Title page

Self-applied home sleep recordings: the future of sleep medicine

Henri Korkalainen Ph.D.[1,2,*,#], Sami Nikkonen Ph.D.[1,2,*], Samu Kainulainen Ph.D.[1,2], Amit Krishna Dwivedi Ph.D.[1,2], Sami Myllymaa Ph.D.[1,2], Timo Leppänen Ph.D.[1,2], Juha Töyräs Ph.D.[1,2,3]

*Co-first authors; #Corresponding author

[1]Department of Applied Physics, University of Eastern Finland, Kuopio, Finland.

[2]Diagnostic Imaging Center, Kuopio University Hospital, Kuopio, Finland.

[3]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia.

Email:

henri.korkalainen@uef.fi; sami.nikkonen@uef.fi; samu.kainulainen@uef.fi; amit.dwivedi@uef.fi; sami.myllymaa@uef.fi; timo.leppanen@uef.fi; juha.toyras@uef.fi

Mailing address (for all authors)

Department of Applied Physics, University of Eastern Finland

P.O. Box 1627, 70211, Kuopio, Finland

Disclosure statement

This study was financially supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 965417, NordForsk (NordSleep project 90458-06111) via Business Finland (5133/31/2018), the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding (5041767, 5041768, 5041794, 5041797, 5041798, and 5041803), the Academy of Finland (323536), Tampere Tuberculosis Foundation, Päivikki and Sakari Sohlberg Foundation, Finnish Cultural Foundation – North Savo Regional Fund, and The Research Foundation of the Pulmonary Diseases. The authors declare no conflicts of interest.

Acknowledgements

We would like to thank Matias Rusanen for his invaluable assistance in preparing the figures.

Clinics care points

- In-lab polysomnography is the gold standard method but suffers from high costs, limited availability, and laborsome manual analysis.
- Home recordings are well-suited especially for screening purposes and follow-up monitoring.
- In the future, simpler home recording devices alongside fully automatic analysis have massive potential for sleep disorder diagnostics
- Current machine learning-based automatic scoring methods can already be considered as accurate as manual scoring; however, more comprehensive validation in heterogeneous populations is needed before widespread clinical adoption.
- Home recordings cannot fully replace the in-lab polysomnography in complex situations and when multiple sleep disorders are suspected.

Synopsis (max 100 words)

Sleep disorders form a massive global health burden and there is an increasing need for simple and cost-efficient sleep recording devices. Recent machine learning-based approaches have already achieved scoring accuracy of sleep recordings on par with manual scoring, even with reduced recording montages. Simple and inexpensive monitoring over multiple consecutive nights with automatic analysis could be the answer to overcome the substantial economical burden caused by poor sleep and enable more efficient initial diagnosis, treatment planning, and follow-up monitoring for individuals suffering from sleep disorders.

Key points (3-5 bullet points)

- Sleep disorders are constantly increasing in prevalence inducing a massive burden to the health care systems. As the current gold standard polysomnography is expensive, requires substantial labor, and has limited availability, there is an increasing need for simple home-based recordings.
- Advancements in machine learning and artificial intelligence allow automatic analysis of sleep recordings even using reduced measurement setups without compromising diagnostic accuracy. These approaches are already reaching accuracy on par with manual scoring by clinical experts.
- There is immense potential in using wearable sensing solutions for screening and long-term monitoring of sleep disorders, and their use would be a highly valuable addition to diagnostics. Combining simple screening devices with automatic analysis would enable cost-efficient monitoring over multiple nights which would, in turn, minimize the first-night effect.

- The pulse oximeter is one of the most potential devices to act as an efficient and accurate stand-alone screening device. Photoplethysmogram and blood oxygen saturation measured with a finger pulse oximeter have already been successfully used to identify sleep stages and assess the severity of obstructive sleep apnea. These automatic analysis approaches could already be easily adapted to all home-based sleep recordings.
- In complex cases or when multiple sleep disorders are suspected, simplified recording setups with a reduced electroencephalography montage, pulse oximetry, leg electromyography, and respiratory measurements at home could be used.

Keywords

sleep disorders, home sleep recordings, machine learning, deep learning, electroencephalography, photoplethysmography, wearables, sensors, medical devices

Introduction

Sleep disorders and inadequate sleep are quickly becoming a substantial global health problem. Poor sleep induces a major economical and social burden cost due to direct health care costs, loss of productivity, and increased risk of accidents and traffic crashes[1]. Meanwhile, sleep disorders have an increasingly high prevalence. Obstructive sleep apnea (OSA) alone is estimated to affect hundreds of millions of individuals[2], whereas insomnia symptoms are prevalent in up to half of the adult population, with 10 – 15% of the population also suffering from daytime impairment[3]. It is evident that efficient diagnostic practices for sleep medicine are crucial. Due to the heavy patient inflow and the limited capacity of sleep laboratories, home-based recordings will most likely have an increasing role in the future.

The current gold standard in diagnosing sleep disorders and studying sleep is the type I polysomnography (PSG) conducted at a specialized sleep laboratory. PSG records the electrical activity of the brain (electroencephalography, EEG), eye movements (electrooculography, EOG), chin and leg muscle tone (electromyography, EMG), and cardiac function (electrocardiography, ECG). In addition, respiratory effort, airflow, blood oxygen saturation, and sleeping position are recorded alongside additional video and audio recordings[4,5]. The signals commonly recorded in modern PSG, the sensors used to record these signals, and the rationale behind the signal inclusion are presented in Table 1.

Table 1. The signals included in modern polysomnography, the commonly used sensor types, and the main diagnostic use of the recorded signals[4–8]

Recording	Sensor types	Main diagnostic use
Airflow	Thermistor, nasal pressure sensor	Respiratory event scoring
Audio	Microphone	Breathing and snoring sounds
Blood oxygen saturation	Pulse oximeter	Desaturation event scoring
Body position	Accelerometer	Identify positional sleep apnea
Electrical activity of the brain	Cup electrodes	Sleep stage and arousal scoring
Eye movements	Adhesive electrodes, cup electrodes	Sleep staging
Cardiac function	Adhesive electrodes	Heart rate
Muscle tone and leg movements	Adhesive electrodes	Sleep staging and periodic limb movements
Respiratory effort	RIP-belts, piezoelectric sensor	Differentiating central, mixed, and obstructive apneas
Video	Video camera	Investigating behavioral patterns and identifying issues in recording

RIP=respiratory inductance plethysmography

Although type I PSG is the most comprehensive diagnostic method and can be especially useful in the differential diagnosis when multiple sleep disorders are suspected, it has several major limitations and shortcomings. One of the primary drawbacks is high cost. A sleep laboratory staffed with professional sleep technologists is required to conduct the PSG, further increasing complexity and cost[9]. PSG also requires substantial labor from professional sleep technologists as the electrodes and measurement devices must be placed meticulously,

the participant must be supervised during the night, and the manual scoring of the recordings is a highly time-consuming process [10].

Moreover, PSG may not always be fully representative of normal sleep and can suffer from a considerable first-night effect where the unfamiliar environment and complex PSG equipment causes discomfort and stress disturbing normal sleep[11,12]. The first night effect is very difficult to eliminate completely even if the sleep laboratory is set up in a more comfortable and less clinical environment, such as in a hotel[12]. Possibly the most effective solution for reducing or even eliminating the first-night effect is to record sleep for multiple consecutive nights. The benefits of recording sleep over multiple nights are also supported by the fact that there exists significant night-to-night variation in sleep and the severity of some common sleep disorders[11,13,14]. Therefore, it is clear that one recording night for PSG may not sufficiently catch all sleep disorders nor provide a good representation of a typical night. Nevertheless, due to above mentioned practical constraints, only a single monitoring night is almost exclusively used in clinical sleep medicine[12].

Compared to the full in-laboratory PSG, simpler ambulatory devices are also available for use in sleep diagnostics[6]. The Task Force of the Standards of Practice Committee of the American Sleep Disorder Association (ASDA) has defined four monitor types for sleep recording[5]. The requirements for each category are presented in Table 2. Type I recording is a standard attended in-lab PSG that is recommended for most sleep studies and often required to diagnose complex sleep disorders[5,15]. Type II recording refers to a full PSG setup conducted unattended in a home environment often missing the video and audio recordings. Type III device is an unattended polygraphy device used to diagnose some sleep disorders, e.g. OSA. However, it cannot be used to fully replace PSG due to the lack of EEG preventing accurate

sleep staging and detecting arousals from sleep[10]. Most recording devices that are accepted to be used for sleep disorder diagnosis are of types I-III. Only a few type IV devices, required to record a single channel, are in diagnostic use.

The main advantage of home-based measurements is the capability of the studied individual to sleep in a familiar environment. Moreover, home recordings do not require healthcare practitioners for active monitoring thus providing a more cost-efficient option over an in-lab PSG[15]. However, ambulatory devices have their limitations as they generally cannot be used fully independently and thus still require at least some training and set up by a professional. For example, a type II PSG usually requires a sleep technologist to place at least the EEG electrodes and set up the device which takes approximately one hour of time[10]. Type III devices may be equipped by the studied individual and only guidance is required from healthcare professionals. However, mistakes made by the patients and incorrect use of the devices in unattended conditions lead to greater failure rates and poorer signal quality compared to full in-lab PSG[15,16]. This in turn facilitates an increased need for retesting and therefore mitigating some of the cost benefits of the ambulatory devices over full in-lab PSG[15,16]. Ambulatory devices also suffer from a higher rate of data loss which can lead to inconclusive results and further necessitate retesting[17].

Table 2. The different sleep monitor types as defined by the Task Force of the Standards of Practice Committee of the American Sleep Disorder Association[5,8,18]. The table presents minimum requirements; however, many modern sleep monitors record more signals.

Sleep study type	Diagnostic purpose	Minimum number of signals	Required signals
Type I in-laboratory PSG	Various sleep disorders	8	EEG, EOG, chin EMG, ECG, airflow, respiratory effort, body position, oxygen saturation
Type II unattended PSG	Various sleep disorders	7	EEG, EOG, chin EMG, heart rate, airflow, respiratory effort, oxygen saturation
Type III unattended polygraphy	Mainly sleep apnea	4	respiratory effort, airflow, heart rate, oxygen saturation
Type IV unattended recording	Mainly monitoring	1	respiratory effort or airflow or oxygen saturation

ECG=electrocardiography, EEG=electroencephalography, EMG=electromyography, EOG=electrooculography, PSG=polysomnography

The classification of sleep monitoring devices was done over two decades ago and it is still the current official specification[5]. However, there have been massive advancements in sensor, signal analysis, and recording technology after this specification. Therefore, there can be a wide range in capability between devices of the same type. It should also be noted that AASM has not included this sleep monitor type classification in their scoring manual and only separates between in-lab PSG (type I) and home sleep apnea test (HSAT) devices (types II-IV).

It is clear that simpler, more affordable, and fully automatic devices, diagnostic methods, and analysis tools are needed in sleep diagnostics. These new approaches could simultaneously mitigate some of the shortcomings of current diagnostic methods while allowing more widespread screening and diagnosis. This would in turn enable treatment and its follow-up monitoring for many who are currently suffering from a sleep disorder but are not diagnosed due to waiting times or limited diagnostic resources. Overall, current technical innovations could be exploited to streamline the diagnostic process and move towards the next-generation self-applied home sleep recordings. Mainly, the advancements in simpler wearable sensors and automated analyses based on artificial intelligence (AI), or more precisely machine learning and deep learning, could allow simpler and more affordable sleep recordings in the future. Below, we present our views on the future of sleep recordings based on the most recent advancements in sensing technology and AI-based analysis methods.

Recording and scoring of sleep stages

Sleep staging is traditionally performed by manually reviewing the PSG recordings in 30-second segments, called epochs. Each epoch of sleep is scored to one of five stages: Stage W (wakefulness), Stage R (rapid eye movement sleep, REM), and three non-REM (NREM) Stages N1, N2, and N3. The most common characteristics of each sleep stage are presented in Table 3. Identifying sleep stages is mainly based on EEG, EOG, and chin EMG signal features, patterns, and waveforms. The gold standard of identifying sleep stages requires recordings of three EEG channels (F4-M1, C4-M1, and O2-M1) with three additional backup channels (F3-M2, C3-M2, and O1-M2), two EOG channels, and a chin-EMG channel[4]. However, it has been reported that using all these recommended channels might not be necessary for accurate

sleep staging as other derivations will only lead to slightly different results in sleep stage scoring[19,20].

Table 3. The common characteristics of sleep stages[4,8].

Sleep Stage	Characterized by	Clarification
Wake	Alpha rhythm	Sinusoidal 8–13 Hz activity in EEG
	Eye blinking	Vertical eye movements at a frequency range of 0.5–2 Hz
	High muscle tone	High chin EMG activity
N1	LAMF activity	LAMF EEG activity mostly at 4–7 Hz
	Slow eye movements	Regular, sinusoidal eye movements with a deflection duration of over 0.5 s
	Varying muscle tone	Chin EMG activity varies but is generally lower than during wake
N2	K complexes	A sharp wave with both negative and positive components
	Sleep spindles	A train of sinusoidal waves (11–16 Hz) with a duration >0.5 s
	Low EOG, varying EMG	Usually minimal eye movements. Chin EMG activity varies but is usually lower than during wake
N3	Slow-wave activity	Slow waves (0.5–2 Hz) with an amplitude >75 μV
	No eye movements	Usually, no visible eye movements and the EOG only displays the same frequencies as the EEG
	Low EMG	Chin EMG activity is usually the lowest of all NREM stages
REM	Rapid eye movements	Irregular, sharp eye movements with a deflection duration <0.5 s
	Low chin muscle tone	EMG activity at the lowest level of all sleep stages
	Transient muscle activity	Short irregular bursts of EMG activity usually with duration <0.25 s
	Sawtooth waves	Trains of triangular, serrated, 2–6 Hz waves

ECG=electrocardiography, EEG=electroencephalography, EMG=electromyography, EOG=electrooculography, LAMF=Low-amplitude mixed-frequency, NREM=non-rapid eye movement, PSG=polysomnography, REM=rapid eye movement

The sleep staging process is currently only possible from type I and type II PSGs and the sleep architecture remains unknown with other types of sleep recordings. This inhibits using type III and IV recording devices when diagnosing most sleep disorders, especially in complex situations with several comorbidities. The self-applied type III recordings are most often used in diagnosing OSA as the respiratory events during the night can mostly be identified without the sleep staging. Type III devices are even the preferred diagnostic method for OSA in some healthcare systems as they are simpler and allow the patient to sleep at home; decreasing costs and increasing patient comfort[21,22]. However, the lack of sleep staging and EEG recording leads to an unreliable estimation of OSA severity since some respiratory events require EEG-based arousal detection to be accurately detected[23]. In addition, the total sleep time is important for accurate estimation of OSA severity and if the total recording time is used instead of the total sleep time, OSA severity can be significantly underestimated[23]. Still, the same thresholds are used for defining the OSA severity and choosing the patients for receiving health insurance- or government-subsidized treatment regardless of the diagnostic device type[24,25]. In addition, the identification of more specific conditions, such as REM-related OSA, cannot be done without sleep staging. Finally, differential diagnosis when multiple sleep disorders are suspected is impossible without sleep staging. Therefore, even a simple assessment of sleep architecture would be highly beneficial in home-based measurements.

Self-applicable electrode sets and wearable EEG devices already exist for measuring EEG at home[26–30, 31]. These do not require any input from a sleep technologist to set up, aside from possibly providing instructions. They usually measure EEG from the forehead area instead of the crown typically done in 10-20 derived systems (Figure 1). For example, a screen-printed self-applicable electrode set has been successfully used in identifying sleep stages with a low

failure rate and high correspondence to type I PSG[20,30]. As the electrode set enables good quality measurements without the need for any skin preparation, it would be well suited for self-applied home measurements[32]. However, these devices still require manual sleep stage scoring. In addition, since these devices use non-standard electrode placements, they require rigorous validation before they can be adopted for clinical use. This is because the forehead EEG signals collected from non-standard locations may not contain exactly the same information as signals from the 10-20 system; thus the standard scoring rules may not be applicable as such without adjustments.

The AASM also considers peripheral arterial tonometry (PAT) devices acceptable for estimating sleep time and for the identification of respiratory events[4]. At least one such device, Watch PAT, is on the market although the accuracy of the device is not on par with standard HSAT methods[33–36].

Recently, there has been an increasing number of consumer-grade wearable devices for assessing sleep quality. These include devices such as wristbands, smartwatches, and rings for sleep tracking. Despite having suffered from low reliability compared to type I PSG during disordered sleep[37], there may be potential to enable simple and comfortable assessment of sleep architecture in the future. However, rigorous clinical validation and appropriate medical approval processes are certainly needed before any of these can be adapted to diagnostic practices.

With the advancement of machine learning methodology, different solutions capable of identifying sleep stages from simple measurements have been introduced. Machine learning methods have enabled highly accurate sleep staging even based on a single frontal EEG channel, with epoch-by-epoch accuracies as high as 83%[38–41]. In comparison, the inter scorer

agreement in the AASM inter scorer reliability program was reported to be 82.6%[42]. Therefore, these machine learning-based models seem to be already comparable to expert manual scoring. Similar approaches have also been able to provide an estimate of the sleep architecture based on simpler surrogate measurements. For example, ECG measurement has been successfully used in differentiating between sleep and wakefulness[43–45]. Even simpler solutions may be used; for example, a photoplethysmogram (PPG) measured with a pulse oximeter has been shown to be a viable option[46–49]. With deep learning applied to PPG, it has been shown that differentiating between the individual sleep stages is possible with a moderate agreement to manual scoring of PSG (64.1%), and good agreement in differentiating between NREM, REM, and wakefulness (80.1%)[50] (Figure 2A). The PPG-based sleep staging could be extremely useful, even if not as accurate as EEG-based sleep staging since it only requires a simple pulse oximeter, which is a cheap, easy-to-use, and reliable sensor that is already integrated into all type I-III recording devices. In addition, it is completely non-invasive and causes minimal disruption to sleep and thus would be highly useful for reduced recording setups such as type III or type IV recordings. Therefore, comparing PPG sleep staging accuracy to the accuracy of full PSG setup may not be always relevant as it would be only used in situations where EEG recording is not available.

The future of sleep recording at home certainly lies in implementing simple measurement devices capable of identifying sleep stages. Whenever a sensitive and accurate measurement is required, the self-applicable electrode sets would be the most viable option due to their simple placement and easy comparison to standard EEG measurements. Consumer-grade wearable devices have high potential but are currently not ready for use in a diagnostic setting before rigorous clinical validation and medical approval process. Wearable devices would provide an even simpler setup compared to single-use self-applicable electrode sets and

would enable monitoring over consecutive nights. While they may be more prone to failed measurements and poor placement, long-term monitoring could make up for these limitations. Finally, whenever the information on the total sleep time and a rough division to NREM/REM/wakefulness are sufficient, the surrogate measures to EEG for sleep staging are the most viable option. For example, PPG is already recorded during practically all home-based sleep measurements and would not require any additional sensors to enable differentiation of sleep stages.

Most sleep analysis software, such as Noxturnal (Nox system, Nox Medical, Iceland), RemLogic (Embla/Embletta systems, Natus, USA), and Sleepware 3G (Philips Alice system, Philips, Netherlands), also allow automatic sleep staging and respiratory event scoring. However, the accuracy of these methods varies and in most cases is not thoroughly tested or even reported at all. Therefore, the automatic scoring methods provided by the sleep scoring software cannot be considered reliable for clinical use without manual correction. In addition, the position of AASM is still that diagnosis and treatment decisions cannot be only based on automatic methods or artificial intelligence-based algorithms and the raw data must be interpreted manually[51,52].

Recording and scoring of respiratory events

The diagnosis of OSA is among the most common reasons for conducting a PSG as it has been estimated that as much as half of the adult population is afflicted by OSA according to current clinical standards[2]. OSA is a nocturnal breathing disorder characterized by frequent obstructions in the upper airways during sleep. These obstructions lead to either complete (apnea) or partial (hypopnea) cessations in breathing, called respiratory events[4]. The diagnosis of OSA is primarily based on the number of apnea and hypopnea events per hour

of sleep, called the apnea-hypopnea index (AHI), which determines the severity of OSA[4]. More precisely, OSA is diagnosed if the AHI ≥ 5 with associated signs or symptoms of sleepiness or if the AHI ≥ 15 even without any symptoms[4,53]. Therefore, the diagnosis of OSA requires the detection of respiratory events from PSG signals. This process is currently done manually by annotating the signals using visual scoring rules, which is very time-consuming. Although most recording software provides initial built-in automated scoring, it is not meant for final diagnosis and still needs to be manually corrected and evaluated. These factors make respiratory event scoring currently very laborious and therefore also expensive[10].

A wide variety of automatic classification methods for the detection of OSA exist[54]. Most of these methods are not capable of scoring individual events or even estimating the AHI, making them unable to accurately evaluate the OSA severity[54]. However, recent studies have shown, that by using machine learning methods, an extremely accurate and fully automatic estimation of the AHI is possible based solely on the oxygen saturation signal[55]. For example, over 90% accuracy in OSA severity estimation with a median AHI error of <1 has been reported in an HSAT dataset using the 4% desaturation criteria for hypopnea scoring[55] (Figure 2B). Similar results have also been shown with cerebrovascular disease patients and in large external test populations indicating good generalizability and robustness of these methods[55,56]. A further advantage of these methods is that they only require the blood oxygen saturation recording during the night; thus they can be easily applied to all types of PSG or home sleep monitoring devices as oxygen saturation is acquired with a pulse oximeter, which is already included in all type I-III devices. As mentioned before, the pulse oximeter is a simple and easy-to-use sensor that requires no calibration or set up. Therefore, the pulse oximetry-based OSA severity estimation could be especially suitable for large-scale screening of OSA and in various settings of scientific research.

While the pulse oximetry-based methods are well suited for tasks such as large-scale screening for OSA, they are only able to provide estimates of the AHI and OSA severity and cannot be used to diagnose more specific variants of sleep apnea, e.g. central sleep apnea (CSA), supine dominant or isolated OSA, or REM-related OSA. Therefore, these automatic methods cannot be directly compared to standard manual respiratory event scoring or used for detailed respiratory event analysis. However, this kind of fully automatic respiratory event scoring based on machine learning methods has also been recently shown possible[8]. The automatic respiratory event scoring has been reported with an agreement of 88.9% (κ=0.728) with manual scoring, which is very close to the inter-scorer agreement reported by the AASM inter-scorer reliability program (93.9%, κ=0.92)[8,42]. Therefore, automatic machine learning-based scoring could also be applied for comprehensive respiratory event scoring when it is required, with an accuracy that is near human expert scoring. One disadvantage of this method is that the machine learning model also requires respiratory effort and airflow signals in addition to the SpO_2, and therefore requires a more complex recording setup[8]. However, all of these signals are relatively easy to record and are present in all type I-III sleep monitors. Therefore, this method could be best suited for reducing manual scoring workload for recording setups where more information on the respiratory events, other than their frequency, is needed.

As with sleep staging and other sleep disorders, there is also considerable night-to-night variation in OSA severity[11,13]. Thus, multiple-night studies could greatly increase the accuracy of the OSA severity evaluation as easy-to-record signals with accompanying automatic analysis could be simply conducted for multiple consecutive nights with minimal additional cost or resources. This would simultaneously eliminate the first-night effect of sleeping with

an unfamiliar and uncomfortable recording setup and the random night-to-night variation in sleep patterns that is always present.

Identifying arousals from sleep

Arousals from sleep are defined as an abrupt shift in EEG frequencies lasting over 3 seconds[4]. Therefore, identifying these requires an EEG recording. Moreover, as the arousals are further differentiated to spontaneous arousals, respiratory arousals, respiratory effort-related arousals (RERAs), and periodic limb movement (PLM) arousals, these require additional measurements, i.e. breathing signals and leg EMG signals. Currently, these all are only included in type I and type II recordings.

Similar to sleep staging, automatic identification of arousals has been conducted based on PSG recordings[57–59]. In 2018, there was a PhysioNet/Computing in Cardiology Challenge to develop an automatic arousal detection software[57]. The winning approach relied on deep learning utilizing EEG, EOG, chin EMG, RIP belt, oxygen saturation, and airflow signals and achieved reasonable accuracy[58]. Besides requiring a comprehensive measurement setup from a full PSG, the arousal identification only relied on classifying 10-second epochs into arousal or not arousal instead of accurately identifying the actual starting and ending times of the arousals. Aside from these, arousal identification has been conducted from ECG recording with promising results[59]. Therefore, there is potential to conduct automatic detection of arousals in a clinical setting both from signals recorded during a PSG and from simpler measurements such as PPG. Arousal identification would be crucial to assess sleep fragmentation and in the diagnosis of OSA, as this would allow the identification of hypopneas related to arousals. However, more research is warranted before the simpler approaches can be applied in routine clinical practice as scoring of arousals suffers from a relatively low

agreement even between manual scorers evidenced by reported arousal index intraclass correlation of as low as 0.54[60,61].

Identifying movements, snoring, and cardiac events

Detecting limb movement events is essential for assessing the presence of periodic limb movements and differentiating between different arousals types. Automatic detection of limb movements has also been successfully conducted from leg EMGs relying on deep learning[62]. However, leg EMG recordings are not consistently included in home-based measurements. Thus, AI-based algorithms could be the solution for identifying limb movements from surrogate measures. Potential alternatives include activity-based measurements and position sensors.

Identifying snoring is mostly done based on audio recordings which can also be conducted automatically[63,64]. Implementing a simple recording of audio to home sleep recordings would be straightforward and could be done, for example, with an ambient microphone or a microphone placed over the trachea. These could be accompanied by automatic analyses to easily assess the snoring tendency, which is often also related to OSA. However, in microphone-based snoring detection, it can sometimes be difficult to isolate background noise or bed partner's snoring from the recording which complicates accurate snoring detection.

The detection of cardiac events conventionally relies on the ECG signal. However, the PPG has immense potential to function as a surrogate for ECG[65–68]. Even though PPG is not a direct measure of the electrical functioning of the heart, the pulse rate and pulse rate variability

metrics correlate well with those derived from ECG[66]. Furthermore, computational solutions for PPG-based atrial fibrillation, ectopic complexes, and extrasystole detection exist[65,67,68]. However, these methods warrant further studies on their usability as a part of routine sleep recordings.

The limitations of home recordings

Home sleep recordings have certain limitations that may be difficult to completely overcome and therefore likely cannot fully replace in-lab PSG. For example, in-lab PSG often incorporates a video recording of the night. This is especially crucial when diagnosing disorders such as parasomnias, behavioral night-time disorders, or nocturnal epilepsy. Diagnosing these in a home environment would either require a portable night-vision video recording device with sufficient quality or other surrogate measurements capable of reliably assessing nighttime behaviors. These could potentially include sonar-based solutions or accelerometers directly measuring the activity or movement of the individual with accompanying algorithms for automatic analysis. Moreover, some sleep studies, such as multiple sleep latency test or maintenance of wakefulness test cannot be conducted at home in their current form. Therefore, an accurate objective assessment of daytime sleepiness still requires an in-lab recording. Finally, many devices are still relatively difficult to use. Thus, more development work needs to still be conducted to make them suitable for all individuals regardless of factors such as mental status or age.

Conclusions

Sleep disorders are a global health burden and more efficient diagnostic practices are sorely needed. Home sleep recordings with accompanying artificial intelligence-based automatic analysis approaches have immense potential to resolve the increasing patient inflows. The specific conclusions related to the future of home sleep recordings are:

1. AI-based sleep staging and OSA severity estimation are already highly accurate and nearly on par with manual scoring. Thus, night-to-night variation, the first-night effect, and incorrect use of devices already cause greater uncertainty in the diagnosis than the current automatic analysis methods. As such, the accuracy of these methods is no longer the limiting issue for their adoption and lack of understanding and trust likely play a much greater role. However, these trust concerns are not entirely unfounded as there is a clear lack of large-scale multi-center validation studies which are certainly needed before adopting any automatic method for general clinical use.
2. Once validated, the automatic analysis methods could considerably improve the diagnostics of sleep and sleep disorders as the automatic methods would not only reduce the workload related to manual scoring but also make the analysis more consistent and improve the overall accuracy and ease of comparison.
3. Simpler recording setups would be highly useful for screening and assisting in the diagnosis of many sleep disorders. Cheap and easy-to-use recording devices could increase the availability of recordings and enable diagnosis and follow-up of treatment for more individuals. Simpler measurement setups would simultaneously be more comfortable and impose a lesser disruption to sleep. Thus, they could allow obtaining a more reliable representation of natural sleep.

4. Combining automatic AI-based analysis and simple screening devices would enable simple and cost-efficient monitoring over multiple consecutive nights. This would minimize the first-night effect and the effect of night-to-night variability and release the limited resources of sleep specialists from manual scoring to other tasks.
5. Based on the current research, we consider that pulse oximeter recording including the PPG and SpO_2 has the most potential for simple and cost-efficient yet accurate screening of sleep disorders. A pulse oximeter is cheap and easy-to-use, requires no calibration or difficult setup, and can be used multiple times with no preparation or a need to replace single-use parts such as electrodes. Furthermore, it is fully non-invasive and causes minimal disruption to sleep. These factors make it an ideal sensor for continuous long-term and multi-night recordings. Finally, and most importantly, the measured signals (oxygen saturation and PPG) are extremely information-rich and can be alone used to automatically evaluate sleep stages and the severity of OSA with high accuracy (Figure 1)[50,55].
6. For more complex sleep disorders and specialized analysis, simplified self-applied recording setups with reduced EEG, pulse oximetry, leg EMG, and respiratory measurements could be used. The signals recorded by the simplified devices could still be scored using automatic machine learning-based methods. Traditional type I PSG could be reserved for only those cases which cannot be reliably studied and diagnosed otherwise.

References

1. Hillman D, Mitchell S, Streatfeild J, Burns C, Bruck D, Pezzullo L. The economic cost of inadequate sleep. *Sleep*. 2018;41(8):zsy083. doi:10.1093/sleep/zsy083

2. Benjafield A, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;7(8):687-698. doi:10.1016/s2213-2600(19)30198-5

3. Schutte-Rodin S, Broch L, Buysse D, Dorsey C, Sateia M, Schutte-Rodin SL. Clinical Guideline for the Evaluation and Management of Chronic Insomnia in Adults. *J Clin Sleep Med*. 2008;4(5):487-504.

4. Berry RB, Albertario CL, Harding SM, et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications.* Version 2.5 Darien, IL: American Academy of Sleep Medicine; 2018.

5. Ferber R, Millman R, Coppola M, et al. ASDA standards of practice: Portable recording in the assessment of obstructive sleep apnea. *Sleep*. 1994;17(4):378-392. doi:10.1093/sleep/17.4.378

6. Collop NA, Anderson WMD, Boehlecke B, et al. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. *J Clin Sleep Med*. 2007;3(7):737-747. doi:10.5664/jcsm.27032

7. Bianchi MT. Sleep devices: wearables and nearables, informational and interventional, consumer and clinical. *Metabolism*. 2018;84:99-108.

doi:10.1016/j.metabol.2017.10.008

8. Nikkonen S, Korkalainen H, Leino A, et al. Automatic respiratory event scoring in obstructive sleep apnea using a long short-term memory neural network. *IEEE J Biomed Heal Informatics*, 2021.doi: 10.1109/JBHI.2021.3064694.

9. Punjabi NM. The epidemiology of adult obstructive sleep apnea. *Proc Am Thorac Soc*. 2008;5(2):136-143. doi:10.1513/pats.200709-155MG

10. Fischer J, Dogas Z, Bassetti CL, et al. Standard procedures for adults in accredited sleep medicine centres in Europe. *J Sleep Res*. 2012;21(4):357-368. doi:10.1111/j.1365-2869.2011.00987.x

11. Newell J, Mairesse O, Verbanck P, Neu D. Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples. *Psychiatry Res*. 2012;200(2-3):795-801. doi:10.1016/j.psychres.2012.07.045

12. Hutchison KN, Song Y, Wang L, Malow BA. Analysis of sleep parameters in patients with obstructive sleep apnea studied in a hospital vs. a hotel-based sleep center. *J Clin Sleep Med*. 2008;4(2):119-122. doi:10.5664/jcsm.27127

13. Bittencourt LRA, Suchecki D, Peres C, et al. The variability of the apnoea-hypopnoea index. *J Sleep Res*. 2001;10(3):245-251.

14. Miettinen T, Myllymaa K, Hukkanen T, Töyräs J, Sipilä K, Myllymaa S. Home polysomnography reveals a first-night effect in patients with low sleep bruxism activity. *J Clin Sleep Med*. 2018;14(8):1377-1386. doi:10.5664/jcsm.7278

15. Golpe R, Jimenéz A, Carpizo R. Home sleep studies in the assessment of sleep

apnea/hypopnea syndrome. *Chest*. 2002;122(4):1156-1161. doi:10.1378/chest.122.4.1156

16. Whittle AT, Finch SP, Mortimore IL, MacKay TW, Douglas NJ. Use of home sleep studies for diagnosis of the sleep apnoea/hypopnoea syndrome. *Pneumologie*. 1998;52(8):467.

17. Ahmed M, Patel NP, Rosen I. Portable monitors in the diagnosis of obstructive sleep apnea. *Chest*. 2007;132(5):1672-1677. doi:10.1378/chest.06-2793

18. Chesson AL, Berry RB, Pack A. Practice parameters for the use of portable monitoring devices in the investigation of suspected obstructive sleep apnea in adults. *Sleep*. 2003;26(7):907-913. doi:10.1093/sleep/26.7.907

19. Ruehland WR, O'Donoghue FJ, Pierce RJ, et al. The 2007 AASM recommendations for EEG electrode placement in polysomnography: Impact on sleep and cortical arousal scoring. *Sleep*. 2011;34(1):73-81. doi:10.1093/sleep/34.1.73

20. Myllymaa S, Muraja-Murro A, Westeren-Punnonen S, et al. Assessment of the suitability of using a forehead EEG electrode set and chin EMG electrodes for sleep staging in polysomnography. *J Sleep Res*. 2016;25(6):636-645. doi:10.1111/jsr.12425

21. Flemons WW, Douglas NJ, Kuna ST, Rodenstein DO, Wheatley J. Access to Diagnosis and Treatment of Patients with Suspected Sleep Apnea. *Am J Respir Crit Care Med*. 2004;169(6):668-672. doi:10.1164/rccm.200308-1124PP

22. Arnardottir ES, Verbraecken J, Gonçalves M, et al. Variability in recording and scoring of respiratory events during sleep in Europe: a need for uniform standards. *J Sleep Res*. 2016;25(2):144-157. doi:10.1111/jsr.12353

23. Bianchi MT, Goparaju B. Potential underestimation of sleep apnea severity by at-home kits: Rescoring in-laboratory polysomnography without sleep staging. *J Clin Sleep Med*. 2017;13(4):551-555. doi:10.5664/jcsm.6540

24. American Academy of Sleep Medicine. Sleep-Related Breathing Disorders in Adults : Recommendations for Syndrome Definition and Measurement Techniques in Clinical Research. *Sleep*. 1999;22(5):662-689.

25. Korkalainen H, Töyräs J, Nikkonen S, Leppänen T. Mortality-risk-based apnea–hypopnea index thresholds for diagnostics of obstructive sleep apnea. *J Sleep Res*. 2019;28(6):1-7. doi:10.1111/jsr.12855

26. Myllymaa S, Muraja-Murro A, Westeren-Punnonen S, et al. Assessment of the suitability of using a forehead EEG electrode set and chin EMG electrodes for sleep staging in polysomnography. *J Sleep Res*. 2016;25(6):636-645. doi:10.1111/jsr.12425

27. Levendowski DJ, Ferini-Strambi L, Gamaldo C, Cetel M, Rosenberg R, Westbrook PR. The accuracy, night-To-night variability, and stability of frontopolar sleep electroencephalography biomarkers. *J Clin Sleep Med*. 2017;13(6):791-803. doi:10.5664/jcsm.6618

28. Younes M, Soiferman M, Thompson W, Giannouli E. Performance of a new portable wireless sleep monitor. *J Clin Sleep Med*. 2017;13(2):245-258. doi:10.5664/jcsm.6456

29. Arnal PJ, Thorey V, Debellemaniere E, et al. The Dreem Headband compared to polysomnography for electroencephalographic signal acquisition and sleep staging. *Sleep*. 2020;43(11):1-13. doi:10.1093/sleep/zsaa097

30. Miettinen T, Myllymaa K, Westeren-Punnonen S, et al. Success Rate and Technical

Quality of Home Polysomnography with Self-Applicable Electrode Set in Subjects with Possible Sleep Bruxism. *IEEE J Biomed Heal Informatics*. 2018;22(4):1124-1132. doi:10.1109/JBHI.2017.2741522

31. Kainulainen S, Korkalainen H, Sigurðardóttir S, et al. Comparison of EEG signal characteristics between polysomnography and Self Applied Somnography setup in a pediatric cohort. *IEEE Access (in press)*. 2021.

32. Kalevo L, Miettinen T, Leino A, et al. Effect of Sweating on Electrode-Skin Contact Impedances and Artifacts in EEG Recordings With Various Screen-Printed Ag/Agcl Electrodes. *IEEE Access*. 2020;8:50934-50943. doi:10.1109/ACCESS.2020.2977172

33. Hedner J, White DP, Malhotra A, et al. Sleep staging based on autonomic signals: A multi-center validation study. *J Clin Sleep Med*. 2011;7(3):301-306. doi:10.5664/JCSM.1078

34. Zou D, Grote L, Peker Y, Lindblad U, Hedner J. Validation a portable monitoring device for sleep apnea diagnosis in a population based cohort using synchronized home polysomnography. *Sleep*. 2006;29(3):367-374. doi:10.1093/sleep/29.3.367

35. Choi JH, Lee B, Lee JY, Kim HJ. Validating the Watch-PAT for diagnosing obstructive sleep apnea in adolescents. *J Clin Sleep Med*. 2018;14(10):1741-1747. doi:10.5664/jcsm.7386

36. Pang KP, Gourin CG, Terris DJ. A comparison of polysomnography and the WatchPAT in the diagnosis of obstructive sleep apnea. *Otolaryngol - Head Neck Surg*. 2007;137(4):665-668. doi:10.1016/j.otohns.2007.03.015

37. Liang Z, Chapa Martell MA. Validity of Consumer Activity Wristbands and Wearable EEG

for Measuring Overall Sleep Parameters and Sleep Structure in Free-Living Conditions. *J Healthc Informatics Res*. 2018;2(1-2):152-178. doi:10.1007/s41666-018-0013-1

38. Korkalainen H, Aakko J, Nikkonen S, et al. Accurate Deep Learning-Based Sleep Staging in a Clinical Population with Suspected Obstructive Sleep Apnea. *IEEE J Biomed Heal Informatics,* 2019:doi: 10.1109/JBHI.2019.2951346. doi:10.1109/JBHI.2019.2951346

39. Mousavi S, Afghah F, Acharya UR. SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach. *PLoS One*. 2019;14(5):e0216456. doi:10.1371/journal.pone.0216456

40. Phan H, Andreotti F, Cooray N, Chen OY, De Vos M. Joint Classification and Prediction CNN Framework for Automatic Sleep Stage Classification. *IEEE Trans Biomed Eng*. 2019;66(5):1285-1296. doi:10.1109/TBME.2018.2872652

41. Phan H, Mikkelsen K, Chén OY, et al. Personalized automatic sleep staging with single-night data: A pilot study with kl-divergence regularization. *Physiol Meas.* 41(6), 064004.

42. Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine inter-scorer reliability program: Respiratory events. *J Clin Sleep Med*. 2014;10(4):447-454. doi:10.5664/jcsm.3630

43. Li Q, Li Q, Liu C, Shashikumar SP, Nemati S, Clifford GD. Deep learning in the cross-time frequency domain for sleep staging from a single-lead electrocardiogram. *Physiol Meas*. 2018;39(12):124005. doi:10.1088/1361-6579/aaf339

44. Fonseca P, Long X, Radha M, Haakma R, Aarts RM, Rolink J. Sleep stage classification with ECG and respiratory effort. *Physiol Meas*. 2015;36(10):2027-2040. doi:10.1088/0967-3334/36/10/2027

45. Willemen T, Van Deun D, Verhaert V, et al. An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification. *IEEE J Biomed Heal Informatics*. 2014;18(2):661-669. doi:10.1109/JBHI.2013.2276083

46. Korkalainen H, Aakko J, Duce B, et al. Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea. *Sleep*. 2020;43(11):1-10. doi:10.1093/sleep/zsaa098

47. Fonseca P, Weysen T, Goelema MS, et al. Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. *Sleep*. 2017;40(7). doi:10.1093/sleep/zsx097

48. Beattie Z, Oyang Y, Statan A, et al. Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. *Physiol Meas*. 2017;38(11):1968-1979. doi:10.1088/1361-6579/aa9047

49. Dehkordi P, Garde A, Dumont GA, Ansermino JM. Sleep/wake classification using cardiorespiratory features extracted from photoplethysmogram. *Comput Cardiol (2010)*. 2016;43:1021-1024. doi:10.22489/cinc.2016.294-147

50. Korkalainen H, Aakko J, Duce B, et al. Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea. *Sleep*. 2020;(May):1-10. doi:10.1093/sleep/zsaa098

51. Rosen IM, Kirsch DB, Carden KA, et al. Clinical use of a home sleep apnea test: An updated American academy of sleep medicine position statement. *J Clin Sleep Med*. 2018;14(12):2075-2077. doi:10.5664/jcsm.7540

52. Goldstein CA, Berry RB, Kent DT, et al. Artificial intelligence in sleep medicine: An

American Academy of Sleep Medicine position statement. *J Clin Sleep Med*. 2020;16(4):605-607. doi:10.5664/jcsm.8288

53. American Academy of Sleep Medicine. *International Classification of Sleep Disorders*. 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014.

54. Uddin MB, Chow CM, Su SW. Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: a systematic review. *Physiol Meas*. 2018;39(3). doi:10.1088/1361-6579/aaafb8

55. Nikkonen S, Afara IO, Leppänen T, Töyräs J. Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea. *Sci Rep*. 2019;9:1-9. doi:10.1038/s41598-019-49330-7

56. Leino A, Nikkonen S, Kainulainen S, et al. Neural network analysis of nocturnal SpO2 signal enables easy screening of sleep apnea in patients with acute cerebrovascular disease. *Sleep Med*. 2020. doi:10.1016/j.sleep.2020.12.032

57. Ghassemi MM, Moody BE, Lehman LWH, et al. You Snooze, You Win: The PhysioNet/Computing in Cardiology Challenge 2018. *Comput Cardiol (2010)*. 2018;2018-Septe:20-23. doi:10.22489/CinC.2018.049

58. Howe-Patterson M, Pourbabaee B, Benard F. Automated Detection of Sleep Arousals from Polysomnography Data Using a Dense Convolutional Neural Network. *Comput Cardiol (2010)*. 2018;2018-Septe:1-4. doi:10.22489/CinC.2018.232

59. Li A, Chen S, Quan SF, Powers LS, Roveda JM. A deep learning-based algorithm for detection of cortical arousal during sleep. *Sleep*. 2020;43(12):1-10. doi:10.1093/sleep/zsaa120

60. Drinnan MJ, Murray A, Griffiths CJ, Gibson GJ. Interobserver Variability in Recognizing Arousal in Respiratory Sleep Disorders. *American journal of respiratory and critical care medicine*, *158*(2), 358-362

61. Whitney CW, Gottlieb DJ, Redline S, et al. Reliability of scoring respiratory disturbance indices and sleep staging. *Sleep*. 1998;21(7):749-757. doi:10.1093/sleep/21.7.749

62. Biswal S, Sun H, Goparaju B, Brandon Westover M, Sun J, Bianchi MT. Expert-level sleep scoring with deep neural networks. *J Am Med Informatics Assoc*. 2018;25(12):1643-1650. doi:10.1093/jamia/ocy131

63. Azarbarzin A, Moussavi ZMK. Automatic and unsupervised snore sound extraction from respiratory sound signals. *IEEE Trans Biomed Eng*. 2011;58(5):1156-1162. doi:10.1109/TBME.2010.2061846

64. Swarnkar VR, Abeyratne UR, Sharan R V. Automatic picking of snore events from overnight breath sound recordings. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS*. 2017:2822-2825. doi:10.1109/EMBC.2017.8037444

65. Pereira T, Tran N, Gadhoumi K, et al. Photoplethysmography based atrial fibrillation detection: a review. *npj Digit Med*. 2020;3(1). doi:10.1038/s41746-019-0207-9

66. Gil E, Orini M, Bailón R, Vergara JM, Mainardi L, Laguna P. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiol Meas*. 2010;31(9):1271-1290. doi:10.1088/0967-3334/31/9/015

67. Saritas T, Greber R, Venema B, et al. Non-invasive evaluation of coronary heart disease in patients with chronic kidney disease using photoplethysmography. *Clin Kidney J*.

2019;12(4):538-545. doi:10.1093/ckj/sfy135

68. Drijkoningen L, Lenaerts F, Van Der Auwera J, et al. Validation of a smartphone based photoplethysmographic beat detection algorithm for normal and ectopic complexes. *Comput Cardiol (2010)*. 2014;41(January):845-848.

Figure captions

Figure 1: Forehead/face (grey) is a common area to attach self-applicable electrodes used in wearable EEG devices[20,29,31] (A). The 10-20 system-derived electrode places are used in conventional in-lab polysomnography (B).

Figure 2: Examples of automatic pulse oximetry-based sleep stage classification[50] (A) and apnea-hypopnea index (AHI) estimation[55] (B) and how they correspond to expert manual scoring. The sleep stages and AHI were automatically determined from the photoplethysmogram (PPG) and the blood oxygen saturation (SpO_2) signals, respectively. As such, these results highlight what can be already achieved utilizing only a single probe- pulse oximetry measurement instead of the full polysomnography setup.