentage error (sMAPE), an extension of the classical MAPE analysis, can quantify the percentage error of the CHT predictions relative to PSG values, preventing extreme errors when values are close to zero (e.g., SL values) [117]. The Standardized Absolute Difference (SAD) can be used to compare the performance of two devices by measuring the absolute difference between their outputs, taking also into account potential differences in the magnitude of the values provided by the devices [117].

There is presently a growing number of well-conducted performance evaluation studies comparing different CHTs with PSG, surpassing the rigor of testing performed when actigraphy became widely used. However, expectations have also been raised. An accompanying comprehensive review of work to date will follow the publication of these recommendations, but the knowledge base is expected to continue to grow.

Table 1. Definition of specific terms used in the context of sleep tracker performance evaluation

Term	Definition in the context of sleep tracker performance evaluation
Synchronization	Precise temporal alignment of the CST and the PSG recordings, usually between lights off and lights on in laboratory studies.
2-Stage Classification	Scored epochs from both the CHT and the PSG are classified as wake (W) or sleep (S)
4-Stage Classification	Scored epochs from the CHT and the PSG are classified as wake, light sleep (N1+N2), deep sleep (N3), and REM sleep.
Bland-Altman Plot	A visual assessment of agreement between measurements derived from CHT and PSG, plotting the difference between the measurements against the PSG measurements. It can be used to assess systematic and proportional biases as well as adherence to limits of agreement
Systematic (Constant) Bias	The mean difference between a parameter derived from the test and reference devices (here, CHT and PSG). It indicates systematic over or underestimation of values across the measurement range.
Proportional Bias	The measure of how the bias of measurements obtained by the test and reference device varies according to the magnitude of the measurement. A negative proportional bias for TST would indicate that the CHT assesses sleep to be even shorter than PSG-measured TST as a participant's TST is lower than average for the sample.
Limits of agreement (LOA)	The limits of agreement estimate the interval that a given proportion of differences (typically 95%) between measurements is likely to lie within. The limits guide as to whether methods can be used interchangeably or if the new device can substitute for the reference (typically PSG) without changing the interpretation of the results. LOAs define the range within which most differences between sleep trackers and PSG are expected to lie. They are calculated as the mean difference (systematic bias) ± 1.96 times the standard deviation of the differences, covering approximately 95% of the differences under the

	assumption of normality (homoscedasticity).
Homoscedasticity	It refers to the condition where the differences between the sleep tracker and the PSG exhibit constant variability across the range of measured values. This indicates that the level of agreement between the tracker and the PSG does not depend on the magnitude of the measurements.
Heteroscedasticity	It refers to the condition where the variability of the differences between the sleep tracker and the PSG changes across the range of measured values. This suggests that the level of agreement between the tracker and the PSG depends on the magnitude of the measurements, indicating proportional bias.
Epoch-by-epoch analysis	EBE is a method used to assess the agreement between the classification (e.g., sleep or wake) provided by the sleep tracker and the PSG for each individual epoch.
Confusion matrix	A table used to evaluate the performance of a classification model. It compares the classifications (binary or categorical) provided by a sleep tracker against the classification provided by the reference (PSG), showing the number of correct (true positive (TP), true negative (TN)) and incorrect predictions (false positive (FP), false negative (FN) for each stage.
Accuracy	The proportion of epochs correctly classified by the CHT: ACC=(TP+TN)/(TP+TN+FP+FN). It ranges from 0 (no epoch correctly classified) to 1 (all epochs correctly classified).
Sensitivity (to sleep)	The proportion of sleep epochs correctly classified by the CHT: Sens=TP/(TP+FN).
Specificity (to wake detection; alternatively, Sensitivity to Wake following sleep onset)	The proportion of wake epochs correctly classified by the CHT: Spec=TN/(TN+FP). It ranges from 0 to 1.
F1 score	F1 score is an index of predictive performance: F1=2TP/(2TP+FP+FN)
Cohen's kappa	A measure of the agreement between two methods (e.g., PSG and sleep tracker) that are classifying epochs into categories (e.g., sleep/wake), which is correct for the agreement that could occur by chance:

	Cohen's Kappa = (Po-Pe)/(1-Pe), where Po is the probability of agreement; Pe: p (Disagreement |chance). It ranges from -1 (agreement lower than expected) to 1 (perfect agreement).
Prevalence-Adjusted and Bias-Adjusted Kappa (PABAK)	A measure of agreement that adjusts Cohen's kappa for the effects of prevalence and bias in the data (e.g., sleep is more prevalent than wake during nocturnal sleep)
Symmetric Mean Absolute Percentage Error (sMAPE)	An approach to quantify the percentage error of a sleep tracker's predictions relative to the PSG values: $$sMAPE = \left(\frac{100}{n}\right) * \sum_{t=1}^{n} \frac{\lvert y_t - x_t \rvert}{\lvert y_t \rvert + \lvert x_t \rvert}$$ *where x* is the reference device estimate (PSG); *y* is the CHT estimate; *n* is the number of simultaneous measurements. S^2 is the variance of the PSG and CHT measurements.

Recommendations

***Researchers*:** (some points relevant to industry as well):

1. As most performance evaluations are performed in the traditional setting of a sleep lab on a single night, multi-night and sleep-at-home studies where night-to-night variation in CHT performance [118-120], including variation in sleep [121], can be assessed, are advised.
2. For both algorithm training and evaluation purposes, use a protocol that inserts periods of in-bed but pre-sleep wakefulness as well as sleep-period interruptions to simulate real-life disrupted nighttime sleep and post-wake lie-in periods [59].
3. Expand nap detection testing [60].
4. Expand testing in adolescents and older adults as increasing numbers of these demographics use CHT.
5. For industrial uses, shift workers and persons on irregular sleep schedules should be recruited for performance testing.
6. For clinical uses, a multi-site combination of data is recommended to achieve sufficient scale to enable collected data to be robust.
7. Use open-source code for standardized performance evaluation of any sleep-tracking device. These exist in Python [34] and R [35] versions.

Manufacturers:

1. Develop an industry-wide standardized benchmarking framework for sleep measures evaluation to enhance comparability and transparency.
2. Provide easily accessible comprehensive white papers on performance evaluation, including important clinical studies that allow for standardized cross-device comparisons.
3. Provide researchers information about the mean absolute percentage errors for your sensor/output measurements where appropriate (relevant for HR, HRV, temperature).

5.2 Populations tested: present samples and future needs

Most commercial algorithms have been trained on healthy or relatively healthy participants obtained through convenience samples [39, 40, 80]. These persons tend to be of working age with normal or relatively normal sleep physiology. Insofar as most CHTs are purchased by this demographic, the derived algorithms are appropriate.

Even for healthy persons, data for children, adolescents, and older adults is relatively limited. In these age groups, physiological measures obtained from different sensors may be differentially modulated, affecting inferences about sleep and other health indicators. Conventional actigraphy requires adjustments to motion sensitivity for more accurate sleep/wake classification as children move more and older adults move less during sleep. Likely, these issues must also be addressed when developing algorithms for CHT.

For persons with sleep disorders and other medical conditions, the mapping from sensor data to clinically useful inference is more complex, involving the interaction between condition and severity of disease and the combination of multiple sensor data.

Some CHT features, such as the detection of atrial fibrillation, falls, febrile illness or massive oxygen desaturation, which are associated with pronounced deviations of sensor readouts from the norm, the demonstration of clinical utility can be more readily achieved. However, with common sleep conditions discussed earlier, additional algorithm training, validation, and refinement will be necessary for robust clinical utility. In some cases, this may be unachievable with current sensor modalities.

5.3 Defining what is 'good enough' for clinical applications

CHTs provide information about sleep that can be partially mapped to PSG for objective sleep monitoring. However, to expect concordance beyond an asymptote is unrealistic as the methods tap into different aspects of sleep physiology. When considering whether a particular CHT should be used for a given use case, the guiding question is, “is the device accurate enough for a specific use(s)?” The answer will depend on the availability of performance evaluation data and the use case.

For example, if a healthy adult opts to use a CHT to monitor their long-term sleep health for health improvement or maintenance, then a CHT with good performance levels will be suitable (Fig. 4). If, however, the intention is to use a CHT to screen for sleep disorder symptoms, then a higher degree of accuracy is often required [12]. As with all diagnostic testing, the acceptable error margin depends on the measure, the effect size that needs to be detected, and the consequences of false negatives.

Studies that compare the accuracy of clinical decision-making facilitated by data from the gold-standard measures versus a CHT are needed. These studies would be helpful for users and clinicians to determine whether CHTs are accurate enough for these higher-risk use cases. However, the currently available evidence suggests that multiple CHTs are as accurate, if not more accurate, than research-grade actigraphy devices and could thus be used for similar purposes.

We expect that, over time, two things will happen. First, data mapping sleep/health outcomes with CHT physiological markers, their cross-modality interaction, their temporal dynamics, and how they interact with relevant participant and environmental features will grow to the point where it exceeds what is possible with PSG except in specific neuroscience applications. Secondly, further development in sensor technology will open the door to new applications.

Beyond these technical issues, other major considerations to determine whether a CHT is currently 'good enough' for clinical applications include patient and clinician acceptance, feasibility, and cost-effectiveness. For example, devices with longer battery lives or simpler displays are easier to use in clinical contexts and often reduce data loss. Healthcare professionals are encouraged to consider the needs of their clients when determining device suitability. Some people do not like wearing rings, others dislike wrist-worn devices, and a few dislike both. CHTs also vary in their sensor composition and quality applications. For example, a device great for measuring nocturnal vascular signals may not be the best for exercise heart rate tracking.

Recommendations

Researchers:

1. Studies that evaluate the clinical utility of CHTs versus gold standard clinical measures would be highly valuable and should be encouraged.
2. Data on wear time and acceptability of usage should be included in longitudinal studies adopting CHT.

Clinicians:

1. Evidence about the clinical utility of CHTs is rapidly evolving; keeping abreast of the latest evidence is a difficult but necessary hurdle towards realizing the unprecedented potential of CHTs for transforming medicine.

2. Validated CHT should be suitable for purposes currently undertaken by actigraphy.
3. The 'right device' for a patient is one they will wear regularly, be willing to pay for and provide sufficiently good data.

M***anufacturers:***

1. Invest in directly mapping short-term and longitudinal device data to relevant health outcomes, e.g. mortality, risk of disease, and rate of disease progression, to develop a fresh framework for assessing the level of accuracy necessary for robust clinical application.
2. Consider applying the framework of equivalence testing [116] to this effort

— Fig. 4 Around here

6.0 TRANSFORMING DATA INTO CLINICALLY USEFUL RECORDS

6.1 Artificial intelligence, 'black-box' issues and upgrades

Artificial intelligence plays a critical role in CHTs. Machine learning is used to train algorithms that process CHT data, remove artifacts, and classify signals into understandable sleep, activity, and cardiac parameters. Additionally, machine learning algorithms are well suited for analysis of the massive datasets that arise from collecting data with CHTs used by a large proportion of the population. Such algorithms may utilize multiple small but consistent differences in signal between groups of interest that humans may not detect e.g. for atrial fibrillation detection using PPG signals [122]. Furthermore, artificial intelligence may power the apps associated with CHTs, with chatbots and virtual agents providing sleep advice.

There are multiple considerations when using artificial intelligence for any medical or wellness purpose, including but not limited to transparency regarding product development, testing, and intended use; performance and potential for performance changes with algorithm retraining or learning during use in the field; security; IT infrastructure; patient, provider, and staff education; and bias. Here, we discuss issues related to the use of machine learning classifiers in processing data acquired by CHT embedded sensors. For a comprehensive discussion of strengths, weaknesses, opportunities, and threats posed by artificial intelligence in sleep medicine, we direct readers to the recent publication in the Journal of Clinical Sleep Medicine [123].

Given the complexity inherent in CHT data collection, pre-processing, processing, analysis and presentation, some degree of tolerance of the 'black-box' nature of CHT is required. A majority of healthcare providers and patients accept that they do not have insights into the sensor and data processing details of medical devices, entrusting these to experts. However, there is disproportionate

concern about CHT which has persisted despite an appeal following discussions on their use as sleep and circadian biomarkers 6 years ago [27]. This could be based on prior experience with traditional actigraphy and associated software. When the original actigraphy algorithms were developed, the principles were intuitive and simple, published in considerable detail [5, 6], and explained in reviews [3, 124]. This allowed researchers to select appropriate algorithms to achieve reasonable performance under sleep disorder-specific constraints. Additionally, given an understanding of transformation of acceleration to activity counts and classification of sleep and wake, sleep assessments with actigraphy were considered rigorous, reproducible, and potentially generalizable across devices. Training algorithms adapted to CHT collected signals in different sleep and medical disorders is complex and requires money and dedicated resources. Any code must be broadly and easily usable, requiring constant upkeep. Few scientist-generated graphical user interfaces can compare with those in commercial products. Additionally, event data across 'research grade' devices are not directly comparable, even in specific applications [125].

Manufacturers constantly improve their products by updating hardware, firmware, and software to remain competitive. Upgrades can have different impacts on sleep assessment. The extent to which a user/researcher can control these also varies. Researchers can use a particular model of the device for the entire duration of a study. Firmware locking may be possible with some manufacturers. However, app or cloud-level software upgrades can change sleep algorithms and cannot be blocked at present. Most software updates relate to user interface or user experience enhancements rather than sleep algorithm changes. Unfortunately, information on how each update could affect the collected data is rarely provided.

Concerns about upgrades arise from the proprietary nature of each component, which could make it challenging to compare previously collected data with current and future data collection efforts. Research/clinical grade actigraphs face similar issues with upgrades; however, disclosure of sleep classifier algorithms allows for backward compatibility and reproducibility over long time periods. Of note, accelerometers may have variable quality across [9] and even within manufacturers over time. Secondly, as with CHT manufacturers, firmware upgrades change how sensor data is read out, resulting in non-identical measurements [126]. Thirdly, different implementations of the same algorithm, a proprietary and an open source version, can give different sleep metrics despite being fed identical acceleration data [7]. Few, if any, older papers record firmware or software versions used. As updates generally correct errors or improve measurement, insisting on older systems for backward compatibility and perceived transparency may be counterproductive. Historical 'research grade' algorithms were designed before the groundbreaking success of modern machine-learning methods and did not generalize well; they needed to be adjusted for age, biological sex, obesity, and health conditions with technical skill and effort to get closer to PSG outputs. Provided the right training data, machine learning excels at those tasks and requires minimum user input.

Recommendations:

Users:

1. Focus on trends rather than absolute values of sleep metrics when relating these to outcomes.
2. End-users and clinicians should accept that only a few individuals understand the entirety of the data chain, spanning physiology to clinical application. Insistence on publicizing proprietary methods is futile, although manufacturers should be encouraged to disclose the principles underlying how their devices measure sleep.

Manufacturers:

1. Provide sufficiently detailed instructions on how data from different sensors are combined to generate health measures of interest. This can be done without revealing the intricacies of how weights or features are combined [39, 40].
2. Evaluate the effects of firmware and software algorithm updates on a range of human participants before public release.
3. Communicate in advance when and how the updates may affect the issued sleep reports.
4. Provide details about what is changed with each change in firmware and software
5. Provide a facility to lock firmware upgrades.

Researchers and developers:

1. Develop tools to harmonize data across manufacturers to allow clinically meaningful comparisons where possible.
2. Pursue novel ways to combine sensor data, exploiting temporal and environmental changes in physiological measurements for novel health applications.

6.2 From CHT dashboard to electronic health record

The flow of data from the wearable dashboard to meaningful outcomes is illustrated in Fig. 5. Current CHT often provide application programming interfaces (APIs), through which wearable data can be extracted and exchanged with third-party, device-agnostic software platforms, for example, Apple Healthkit, Google Fit, and Samsung Health. Device-agnostic platforms may provide certified data analytics services such as sleep scoring [127] (e.g. https://sleepacta.com/en/) . Some platforms also enable integrating health data from multiple devices with different manufacturers. Data integration with electronic health record systems (EHRs) is rapidly evolving. Epic, for example, integrates patient-generated data via Apple Healthkit and Google Fit in discrete fields, although sleep parameters are not

supported as of February 2025. Third-party data integrators (e.g., Terra https://tryterra.co/ or Validic https://www.validic.com/) work with manufacturer specific data fields and API's to transfer CHT data into EHR, but they do not address the issue of heterogeneity of sleep data across different manufacturers.

At present, there is neither a unified, cross-industry data lexicon (which focuses on the definition of sleep measurements and their underlying concepts) nor a data dictionary (which specifies the structure, format, and relationships of data elements within a database). This gap motivated the creation of ANSI/CTA/NSF-2052.1-A [32], whose essence is captured in Section 3.2 in this recommendation. Agreeing on the constitution of these two collections is critical for incorporating CHT data into EHRs and, thereafter, into clinical workflows.

The integration data from different sources has been led by scientists from genomics and brain imaging fields who have employed statistical and deep learning methods to achieve data harmonization. They developed widely disseminated tools like ComBat [128]. Multi-site brain imaging studies often scan a few participants using both old and new protocols to allow differences related to upgrades to be reconciled. The sleep field can learn from these approaches. However, It should be noted that the development and maintenance of such technological infrastructure will require significant investment in expertise and effort.

Additionally, education on the use of CHT data and workflows is required. Without reimbursement for the utilization of CHT data in clinical settings, these costs could prove challenging for manufacturers to absorb. Therefore, while promising, wearable sleep tracking in clinical sleep medicine requires careful consideration even beyond device performance. This will include assessment of clinical utility, cost effectiveness, and the generation of false positive results that could result in the need for further testing, diagnosis, and treatment.

— Fig. 5 Around here

Recommendations

Manufacturers:

1. Adopt common definitions of selected 'fundamental sleep measures' as described in Section 3.2 for the initial fields used by EHR systems.
2. Provide evidence of compliance with minimal data quality standards for specific applications to qualify for incorporation into EHR for clinical decision support (Section 5.3).

7.0 MEDICAL REGULATORY STATUS: CURRENT STATUS AND FUTURE DIRECTIONS

Medical devices are regulated by national agencies such as the US FDA, EU MDR /CE, UK UKCA Mark, Canadian MDL, Australian TGA, Japanese PDMA, and Chinese NMPA. Each shares common principles of ensuring medical benefits and mitigating risks through classifying medical products into device classes according to the risks they pose. CHTs usually lie in the low risk category and are not scrutinized as closely as medical devices. This allows them to be sold as consumer goods. Insofar as they provide the sleep measures indicated in this document and make no claims about diagnosing and treating sleep disorders such as insomnia and sleep apnea, manufacturers are not obliged to seek medical device certification.

Most manufacturers will elect not to seek certification for commercial reasons as significant time, effort, and costs involved in the process outweigh commercial benefits when these devices are primarily used for wellness promotion (e.g., self-discovery). This is true even when the sleep metrics provided by CHTs overlap with those offered by medical-grade sleep devices, such as total sleep time and sleep efficiency.

Notably, the distinction between CHT and traditional medical devices is increasingly blurred, creating a complex certification landscape that may be difficult for users to assess. Empatica, for example, is a wearable medical device with FDA-approved oximeter, temperature, pulse rate, pulse rate variability, EDA, and an accelerometer. However, its sleep measurement is not certified.

Device quality is a priority for many manufacturers, and the quality of some CHTs already rivals that of certified instruments and can potentially assist in clinical decision-making. However, without certification, they cannot be used in more sensitive clinical trials or independently for diagnostic decision-making, particularly in the EU.

Software as a medical device (SaMD) is a regulatory pathway to certify consumer device features like atrial fibrillation detection. This pathway is increasingly being adopted to accommodate the rapid development of new technologies.

The regulatory landscape around AI and data protection for health data and privacy is rapidly evolving. As manufacturers exploit the potential of consumer electronics and machine learning-based algorithms for health applications (sleep being one), we can expect more manufacturers to seek certification.

Another area where we expect rapid evolution is in multi-device integration. Some manufacturers, such as Withings and Samsung, are moving along this path. The collection of redundant information

about sleep through different, complementary sensing technologies that probe diverse aspects of physiology can be expected to expand clinical applications in the future.

When the intended function of a device component is to provide 'good enough' data for clinical decision making, medical certification is sought. This entails a detailed documentation process where data quality has to match or exceed either prescribed standards set by the regulator or the performance of a device that has already been certified (as in the case of 510(k) filings in the US). Atrial fibrillation and sleep apnea diagnosis are examples of such certifiable functions. Meeting quality and safety standards for specific features does not warrant other device functions, and users should be educated about this distinction.

Recommendations

All:

1. Current CHTs should not be solely relied upon for accurate diagnosis or treatment of sleep disorders.
2. Device features used for medical purposes should remain regulated, but this can and should be distinguished from features used for general sleep, physical activity, and other physiology tracking features that should remain unregulated but where quality standards should be upheld.
3. CHT features that are not explicitly 'medically certified' should carry a statement cautioning unsupervised interpretation of device outputs. In particular, their uncertain performance for sleep measurement in contexts known to reduce accuracy should be declared unless solutions to address these are in place (i.e. extremes of age, persons who have difficulties with sleep initiation or maintenance, or known sleep or chronic medical disorders).

8.0 SUMMARY AND FUTURE RECOMMENDATIONS

These World Sleep Society Recommendations aim to provide timely and practical recommendations to users, researchers, clinicians, and manufacturers on consumer-centric wearable health trackers (CHTs) for sleep monitoring. As outlined, multiple factors must be considered when using CHTs for sleep tracking, including understanding how sleep is deduced from physiological signals and factors that influence these signal outputs (e.g., conditions that impact movement), and differences in variable definitions across devices (e.g., clarity around 'sleep scores').

Despite these challenges, this task force believes that many CHTs have various uses for healthy adults (e.g., self-discovery, preventive health), people with sleep symptoms (e.g., identifying characteristics of disordered sleep), and people with chronic sleep disorders or medical conditions (e.g., screening for

sleep-disordered breathing, engaging patients). Clinical implementation of CHTs has the potential to transform sleep medicine, and there are considerable research opportunities to test these devices in clinical settings. The wealth and multidimensionality of physiological and behavioral data that can be gaged with CHT create unprecedented opportunities for manufacturers and researchers to develop innovative sleep metrics that predict human performance and health outcomes more accurately than traditional approaches.

Performance evaluation will be continuously needed as these technologies are developed and refined. Key gaps include 1) performance evaluation in younger and older people and people with chronic health conditions; 2) performance evaluation of other emerging device classes, e.g. nearable health trackers, which have been investigated much less than wearable health trackers.

These recommendations also support some level of standardization across the rapidly growing industry, with an aspiration for manufacturers to adopt a common set of core sleep metrics (like the proposed FSM) and standardized performance evaluation. This will enable the development of normative values for sleep and respiratory metrics provided by CHTs to help identify abnormal cases for medical investigation.

In summary, CHTs present an unprecedented opportunity to engage individuals in learning about their sleep patterns and observing the impacts on their performance and health. Given that self-management is the cornerstone of behavior change, technologies that provide motivation and insights into healthy behavior change should be applauded and encouraged. While hurdles remain in device development, integration, and regulation, it is becoming increasingly apparent that CHTs (and inter-device integration) will be critically useful for individuals to manage their sleep health, for researchers to gain new insights into global sleep patterns, and for healthcare professionals to support the clinical management of their patients. Co-creation between manufacturers/researchers/clinicians is needed to identify and support use cases for the betterment of sleep health globally. Alignment by various parties to the principles outlined in this document is an important step in that direction.

Acknowledgements

The authors thank Max de Zambotti from Oura Health, Conor Heneghan from Google/Fitbit for detailed comments. Siqi Hao from the HUAWEI Health Lab, and the Xiaomi Wearables Team, provided insightful feedback. Representatives from Garmin and Withings read the content and expressed appreciation but did not provide specific comments. Opportunity to provide feedback on the document was also solicited from Apple, Samsung and the National Sleep Foundation. Feedback where provided was aggregated, reflect the views of individuals as independent scientific contributors and do not represent official positions held by their companies. The Task Force takes final responsibility for the content of the document.

References

[1] Research GV. Fitness tracker market size, share & tends analysis report by Type (Smart Watches, Smart Bands, Smart Clothing), by application, by wearing Type, by distribution channel, by region, and segment forecasts, 2025 - 2030. 2025.

[2] Foster F, Kupfer D, Weiss G, Lipponen V, McPartland R, Delgado J. Mobility recording and cycle research in neuropsychiatry. Biol Rhythm Res. 1972;3:61-72.

[3] Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep. 2003;26:342-92.

[4] Ancoli-Israel S, Martin JL, Blackwell T, Buenaver L, Liu L, Meltzer LJ, et al. The SBSM Guide to Actigraphy Monitoring: Clinical and Research Applications. Behav Sleep Med. 2015;13 Suppl 1:S4-S38.

[5] Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep. 1992;15:461-9.

[6] Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep. 1994;17:201-7.

[7] Cepni AB, Burkart S, Zhu X, White J, Finnegan O, Nelakuditi S, et al. Evaluating the performance of open-source and proprietary processing of actigraphy sleep estimation in children with suspected sleep disorders: A comparison with polysomnography. Sleep. 2024:10.1093/sleep/zsae267.

[8] Patterson MR, Nunes AAS, Gerstel D, Pilkar R, Guthrie T, Neishabouri A, et al. 40 years of actigraphy in sleep medicine and current state of the art algorithms. NPJ Digit Med. 2023;6:51.

[9] Weaver RG, Chandrashekhar MVS, Armstrong B, White Iii JW, Finnegan O, Cepni AB, et al. Jerks are useful: extracting pulse rate from wrist-placed accelerometry jerk during sleep in children. Sleep. 2025;48:10.1093/sleep/zsae099.

[10] Ryals S, Chiang A, Schutte-Rodin S, Chandrakantan A, Verma N, Holfinger S, et al. Photoplethysmography-new applications for an old technology: a sleep technology review. J Clin Sleep Med. 2023;19:189-95.

[11] Chinoy ED, Cuellar JA, Huwa KE, Jameson JT, Watson CH, Bessman SC, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep. 2021;44:zsaa291.

[12] Ong JL, Golkashani HA, Ghorbani S, Wong KF, Chee N, Willoughby AR, et al. Selecting a sleep tracker from EEG-based, iteratively improved, low-cost multisensor, and actigraphy-only devices. Sleep Health. 2024;10:9-23.

[13] Schwartz LP, Devine JK, Choynowski J, Hursh SR. Consumer preferences for sleep-tracking wearables: The role of scientific evaluation and endorsement. Sleep Health. 2024;10:163-70.

[14] 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;1:40-3.

[15] Chaput JP, Dutil C, Featherstone R, Ross R, Giangregorio L, Saunders TJ, et al. Sleep duration and health in adults: an overview of systematic reviews. Appl Physiol Nutr Metab. 2020;45:S218-S31.

[16] Lim DC, Najafi A, Afifi L, Bassetti C, Buysse DJ, Han F, et al. The need to promote sleep health in public health agendas across the globe. Lancet Public Health. 2023;8:e820-e6.

[17] Buysse DJ. Sleep health: can we define it? Does it matter? Sleep. 2014;37:9-17.

[18] Bei B, Wiley JF, Trinder J, Manber R. Beyond the mean: A systematic review on the correlates of daily intraindividual variability of sleep/wake patterns. Sleep Med Rev. 2016;28:108-24.

[19] Gombert-Labedens M, Alzueta E, Perez-Amparan E, Yuksel D, Kiss O, de Zambotti M, et al. Using Wearable Skin Temperature Data to Advance Tracking and Characterization of the Menstrual Cycle in a Real-World Setting. J Biol Rhythms. 2024;39:331-50.

[20] Lyzwinski L, Elgendi M, Menon C. Innovative Approaches to Menstruation and Fertility Tracking Using Wearable Reproductive Health Technology: Systematic Review. J Med Internet Res. 2024;26:e45139.

[21] Bruce LK, Kasl P, Soltani S, Viswanath VK, Hartogensis W, Dilchert S, et al. Variability of temperature measurements recorded by a wearable device by biological sex. Biol Sex Differ. 2023;14:76.

[22] Mason AE, Hecht FM, Davis SK, Natale JL, Hartogensis W, Damaso N, et al. Detection of COVID-19 using multimodal data from a wearable device: results from the first TemPredict Study. Sci Rep. 2022;12:3463.

[23] Mannhart D, Lischer M, Knecht S, du Fay de Lavallaz J, Strebel I, Serban T, et al. Clinical Validation of 5 Direct-to-Consumer Wearable Smart Devices to Detect Atrial Fibrillation: BASEL Wearable Study. JACC Clin Electrophysiol. 2023;9:232-42.

[24] Konstantinidis D, Iliakis P, Tatakis F, Thomopoulos K, Dimitriadis K, Tousoulis D, et al. Wearable blood pressure measurement devices and new approaches in hypertension management: the digital era. J Hum Hypertens. 2022;36:945-51.

[25] Yilmaz G, Ong JL, Ling LH, Chee MWL. Insights into vascular physiology from sleep photoplethysmography. Sleep. 2023;46:10.1093/sleep/zsad172.

[26] Ferizoli R, Karimpour P, May JM, Kyriacou PA. Arterial stiffness assessment using PPG feature extraction and significance testing in an in vitro cardiovascular system. Sci Rep. 2024;14:2024.

[27] Depner CM, Cheng PC, Devine JK, Khosla S, de Zambotti M, Robillard R, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. Sleep. 2020;43:10.1093/sleep/zsz254.

[28] de Zambotti M, Cellini N, Menghini L, Sarlo M, Baker FC. Sensors Capabilities, Performance, and Use of Consumer Sleep Technology. Sleep Med Clin. 2020;15:1-30.

[29] de Zambotti M, Goldstein C, Cook J, Menghini L, Altini M, Cheng P, et al. State of the science and recommendations for using wearable technology in sleep and circadian research. Sleep. 2024;47:10.1093/sleep/zsad325.

[30] 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;14:877-80.

[31] Schutte-Rodin S, Deak MC, Khosla S, Goldstein CA, Yurcheshen M, Chiang A, et al. Evaluating consumer and clinical sleep technologies: an American Academy of Sleep Medicine update. J Clin Sleep Med. 2021;17:2275-82.

[32] Association CT. Definitions and Characteristics for Wearable Sleep Monitors (ANSI/CTA-NSF-2052.1). Consumer Technology Association; 2022.

[33] Association CT. Methodology of Measurements for Feature In Sleep Tracking Consumer Technology Devices and Applications (ANSI/CTA/NSF-2052.2-A). Consumer Technology Association; 2024.

[34] Benedetti D, Menghini L, Vallat R, Mallett R, Kiss O, Faraguna U, et al. Call to action: an open-source pipeline for standardized performance evaluation of sleep-tracking technology. Sleep. 2023;46:zsac304.

[35] Menghini L, Cellini N, Goldstone A, Baker FC, de Zambotti M. A standardized framework for testing the performance of sleep-tracking technology: step-by-step guidelines and open-source code. Sleep. 2021;44:10.1093/sleep/zsaa170.

[36] Imtiaz SA. A Systematic Review of Sensing Technologies for Wearable Sleep Staging. Sensors (Basel). 2021;21.

[37] Rentz LE, Ulman HK, Galster SM. Deconstructing Commercial Wearable Technology: Contributions toward Accurate and Free-Living Monitoring of Sleep. Sensors (Basel). 2021;21:5071.

[38] Kwon S, Kim H, Yeo WH. Recent advances in wearable sensors and portable electronics for sleep monitoring. iScience. 2021;24:102461.

[39] Beattie Z, Oyang Y, Statan A, Ghoreyshi A, Pantelopoulos A, Russell A, et al. Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. Physiol Meas. 2017;38:1968-79.

[40] Altini M, Kinnunen H. The Promise of Sleep: A Multi-Sensor Approach for Accurate Sleep Stage Detection Using the Oura Ring. Sensors (Basel). 2021;21:4302.

[41] Miglis MG. Autonomic dysfunction in primary sleep disorders. Sleep Med. 2016;19:40-9.

[42] Kim H, Jung HR, Kim JB, Kim DJ. Autonomic Dysfunction in Sleep Disorders: From Neurobiological Basis to Potential Therapeutic Approaches. J Clin Neurol. 2022;18:140-51.

[43] Weaver RG, de Zambotti M, White J, Finnegan O, Nelakuditi S, Zhu X, et al. Evaluation of a device-agnostic approach to predict sleep from raw accelerometry data collected by Apple Watch

Series 7, Garmin Vivoactive 4, and ActiGraph GT9X Link in children with sleep disruptions. Sleep Health. 2023;9:417-29.

[44] Walch O, Chee MWL. Revisiting customized algorithms for research grade devices. Sleep. 2025:10.1093/sleep/zsaf011.

[45] Apple. Estimating Breathing Disturbances and Sleep Apnea Risk from Apple Watch. 2024. p. https://www.apple.com/health/pdf/sleep-apnea/Sleep_Apnea_Notifications_on_Apple_Watch_September_2024.pdf.

[46] Charlton PH, Kyriaco PA, Mant J, Marozas V, Chowienczyk P, Alastruey J. Wearable Photoplethysmography for Cardiovascular Monitoring. Proc IEEE Inst Electr Electron Eng. 2022;110:355-81.

[47] Jiang Y, Spies C, Magin J, Bhosai SJ, Snyder L, Dunn J. Investigating the accuracy of blood oxygen saturation measurements in common consumer smartwatches. PLOS Digit Health. 2023;2:e0000296.

[48] Hartmann V, Liu H, Chen F, Qiu Q, Hughes S, Zheng D. Quantitative Comparison of Photoplethysmographic Waveform Characteristics: Effect of Measurement Site. Front Physiol. 2019;10:198.

[49] Baumert M, Phan H. A perspective on automated rapid eye movement sleep assessment. J Sleep Res. 2025;34:e14223.

[50] Sun H, Ganglberger W, Panneerselvam E, Leone MJ, Quadri SA, Goparaju B, et al. Sleep staging from electrocardiography and respiration with deep learning. Sleep. 2020;43.

[51] Quigley KS, Gianaros PJ, Norman GJ, Jennings JR, Berntson GG, de Geus EJC. Publication guidelines for human heart rate and heart rate variability studies in psychophysiology-Part 1: Physiological underpinnings and foundations of measurement. Psychophysiology. 2024;61:e14604.

[52] Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med. 2020;3:18.

[53] Krauchi K. The human sleep-wake cycle reconsidered from a thermoregulatory point of view. Physiol Behav. 2007;90:236-45.

[54] Harding EC, Franks NP, Wisden W. Sleep and thermoregulation. Curr Opin Physiol. 2020;15:7-13.

[55] Boucsein W, Fowles DC, Grimnes S, Ben-Shakhar G, Roth WT, Dawson ME, et al. Society for psychophysiological research ad hoc committee on electrodermal measures. Publication recommendations for electrodermal measurements. Psychophysiol. 2012;49:1017-34.

[56] Herlan A, Ottenbacher J, Schneider J, Riemann D, Feige B. Electrodermal activity patterns in sleep stages and their utility for sleep versus wake classification. J Sleep Res. 2019;28:e12694.

[57] Sano A, Picard RW, Stickgold R. Quantitative analysis of wrist electrodermal activity during sleep. Int J Psychophysiol. 2014;94:382-9.

[58] 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. 2025:10.1093/sleep/zsae290.

[59] Willoughby AR, Golkashani HA, Ghorbani S, Wong KF, Chee N, Ong JL, et al. Performance of wearable sleep trackers during nocturnal sleep and periods of simulated real-world smartphone use. Sleep Health. 2024;10:356-68.

[60] Chinoy ED, Cuellar JA, Jameson JT, Markwald RR. Daytime Sleep-Tracking Performance of Four Commercial Wearable Devices During Unrestricted Home Sleep. Nat Sci Sleep. 2023;15:151-64.

[61] Sun J, Ma C, Zhao M, Magnussen GC, Xi B. Daytime napping and cardiovascular risk factors, cardiovascular disease, and mortality: A systematic review. Sleep Med Rev. 2022;65:101682.

[62] Lau T, Ong JL, Ng BKL, Chan LF, Koek D, Tan CS, et al. Minimum number of nights for reliable estimation of habitual sleep using a consumer sleep tracker. Sleep Adv. 2022;3:zpac026.

[63] Kaplan KA, Hirshman J, Hernandez B, Stefanick ML, Hoffman AR, Redline S, et al. When a gold standard isn't so golden: Lack of prediction of subjective sleep quality from sleep polysomnography. Biol Psychol. 2017;123:37-46.

[64] Stephan AM, Lecci S, Cataldi J, Siclari F. Conscious experiences and high-density EEG patterns predicting subjective sleep depth. Curr Biol. 2021;31:5487-500 e3.

[65] Buysse JD, Reynolds FC, Monk HT, Berman RS, Kupfer JD. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. Psychiatry Research. 1989;28:193-213.

[66] Knutson LK, Phelan J, Paskow JM, Roach A, Whiton K, Langer G, et al. The National Sleep Foundation's Sleep Health Index. Sleep Health. 2017;3:234-40.

[67] Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, Buysse D, 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. Sleep. 2015;38:843-4.

[68] Fox K, Borer JS, Camm AJ, Danchin N, Ferrari R, Lopez Sendon JL, et al. Resting heart rate in cardiovascular disease. J Am Coll Cardiol. 2007;50:823-30.

[69] Nwabuo CC, Appiah D, Moreira HT, Vasconcellos HD, Aghaji QN, Ambale-Venkatesh B, et al. Temporal Changes in Resting Heart Rate, Left Ventricular Dysfunction, Heart Failure and Cardiovascular Disease: CARDIA Study. Am J Med. 2020;133:946-53.

[70] Faust L, Feldman K, Mattingly SM, Hachen D, N VC. Deviations from normal bedtimes are associated with short-term increases in resting heart rate. NPJ Digit Med. 2020;3:39.

[71] Tasnim S, Tang C, Musini VM, Wright JM. Effect of alcohol on blood pressure. Cochrane Database Syst Rev. 2020;7:CD012787.

[72] Souza HCD, Philbois SV, Veiga AC, Aguilar BA. Heart rate variability and cardiovascular fitness: what we know so far. Vasc Health Risk Manag. 2021:701-11.

[73] Camm AJ, Malik M, Bigger JT, Breithardt G, Cerutti S, Cohen RJ, et al. Heart rate variability - Standards of measurement, physiological interpretation, and clinical use. Circulation. 1996;93:1043-65.

[74] Baumert M, Linz D, Stone K, McEvoy RD, Cummings S, Redline S, et al. Mean nocturnal respiratory rate predicts cardiovascular and all-cause mortality in community-dwelling older men and women. Eur Resp J. 2019;54.

[75] Nicolò A, Massaroni C, Schena E, Sacchetti M. The importance of respiratory rate monitoring: From healthcare to sport and exercise. Sensors. 2020;20:6396.

[76] Dünnwald T, Kienast R, Niederseer D, Burtscher M. The Use of Pulse Oximetry in the Assessment of Acclimatization to High Altitude. Sensors. 2021;21:1263.

[77] Cabanas AM, Fuentes-Guajardo M, Latorre K, León D, Martín-Escudero P. Skin Pigmentation Influence on Pulse Oximetry Accuracy: A Systematic Review and Bibliometric Analysis. Sensors. 2022;22:3402.

[78] Poorzargar K, Pham C, Ariaratnam J, Lee K, Parotto M, Englesakis M, et al. Accuracy of pulse oximeters in measuring oxygen saturation in patients with poor peripheral perfusion: a systematic review. J Clin Monit Comput. 2022;36:961-73.

[79] Galland BC, Short MA, Terrill P, Rigney G, Haszard JJ, Coussens S, et al. Establishing normal values for pediatric nighttime sleep measured by actigraphy: a systematic review and meta-analysis. Sleep. 2018;41:10.1093/sleep/zsy017.

[80] Zhang X, Kou W, Chang EI, Gao H, Fan Y, Xu Y. Sleep stage classification based on multi-level feature learning and recurrent neural networks via wearable device. Comput Biol Med. 2018;103:71-81.

[81] Olsen M, Zeitzer JM, Richardson RN, Davidenko P, Jennum PJ, Sorensen HBD, et al. A Flexible Deep Learning Architecture for Temporal Sleep Stage Classification Using Accelerometry and Photoplethysmography. IEEE Trans Biomed Eng. 2023;70:228-37.

[82] Baron KG, Abbott S, Jao N, Manalo N, Mullen R. Orthosomnia: Are Some Patients Taking the Quantified Self Too Far? J Clin Sleep Med. 2017;13:351-4.

[83] Vestergaard CL, Simpson MR, Sivertsen B, Kallestad H, Langsrud K, Scott J, et al. Weekday-to-weekend sleep duration patterns among young adults and outcomes related to health and academic performance. Sleep Sci Pract. 2024;8:15.

[84] Wing YK, Li SX, Li AM, Zhang J, Kong AP. The effect of weekend and holiday sleep compensation on childhood overweight and obesity. Pediatrics. 2009;124:e994-e1000.

[85] Akerstedt T, Ghilotti F, Grotta A, Zhao H, Adami HO, Trolle-Lagerros Y, et al. Sleep duration and mortality - Does weekend sleep matter? J Sleep Res. 2019;28:e12712.

[86] Lok R, Weed L, Winer J, Zeitzer JM. Adverse effects of late sleep on physical health in a large cohort of community-dwelling adults. Eur J Intern Med. 2024.

[87] Lok R, Weed L, Winer J, Zeitzer JM. Perils of the nighttime: Impact of behavioral timing and preference on mental health in 73,888 community-dwelling adults. Psychiatry Res. 2024;337:115956.

[88] Himali JJ, Baril AA, Cavuoto MG, Yiallourou S, Wiedner CD, Himali D, et al. Association Between Slow-Wave Sleep Loss and Incident Dementia. JAMA Neurol. 2023;80:1326-33.

[89] Qin S, Leong RLF, Ong JL, Chee MWL. Associations between objectively measured sleep parameters and cognition in healthy older adults: A meta-analysis. Sleep Med Rev. 2023;67:101734.

[90] Scullin MK, Bliwise DL. Sleep, cognition, and normal aging: integrating a half century of multidisciplinary research. Perspect Psychol Sci. 2015;10:97-137.

[91] Palagini L, Baglioni C, Ciapparelli A, Gemignani A, Riemann D. REM sleep dysregulation in depression: state of the art. Sleep Med Rev. 2013;17:377-90.

[92] Ross R, Chaput JP, Giangregorio LM, Janssen I, Saunders TJ, Kho ME, et al. Canadian 24-Hour Movement Guidelines for Adults aged 18-64 years and Adults aged 65 years or older: an integration of physical activity, sedentary behaviour, and sleep. Appl Physiol Nutr Metab. 2020;45:S57-S102.

[93] Willoughby AR, Alikhani I, Karsikas M, Chua XY, Chee MWL. Country differences in nocturnal sleep variability: Observations from a large-scale, long-term sleep wearable study. Sleep Med. 2023;110:155-65.

[94] Riemann D, Espie CA, Altena E, Arnardottir ES, Baglioni C, Bassetti CLA, et al. The European Insomnia Guideline: An update on the diagnosis and treatment of insomnia 2023. J Sleep Res. 2023;32:e14035.

[95] Zhang Y, Ren R, Lei F, Zhou J, Zhang J, Wing YK, et al. Worldwide and regional prevalence rates of co-occurrence of insomnia and insomnia symptoms with obstructive sleep apnea: A systematic review and meta-analysis. Sleep Med Rev. 2019;45:1-17.

[96] Harris J, Lack L, Kemp K, Wright H, Bootzin R. A randomized controlled trial of intensive sleep retraining (ISR): a brief conditioning treatment for chronic insomnia. Sleep. 2012;35:49-60.

[97] Mair A, Scott H, Lack L. Intensive sleep retraining treatment for insomnia administered by smartphone in the home: an uncontrolled pilot study. J Clin Sleep Med. 2022;18:1515-22.

[98] Kang SG, Kang JM, Cho SJ, Ko KP, Lee YJ, Lee HJ, et al. Cognitive Behavioral Therapy Using a Mobile Application Synchronizable With Wearable Devices for Insomnia Treatment: A Pilot Study. J Clin Sleep Med. 2017;13:633-40.

[99] Spina MA, Andrillon T, Quin N, Wiley JF, Rajaratnam SMW, Bei B. Does providing feedback and guidance on sleep perceptions using sleep wearables improve insomnia? Findings from "Novel Insomnia Treatment Experiment": a randomized controlled trial. Sleep. 2023;46:10.1093/sleep/zsad167.

[100] Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. Chest. 2014;146:1387-94.

[101] Baumert M, Immanuel SA, Stone KL, Litwack Harrison S, Redline S, Mariani S, et al. Composition of nocturnal hypoxaemic burden and its prognostic value for cardiovascular mortality in older community-dwelling men. Eur Heart J. 2020;41:533-41.

[102] Azarbarzin A, Sands SA, Stone KL, Taranto-Montemurro L, Messineo L, Terrill PI, et al. The hypoxic burden of sleep apnoea predicts cardiovascular disease-related mortality: the Osteoporotic Fractures in Men Study and the Sleep Heart Health Study. European heart journal. 2019;40:1149-57.

[103] Ahmadzadeh S, Luo J, Wiffen R. Review on biomedical sensors, technologies and algorithms for diagnosis of sleep disordered breathing: Comprehensive survey. IEEE Reviews in Biomedical Engineering. 2020;15:4-22.

[104] Roeder M, Bradicich M, Schwarz EI, Thiel S, Gaisl T, Held U, et al. Night-to-night variability of respiratory events in obstructive sleep apnoea: a systematic review and meta-analysis. Thorax. 2020;75:1095-102.

[105] Mendelson M, Bailly S, Marillier M, Flore P, Borel JC, Vivodtzev I, et al. Obstructive sleep apnea syndrome, objectively measured physical activity and exercise training interventions: a systematic review and meta-analysis. Frontiers in Neurology. 2018;9:73.

[106] Medicine AAoS. The AASM International Classification of Sleep Disorders – Third Edition, Text Revision (ICSD-3-TR). 2023. p. https://aasm.org/clinical-resources/international-classification-sleep-disorders/.

[107] Smith MT, McCrae CS, Cheung J, Martin JL, Harrod CG, Heald JL, et al. Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Clinical Practice Guideline. J Clin Sleep Med. 2018;14:1231-7.

[108] Marler MR, Gehrman P, Martin JL, Ancoli-Israel S. The sigmoidally transformed cosine curve: a mathematical model for circadian rhythms with symmetric non-sinusoidal shapes. Stat Med. 2006;25:3893-904.

[109] Mayer C, Walch O, Forger DB, Hannay K. Impact of Light Schedules and Model Parameters on the Circadian Outcomes of Individuals. J Biol Rhythms. 2023;38:379-91.

[110] Blume C, Santhi N, Schabus M. 'nparACT' package for R: A free software tool for the non-parametric analysis of actigraphy data. MethodsX. 2016;3:430-5.

[111] Huang Y, Mayer C, Cheng P, Siddula A, Burgess HJ, Drake C, et al. Predicting circadian phase across populations: a comparison of mathematical models and wearable devices. Sleep. 2021;44:10.1093/sleep/zsab126.

[112] Kim DW, Mayer C, Lee MP, Choi SW, Tewari M, Forger DB. Efficient assessment of real-world dynamics of circadian rhythms in heart rate and body temperature from wearable data. J R Soc Interface. 2023;20:20230030.

[113] Bowman C, Huang Y, Walch OJ, Fang Y, Frank E, Tyler J, et al. A method for characterizing daily physiology from widely used wearables. Cell Rep Methods. 2021;1.

[114] Emens JS, Burgess HJ. Effect of Light and Melatonin and Other Melatonin Receptor Agonists on Human Circadian Physiology. Sleep Med Clin. 2015;10:435-53.

[115] Hertenstein E, Gabryelska A, Spiegelhalder K, Nissen C, Johann AF, Umarova R, et al. Reference data for polysomnography-measured and subjective sleep in healthy adults. J Clin Sleep Med. 2018;14:523-32.

[116] Lakens D, Scheel AM, Isager PM. Equivalence Testing for Psychological Research: A Tutorial. Adv Meth Pract Psychol Sci. 2018;1:259-69.

[117] KK GR, Della Monica C, Atzori G, Lambert D, Hassanin H, Revell V, et al. Three Contactless Sleep Technologies Compared With Actigraphy and Polysomnography in a Heterogeneous Group of Older Men and Women in a Model of Mild Sleep Disturbance: Sleep Laboratory Study. JMIR Mhealth Uhealth. 2023;11:e46338.

[118] Chinoy ED, Cuellar JA, Jameson JT, Markwald RR. Performance of Four Commercial Wearable Sleep-Tracking Devices Tested Under Unrestricted Conditions at Home in Healthy Young Adults. Nat Sci Sleep. 2022;14:493-516.

[119] Ghorbani S, Golkashani HA, Chee N, Teo TB, Dicom AR, Yilmaz G, et al. Multi-Night at-Home Evaluation of Improved Sleep Detection and Classification with a Memory-Enhanced Consumer Sleep Tracker. Nat Sci Sleep. 2022;14:645-60.

[120] Chee N, Ghorbani S, Golkashani HA, Leong RLF, Ong JL, Chee MWL. Multi-Night Validation of a Sleep Tracking Ring in Adolescents Compared with a Research Actigraph and Polysomnography. Nat Sci Sleep. 2021;13:177-90.

[121] Chouraki A, Tournant J, Arnal P, Pepin JL, Bailly S. Objective multi-night sleep monitoring at home: variability of sleep parameters between nights and implications for the reliability of sleep assessment in clinical trials. Sleep. 2023;46:10.1093/sleep/zsac319.

[122] Harmon DM, Sehrawat O, Maanja M, Wight J, Noseworthy PA. Artificial Intelligence for the Detection and Treatment of Atrial Fibrillation. Arrhythm Electrophysiol Rev. 2023;12:e12.

[123] Bandyopadhyay A, Oks M, Sun H, Prasad B, Rusk S, Jefferson F, et al. Strengths, weaknesses, opportunities, and threats of using AI-enabled technology in sleep medicine: a commentary. J Clin Sleep Med. 2024;20:1183-91.

[124] Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev. 2011;15:259-67.

[125] Liu F, Schrack J, Wanigatunga SK, Rabinowitz JA, He L, Wanigatunga AA, et al. Comparison of sleep parameters from wrist-worn ActiGraph and Actiwatch devices. Sleep. 2024;47:10.1093/sleep/zsad155.

[126] John D, Sasaki J, Hickey A, Mavilia M, Freedson PS. ActiGraph activity monitors: "the firmware effect". Med Sci Sports Exerc. 2014;46:834-9.

[127] Topalidis P, Heib DP, Baron S, Eigl E-S, Hinterberger A, Schabus M. The virtual sleep lab—a novel method for accurate four-class sleep staging using heart-rate variability from low-cost wearables. Sensors. 2023;23:2390.

[128] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118-27.

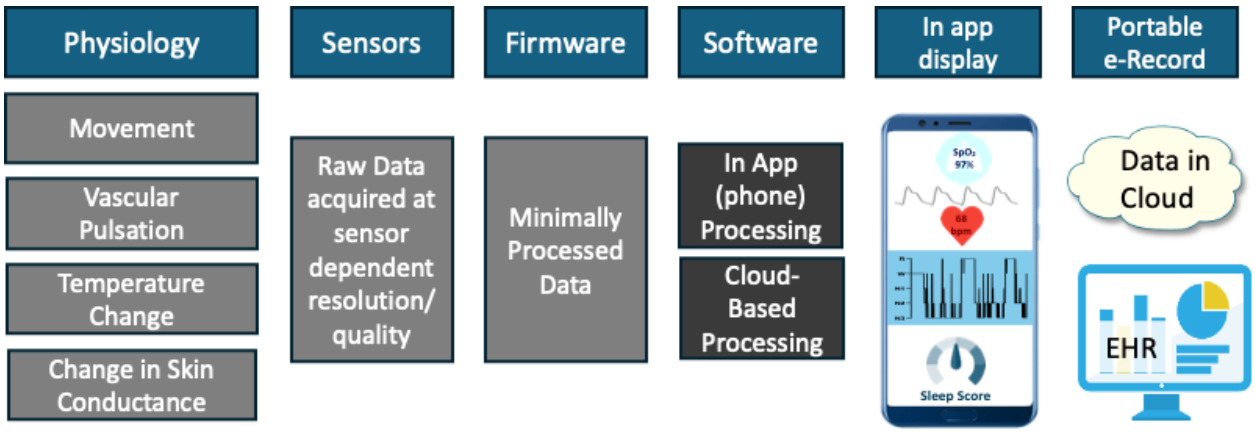
Figure captions

Fig. 1 Schematic showing data flow from participant to a portable e-record and suggestions to improve data handling at each stage of the chain. EHR: Electronic Health Record.

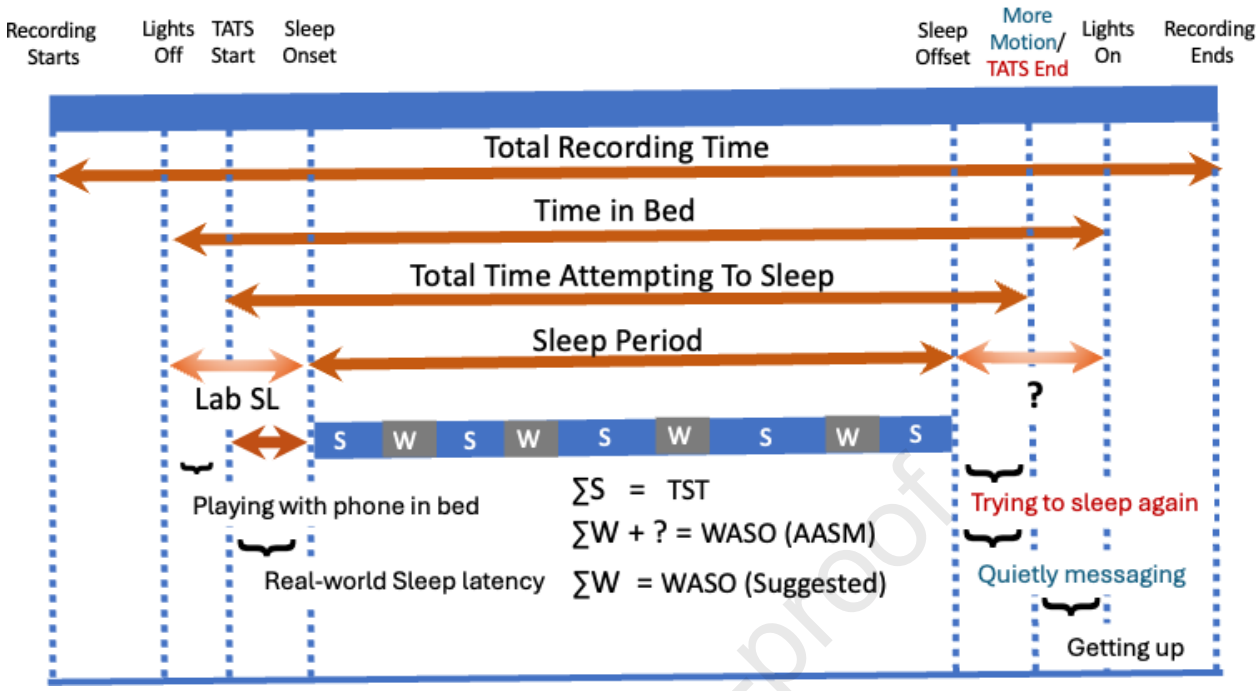
Fig. 2 Schematic showing measures used in wake-sleep classification as inherited from laboratory studies and the challenges in inferring "Lights off" (a proxy for intention to sleep) and "Time Attempting to Sleep (TATS)" automatically in real life as illustrated by an individual using their phone in bed after Lights Off. This complicates the determination of real-world 'sleep latency' outside the lab. Following the last epoch of sleep, which marks the end of the sleep period, some people lie in bed quietly before moving more. Periods of resting wakefulness that flank the sleep may result in the incorrect classification of wake as sleep. Others try to fall back asleep, complicating the assessment of Wake After Sleep Onset (WASO) and TATS End. 'Lights On' is a proxy for the end of 'Time in Bed' but inferring this timing is again difficult / impossible without some user or additional sensor input. AASM: American Academy of Sleep Medicine; TST: Total Sleep Time.

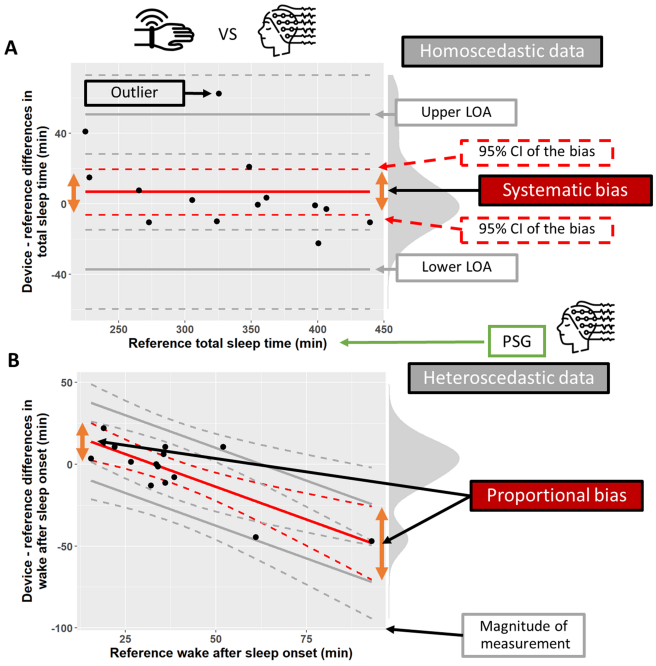
Fig. 3 A. Bland-Altman plot of a comparison of the assessment of TST by a CHT with PSG showing low systematic bias and no proportional bias. B Bland-Altman plot showing minimal systematic bias but large negative proportional bias. C. Epoch-by-epoch 4-class sleep stage agreement D. 4-Class Confusion matrix.

Fig. 4 Suggested guidelines for acceptable CHT sleep staging performance according to intended device use. Note that gaps between kappa value specifications between classes are left as these values are 'gray'. All the values provided here should be considered 'indicative'. (adapted from [12])



Proper device wear Contextual tagging Add medical info	Provide Sensor specs.	Implement update control, Specify effect of updates	Forewarn changes in algos, Specify effect of updates	Report a few standard measures in addition to proprietary ones	Create selected common fields across makers for EHR use





C

Epoch	PSG	Sleep Tracker	Agreement (2-stages)	Agreement (4-stages)
1	Wake	Wake	True Wake	True Wake
2	Wake	Light	False Sleep	False Light
3	N1	Light	True Sleep	True Light
4	N2	Light	True Sleep	True Light
5	N2	Light	True Sleep	True Light
6	N2	Deep	True Sleep	False Deep
7	N3	Deep	True Sleep	True Deep
8	N3	Deep	True Sleep	True Deep
....	....	....	....	....
180	REM	REM	True Sleep	True REM
181	REM	N1	True Sleep	False N1
182	N1	Wake	False Wake	False Wake
183	Wake	N1	False Sleep	False N1
....	....	....	....	....

The tracker wrongly identifies this epoch as sleep

The tracker correctly identifies this epoch as sleep, but providing the wrong sleep stage

D

		Sleep Tracker (Device)				Reference
		Wake	Light	Deep	REM	Total
PSG (Reference)	Wake	**1031** / **0.65**	467 / 0.29	29 / 0.02	71 / 0.04	**1598**
	Light (N1+N2)	272 / 0.05	**3969** / **0.78**	359 / 0.07	501 / 0.1	**5101**
	Deep (N3)	31 / 0.02	1059 / 0.53	**885** / **0.45**	13 / 0.01	**1988**
	REM	30 / 0.02	445 / 0.31	49 / 0.03	**898** / **0.63**	**1422**
Device Total		**1364**	**5940**	**1322**	**1483**	**10109**

The tracker wrongly identifies 501 epochs, defined as Light sleep by PSG as REM sleep. These epochs are the 10% of all epochs scored as REM by the PSG

The tracker correctly identifies 898 REM sleep epochs as defined by the PSG. These epochs are the 63% of all epochs scored as REM by the PSG

PRELIMINARY RECOMMENDATIONS FOR DEVICE QUALITY ACCORDING TO USE CASE

EEG-based wearable

2-stage kappa: ≥0.75
4-stage kappa: ≥0.75

Good for: Users who require highly accurate sleep staging performance in patient studies

- Clinicians/researchers studying samples with sleep disorders and fragmented sleep.
- Clinical trialists requiring uncompromising quality.
- Not available to ordinary consumers.

Iteratively improved CHT

2-stage kappa: 0.55-0.70
4-stage kappa: 0.45-0.70

Good for: Users assessing sleep longitudinally in persons without highly disordered sleep

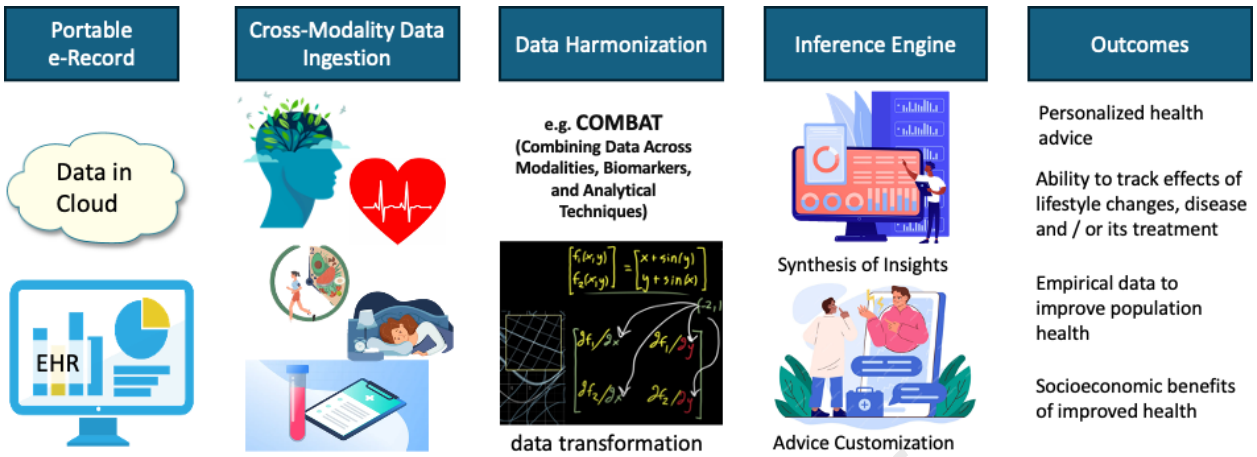
- Researchers evaluating population health involving mostly persons with non-disordered sleep.
- Consumers who desire highest quality sleep measurement.

Low-cost CHT

2-stage kappa: < 0.35
4-stage kappa: < 0.35

Good for: Cost conscious users requiring only basic sleep period logging

- Users satisfied with logging of time-in-bed and can tolerate poorer sleep measurement accuracy.
- Not suitable as an independent device for research.



Declaration of interest statement: MCWL is on the medical advisory boards of Oura and Quantactions. MB has received research funding from ZOLL Respicardia and the ResMed Foundation. HS has a patent regarding sleep onset detection sponsored by Re-Time Pty Ltd and has received research support from Re-Time Pty Ltd, Compumedics Ltd, Withings Ltd, and the American Academy of Sleep Medicine Foundation. CG is on the medical advisory boards of Huxley Medical and Apnimed. KB is a consultant for the National Sleep Foundation and has received research funding from Google and Aether Mindtech. SAI is a co-founder and CTO of Acurable. TP is on the medical advisory board of Bayer Healthcare and a consultant to Cerebra, Sleepimage, and has received speaker fees from AGB-Pharma, Bioprojet, Idorsia, Somnico, Philips, and Löwenstein Medical. CAK is a consultant to Cerebra, Avadel Pharmaceuticals, Lilly, Morgan Stanley, Oxama Medical, Restful Robotics, Samsung, SoundHealth Inc., Vivos Therapeutics, Genentech, and Alkermes.