Article Title: The future of sleep measurements – a review and perspective

Author Names, Degrees, and Affiliations

Erna Sif Arnardottir, PhD[1,2], Anna Sigríður Islind, PhD[1,3], María Óskarsdóttir, PhD[1,3].

[1] Reykjavik University Sleep Institute, School of Technology, Reykjavik University, Reykjavik, Iceland.

[2] Internal Medicine Services, Landspitali University Hospital, Reykjavik, Iceland.

[3] Department of Computer Science, Reykjavik University, Reykjavik, Iceland.

Corresponding author

Dr. Erna Sif Arnardottir

Reykjavik University

Menntavegi 1

102 Reykjavik

Iceland.

ernasifa@ru.is

Disclosure Statement

Dr. Arnardottir discloses lecture fees from Nox Medical, Philips and ResMed. Dr. Islind and Dr. Oskarsdottir have nothing to disclose.” The work of the authors is sponsored in part by the Sleep Revolution which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 965417 and by NordForsk (NordSleep project 90458-06111) via the The Icelandic Centre for Research.

Key Words sleep measurement, subjective data, objective data, sleep diary, co-design, machine learning, data management platform, data science.

Key Points

- We argue for the need for improved subjective and objective assessment for future sleep studies and discuss the current use, limitations and potentials of various types of sleep assessment.
- Data from wearables and non-wearables outline a future potential for informing sleep research and clinical practice to assess long-term effects on sleep but these devices need to be validated further.
- Co-designing with patients and healthcare professionals can enable improved diagnostics and collaborative care, informed by data.
- Data management platforms designed and developed to securely display a variety of data for i) healthcare professionals, ii) patients and iii) researchers can enable the synergy of objective and subjective data in one place.
- Data science and machine learning techniques applied to a variety of sleep data can facilitate the discovery of new and important patterns and novel insights in sleep data.

Clinical Care Points

- When performing a sleep assessment, do not rely only on objective sleep testing as the subjective experience of the patient is equally important.
- Do not rely only on a single screening questionnaire to decide who needs further sleep testing, as most of these questionnaires are markedly flawed and validated in a limited way
- Consider providing care via telemedicine where possible, such as with electronic questionnaires and sleep diaries.
- Be aware of the potential first night effect of sleeping with a device, the night-to-night variability in different sleep parameters and the limitations of different measurement types, that may affect clinical diagnosis.Currently, manual scoring is still needed for sleep study analysis and automatic analysis alone is not recommended for clinical use.

- Currently, wearable and non-wearable data is not accepted for clinical use and patients bringing such data to their healthcare personnel need to be educated about the limitations of this data.

Synopsis:

This paper provides an overview of the current use, limitations and future directions of the variety of subjective and objective sleep assessment available. We argue for various ways and sources of collecting, combining and using data to enlighten clinical practice and the sleep research of the future. We highlight the prospects of digital management platforms to store and present the data, the importance of co-design when developing such platforms and other new instruments. We also discuss the abundance of opportunities that data science and machine learning open for the analysis of the data.

INTRODUCTION

Sleep assessment depends both on the subjective experience of the individual and objective measurements which are traditionally collected through an overnight sleep study. In addition, wearable and non-wearable devices are increasingly being used to collect objective data over a longer period and thus offer new ways to assess sleep and its long-term effect on health and well-being. The various types of data all tell an important story for the sleep diagnosis as each type represents a different side of the same coin. Therefore, to have the most complete picture of someone's sleep, these various types of data need to be considered and analysed, as has been shown in other areas of research as well.[1]

DIFFERENT TECHNIQUES AVAILABLE TO MEASURE SLEEP

Subjective assessment

A crucial part of a sleep assessment is the subjective experience of the patient, which in many cases is sufficient to make a clinical diagnosis without the need for any objective sleep testing. This includes the diagnosis and treatment of e.g., insomnia (difficulties initiating or maintaining sleep) and restless legs syndrome (an urge to move legs and/or uncomfortable/unpleasant[2] sensation in legs) as well as numerous other sleep disorders.[3] For other disorders, such as obstructive sleep apnea (OSA), the subjective experience of the patient is usually very limited. A person typically only knows that he or she snores or has apneas during sleep if told by a bed partner. Therefore, in OSA, an objective sleep measurement is necessary for diagnosis.[3] The subjective part is often overlooked and dismissed as less important than objective measurements of sleep, especially by those less experienced in the relevant sleep disorder in question.

<u>Questionnaires</u>

A vast number of questionnaires have been created in the last decades to assess sleep, sleep problems and their effects on daytime functioning. Shahid and colleagues reviewed over a hundred of sleep questionnaires and an extensive number of others

are additionally used.[4] Currently, the most validated and popular screening questionnaire for OSA is the STOP-Bang, which stands for Snoring, Tired, Observed, Pressure, Body Mass index, Age, Neck Size and Gender.[5-8] The actual list of available OSA screening questionnaires is much larger and includes the Berlin questionnaire[9] and The Neck, Obesity, Snoring, Age, Sex (NoSAS) questionnaire.[10] Le Grande and colleagues identified 21 instruments designed to assess the likelihood of OSA for in cardiac patients.[11]To assess excessive daytime sleepiness, the Epworth Sleepiness Scale[12-14] is most popular but others such as the Karolinska Sleepiness Scale[15] and Stanford Sleepiness Scale[16] are also widely used. For overall sleep quality, the Pittsburgh Sleep Quality Index[17] with 19 items in seven subcategories is frequently used but other exist as well.

Limitations

The majority of sleep questionnaires are markedly flawed. Many of them are not properly validated against a diagnosis of a clinical sleep disorder, the asked time frame varies widely and many use weasel wording such as “Recently” and “Frequently” instead of using a specific time frame for answering. Also, often a “Don't know” option is not included.[18] This makes it difficult for many people to answer questions about occurrences during sleep such as snoring frequency and loudness, leg kicks and other movements, especially for those who do not have a bed partner. It is then impossible to know whether a person forgot to answer or did not know how to answer specific questions. Other questions may be poorly designed as they are “double-barrelled”, asking about more than one issue in the same questions. E.g., in the Pittsburgh Sleep Quality Index, the question *“During the past month, how often have you had trouble sleeping because you cough or snore loudly?”* addresses two separate issues.

Some questionnaires are biased towards male responses. The Epworth Sleepiness Scale measures the ability to fall asleep or doze in different situations. The Epworth captures sleepiness better for males, relating more strongly with reported feelings of sleepiness or being unrested, than for females.[19] Females also less often have a total Epworth score >10, which is the clinical cut-off for excessive daytime sleepiness, despite females reporting feeling sleepy as often as males.[19] Females are less likely to report classical OSA symptoms such as snoring and witnessed apneas than males.[20] However, they are more likely than males to report tiredness, sleep onset

insomnia and morning headaches.[20] The results of this are numerous, including screening questionnaires such as the STOP-Bang being less sensitive to female OSA and the need for sex-specific screening.[21] STOP-Bang additionally has one extremely poorly worded question: *"Gender = Male? Yes or No"*. In the authors experience, many females are offended by the setup of this question and a better worded question for addressing gender in modern society is sorely needed.[22]

Finally, the reliability and validity of different sleep questionnaires have often been assessed to a limited extent.[4,23] Reliability assessment includes internal consistency, test-retest reliability and when appropriate inter-rater reliability. The validity assessment includes content and construct validity. A questionnaire that has been validated e.g., in a general population may have very different sensitivity and specificity to detect the relevant sleep problem than in a sleep clinic population.[23] The format of the questionnaire is important as well, whether is administered by interview or is self-administered using paper-and-pencil or digital format.[4,24]

To capture the different aspects needed to understand a person's sleep pattern, a clinician or researcher needs to mix and match from the large number of available questionnaires.[4] E.g., a number of different questionnaires is needed to assess a person's regular sleep-wake patterns, daytime sleepiness, as well as screen for common sleep disorders such as OSA and insomnia. Other important items such a smoking history, exercise level, caffeine consumption, co-morbidities and medication need to be assessed by a different set of questions. Therefore, the patient burden is increased with longer times to answer repetitive questions in different sections and with varying question formats and instructions in a single test battery. Some efforts to develop a common test battery based on compilation of questionnaires, such as the SAGIC questionnaire[25] and the Western Australian Sleep Health Study Questionnaire have been published but validation of their psychometric properties is still lacking.

Future potential

Most questionnaires are currently designed for paper-and-pencil answering and inter-method reliability needs to be validated for digital format. Also designing questionnaires specifically for digital format allows for a different structure for follow-up questions. E.g., now a broad general question about a specific category can be

asked. If the patient has no perceived issues in the category, he or she can move onto the next category but if an issue is detected, a much deeper probing can be performed. Also, further screening questionnaire development such as the STOP-Bang and NoSAS score[10] designed to exclude OSA in subjects who have low probability of disease and do not need further objective testing, is needed. The reliability and validity of further questionnaire development need to be assessed for different populations, including clinical and general population, females and males separately as well as in different ethnic groups and age categories. These questionnaires should be co-designed with both patients and healthcare workers (see details below) and be reviewed thoroughly by relevant experts for the different design flaws described above.

Interviews (structured, semi-structured etc)

The gold standard sleep assessment includes an interview with a qualified sleep physician (or somnologist). European standards indicate a 1-hour interview with a sleep physician to review patient history, a physical exam and review of questionnaires.[26] The importance of conducting a detailed clinical interview including a physical examination is highlighted in the Sleep Medicine Textbook of the European Sleep Research Society (ESRS).[27] The American Academy of Sleep Medicine (AASM) has also provided a detailed list of screening questions for the sleep history and physical examination on their website. It is, however, unclear who is responsible for this text, date of publication and whether the interview has been validated.[28] The structure and length of the clinical interview will likely differ widely between countries and sleep laboratories.

A few structured interviews, including the *Diagnostic Interview for Sleep Patterns and Disorders (DISP)*[29] and the *Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) Sleep Disorders (SCISD)*[30] have been generated. These can be administered by a trained interviewer without sleep expertise, take about 10-30 minutes, and have been validated to some extent.[29,30]

Limitations

Access to a qualified sleep physician is scarce and waiting lists worldwide for clinical interviews are typically very long. Therefore, the majority of patients complaining of

sleep problems will never meet such a specialist. Further validation of structured interviews that can be administered by trained non-sleep experts, is needed to facilitate a wider use of such instruments.

Future potential

The generation of a simple-to-use, clinically validated structured clinical interview as a test battery for a variety of sleep problems to improve the patient care is highly important. This clinical interview needs to assess sleep hygiene, insomnia, delayed sleep phase, OSA as well as other potential sleep disorders. Such a tool would allow general practitioners, cardiologists, psychologists and other healthcare professionals to thoroughly assess a person's sleep profile and direct them to the needed diagnostics tests and treatments to improve their sleep. Patients with continued problems could then be referred to a more specialized sleep expert. Preferably the structured clinical interview could be available online, free of charge, for all healthcare professionals to use, providing automatic calculations of the risk for different sleep disorders, with guidelines for the healthcare professional for the next steps. Finally, a comparison between the validity of a structured clinical interview and a digital self-administered questionnaire with the same items would be advisable. If a digital questionnaire has similar validity to the structured clinical interview, valuable healthcare personnel time can then be saved for other purposes.

We would also like to emphasize the need for in-depth, structured or semi-structured interviews, observations and focus groups to gather empirical data to improve the different tools used for subjective measurements described in this chapter. These interviews and focus groups transcripts could then be analysed through thematic analysis, or through content analysis[31] to shed light on new patterns in sleep disorders or subjective feelings which could be revealed through such data gathering. Thematic analysis includes closely examining the data to identify recurrent and common themes. These themes can consist of topics, ideas or patterns that appear frequently. Although the process can differ, typically it is divided into six phases: i) familiarisation, ii) coding, iii) generating themes, iv) reviewing themes, v) defining and naming themes and vi) writing up.[31-33]

Sleep diaries

Sleep diaries have an important role in the diagnosis and treatment monitoring of insomnia and another sleep disorders such as circadian rhythm disturbances.[34] The typical sleep diary is used for 1-2 weeks to assess a person's overall sleep length, timing and quality as well as factors that may affect sleep such as daytime naps, sleeping pills, caffeine and alcohol use. The person answering then estimates last night's sleep the following morning by answering questions about when they went to bed, how long it took them to fall asleep and if they woke up during the night. The sleep diary gives an important overview to the sleep expert of the overall subjective sleep experience of the patient or research subject.

However, there is also a consensus regarding the need to standardize sleep diaries as many different versions are available and used by different sleep experts.[35,36] There are various aspects that vary currently within the literature reporting on sleep diaries: i) the wording of questions, ii) the number of questions iii) the format of delivery (see Figure 1) iv) the length of the data collection (typically performed for 1-2 weeks but can vary further, v) answering once or twice per day and vi) whether it should be paper-and-pencil or digital format.[36-38] Therefore, the generation of the Consensus Sleep Diary in 2012[36] was a major step towards standardization of this important tool.

Limitations

Even the Consensus Sleep Diary generated by an expert panel includes three different versions, with a core part designed to fit two sides of a single sheet of paper and an extended, optional version.[36] Focus group feedback indicated that some participants preferred a graphical format, such as clock faces or time charts. Also, participants commented that specific aspects of the sleep of a given night could not be well described.[36] Therefore, as indicated by the authors, further ways to standardize and improve sleep diary assessment is needed.

A very important limitation to sleep diaries is compliance, i.e., the participants forget to fill in the sleep diary. Some patients and research participants may then attempt to fill in entries for several days in a row, which is especially bad in a paper-and-pencil format where there is no way to know when the participant did the assessment.

Another limitation is the subjective nature of the sleep diary which can be affected by e.g., memory bias.

Future potential

Future potential could include sleep diaries in mobile applications (or app). Using smartphones to deliver sleep diaries through mobile apps, could be a way to i) ensure more compliance in answering through nudging the participants with screen notifications, ii) know when the participants fill in the sleep diaries, iii) limiting when an event can be registered. An additional aspect that could be added to sleep diaries is the impact of continuous self-assessment on patient self-care and empowerment, by evaluating whether the participants change their lifestyle over time, through being asked about their lifestyle every day. Future potentials include adding questions that on the subjective experience of the participant and regular objective alertness tests. This could give a person insight into how different aspects such as short sleep duration affects their next day functioning or excessive caffeine/alcohol use affects their next night sleep.

Objective assessment

The number of objective ways to assess sleep and sleep disorders is high (see Figure 2). The gold standard diagnostic method is generally considered an in-laboratory polysomnography (PSG).[39,40] However, this method is not the right tool for every sleep problem and cannot measure the subjective experience of the patient as described above. A PSG without any subjective information in many cases may not yield any meaningful clinical results for a given patient, e.g., for the diagnosis of insomnia.[41]

Type 1-4 sleep studies

The gold standard sleep study is considered an attended, in-laboratory PSG, or type 1 sleep study.[2] This refers to the patient sleeping in a hospital or laboratory environment, monitored by staff the whole night to ensure the quality of the study. The study includes six channels to measure electroencephalography (EEG) for brain wave activity, left and right electrooculography (EOG) to measure eye movements and chin electromyography (EMG) for muscle tone. Two reference channels are added on the

mastoid bone behind both ears for the EEG and EOG assessment. Together, the EEG, EOG and EMG allow for the assessment of different sleep stages and wake periods measured in 30 second epochs throughout the night and arousals from sleep. PSG also includes respiratory flow assessment via a nasal cannula and thermistor to assess nasal and mouth breathing. Respiratory movements are measured via thorax and abdomen belts and oxygen saturation via pulse oximeter. Together, these measurements allow for assessment of sleep apnea severity and subtypes (obstructive, mixed or central). Additionally, an electrocardiography (ECG), leg EMG (for periodic leg movement assessment), body position, synchronized video and audio is included. The PSG is manually edited or "scored" for different events including sleep stages, arousals, respiratory events, periodic leg movements etc. This scoring is described in detail in "The AASM Manual for the Scoring of Sleep and Associated Events"[39], the most widely used standard for sleep scoring albeit some major differences exist in scoring standards worldwide that impact all attempts for automation of sleep scoring. For more details, see review papers.[42,43]

Other types of objective sleep studies are referred to as home sleep apnea testing (HSAT) by the AASM. Therefore, this term includes every type of sleep study performed at home/portable, which are in fact very diverse in nature. Another terminology which is more detailed is Type 1 - Type 4 sleep studies. Type 1 is then an attended in-lab PSG and types 2 - 4 are described below and in Figure 2.

A type 2 sleep study is a home PSG which includes all the same channels as an in-lab PSG except for the video recording and depending on the device used, an audio may be included or not. Typically, the patient goes to the sleep laboratory where a sleep technologist sets up the study (takes about 1 hour), and the patient is then sent home to sleep.[39]

A type 3 sleep study is also called a polygraphy, portable monitoring or cardiorespiratory study. A type 3 study has the same channels as recorded in a PSG excluding the sensor required for an actual sleep measurement (EEG, EOG and chin EMG). The study comprises of sensors to assess breathing, oxygen saturation, pulse and body position. In some devices, additional sensors such as EKG and leg EMG can be added. Type 3 sleep studies are currently considered for OSA diagnostics in

patients with a high pre-test probability.[39] However, the definition of what is considered a high pre-test probability is often not clearly stated.[40] Some papers have defined a high pre-test probability using OSA screening questionnaires which have marked flaws as discussed above.[44,45]

A type 4 sleep study a limited-channel monitoring with 1-3 parameters. One type of such studies, which includes peripheral arterial tone (PAT), pulse oximetry and wrist activity is accepted by the AASM[39]. Guidelines from the European Sleep Research Society (ESRS) and Assembly of the National Sleep Societies (ANSS) in Europe do not accept the use of such studies for a final diagnosis of any sleep disorders albeit an update of these guidelines is needed since they are now almost 10 years old.[46]

Limitations

A major limitation to all sleep studies is that they are typically only applied for a single night, both for clinical and research studies. Therefore night-to-night variability in different sleep parameters and the potential first night effect of sleeping with a device are not measured.[47-49] Type 1 studies have also been criticized as the 'gold standard' measurement due to the unnatural sleep environment of the patient, who is expected to sleep normally during one night in a hospital environment with none of the normal home routines for the evening or night. Also, due to the number of wires attached to the patient and often to a device on the bedside, patients may sleep more in the supine position than normally. This can e.g., increase OSA severity in patients with supine OSA to clinical levels, albeit the patient would only sleep non-supine at home.[50] Type 2 - 4 sleep studies are more likely to have some quality issues than type 1 studies as some issues may come up before the patient goes to sleep and or during the night as they are not continuously monitored.[51] Type 3 and 4 sleep studies lack measurement of actual sleep. Therefore, the total sleep time and wake periods cannot be assessed except in some devices via surrogate markers.[52,53] Also, phenotypes such as REM-related OSA cannot be assessed.[48,49] Further work to convince the majority of sleep clinicians that these surrogate markers are adequate to measure sleep on an epoch-by-epoch basis are needed.[43] Also, arousals as markers of sleep fragmentation cannot be assessed. This additionally affects other event scoring e.g., hypopneas followed by arousals without oxygen desaturation,[39] which impacts the measured OSA severity. Type 4 sleep studies then additionally lack measurement of actual respiration and rely

on surrogate markers to assess breathing stops and typically rely on automatic analysis, albeit some efforts for manual scoring have started.[52] Type 4 sleep studies relying on PAT technology have on average a very high accuracy compared to PSG,[54] but care should be taken for the individual patient results, which can differ extensively from the 'gold standard' measurement. Again, event-by-event comparison needs to be performed and published to adequately validate this technology. Furthermore, special care needs to be taken for patients with different comorbidities and medication use in choosing relevant patients for both type 3 and type 4 sleep studies, again highlighting the role of a thorough subjective assessment.[2]

Finally, type 3 and type 4 studies can be marked by the 'Law of the instrument'[55], to be over-reliant on a familiar tool or a narrow skill set, especially for physicians who are not sleep specialists and have a limited knowledge of other sleep disorders or how to screen for them. Performing such a sleep study may result in incidental findings of OSA, which are not the main complaint of the patient.[42]

Future potential

Some efforts have been made to create self-applied PSGs,[56-58] which the patient can then apply him/herself at home saving valuable clinic time. Self-applied PSG could also allow continued sleep services during periods such as the Covid-19 epidemic which mostly halted in-lab services and PSG setups.[59] Self-applied setups, however, do require extensive testing prior to their acceptance in clinical use. The EEG is typically located only on the forehead (F-channels), with no electrodes located on the top of the scalp (C-channels) and back of the head (O-channels) as in traditional PSG.[39] Research on the number of failed sleep studies and a quality review of signals compared to home PSG set up by a sleep technologist needs to be done as well as patient feedback collected for optimizing the setup (see Co-design below for details). Self-applied setups could also facilitate the use of multi-night studies to capture night-to-night variability and would then be less costly than in-lab or home PSG setup by sleep technologists.

In an era when we have started to rethink many of the current standards used in sleep scoring, such as the fixed 30 second sleep epoch[60] and the counting of respiratory events by the apnea-hypopnea index (AHI) to measure OSA severity,[42,61] type 3 and

type 4 studies may have limited value in the future as they were validated against potentially outdated diagnostic. A type 1 or 2 PSG is important for current research purposes to understand better the essential signals needed for improved future sleep diagnostics.

Actigraphy

To assess “free-living” sleep conditions, multiple night recordings in the home environment need to be performed.[62] Actigraphy studies outline a methodological approach for inferring both sleep and wake patterns primarily based on movement data from small sensors.[63] In the medical field, studies that use actigraphy’s are typically accomplished by using a wrist actigraphy, a small watch-like device embedding an accelerometer, and that often also records ambient light and skin temperature.[62] The use of actigraphy is accompanied by a subjective sleep diary like those discussed above. Clinical guidelines recommend that the subject wears the actigraphy for 7-14 days, but 72 hours of recording is generally sufficient to bill for the testing in the United States.[62] For research purposes, 5-7 day of actigraphy measurement is also often used to assess sleeping behavior.[64] This data can then be used to assess e.g., average sleep duration, chronotype (morningness vs. eveningness) and other sleep parameters of interest.

Limitations

In an epoch-by-epoch validation study of actigraphy against PSG, it was shown that they are a useful and valid means for estimating total sleep time and wakefulness after sleep onset but are limited in terms of specificity.[65] Also, an actigraphy will not measure sleep stages or arousals. In studies with actigraphy, patients need to always wear the watch, while also keeping a sleep diary, which requires high compliance.[66] False positives are also quite common, that is, the watch is good at detecting sleep but worse at detecting wakefulness.[65] Finally, actigraphy is prone to technical malfunctions which as a result, can entail lost data which will not be discovered until the patient or research participant returns the device for downloading.[66]

Future potential

A systematic literature review showed that the accuracy of actigraphy devices is often not significantly different from PSG, but they have a high variability for the same individual. In contrast, EEG-based devices are both more accurate and have less variability, whereas devices that measure behavioural aspects of sleep onset consistently overestimated their PSG counterpart.[67] Since actigraphy is essentially wearable devices, there is much potential in equipping them with additional functionalities, such as sweating and heartrate variability or even using them as an addition to other type of wearables. However, adding any additional measure could improve their accuracy but would require them to be validated again.

Wearables and non-wearables

Self-trackers collect vital signs including sleep measurements on a continuous basis for millions of individuals in an unprecedented way. Wearable technology outlines an umbrella term for body-worn sensors with the ability to send and receive data that can be exchanged between the sensor and a network, such as smartwatches which capture the same information as an actigraphy, while also collecting a wider range of physiological signal data such as pulse.[68-70] Other connected devices, called non-wearables, are placed in a location near the body and used to monitor physiological signals. They include e.g., connected mattresses to monitor sleep patterns.[71] These are growingly used in conjunction with wearables in health-related research studies.[72,73] The use of wearable and non-wearable self-trackers is becoming increasingly popular among the general public for personal monitoring of health and fitness and measure vital signs on a continuous basis. As such, the devices are collecting unprecedented volumes of individuals' vital signs at a very granular level, including movements, physical activity, step count, heart rate, sleep duration and sleep quality, 24 hours a day, seven days a week. The devices typically come with a mobile application, designed and distributed by the company that produces the self-trackers, where the user can see their various measurements and trends over time. In some cases, aggregated data can be retrieved for research purposes with the consent of the participants.

Limitations

To assess the capabilities and accuracy of self-trackers to measure sleep, they need to be validated. To achieve this, they are compared to the gold standard PSG or actigraphy in a double setup either in a clinical setting or free-living environment for one to three nights.[74-76] Numerous validation studies exist, as well as a few meta-studies that systematically investigate the results of such validation studies. The results show that wearables tend to overestimate sleep time and sleep efficiency compared to PSG.[77,78] A systematic review of wearables' estimation of sleep onset showed that the estimate is adequate.[79,80] Further epoch-by-epoch analysis of sleep stages and event-by-event analysis of e.g., breathing stops is needed to validate these devices and for them to be accepted in clinical use.

It is worth noting that these sensors, independent of whether they are body worn or kept close to the body, are consumer products that are designed and developed for the general public and not for specific sleep disorders or for treatment purposes. The same holds for the accompanying digital platforms. What is concerning is that there are no standards, neither regarding the way the data is collected, nor in the way the data is analysed and visualized for the customers using the self-trackers. The algorithms used to analyse the data are black-boxes that are usually not disclosed since they are considered 'business secrets' in this competitive market.[67] There is a certain secrecy in the way the companies estimate sleep stages, and the sleep stage estimation varies between the companies so depending on the company's definition behind the algorithms, the sleep stage can vary. In addition, processes and algorithms are frequently updated without the knowledge of the user, which means that the results shown in the digital platform may not be comparable over time. Moreover, there is an algorithmic bias[81] in the analysis generated by the companies providing the sensors. However, if each participant wears the same device through the entirety of a study, and as long as the same type of devices are compared, they can be useful.

Another point of importance is the potential for such self-trackers to induce anxiety or obsession with sleep with an overreliance on this data. This can cause the nicely coined "orthosomnia.[82]

Finally, there is a clear discrepancy between the various types of self-trackers in the way they collect data and display it to the user.[83] What is also alarming is the gap between the sleep community and the wearable technology companies regarding the way sleep is measured and communicated to the user. This gap could be bridged with

more communication between the stakeholders and adoption of standard metrics for validation of consumer self-trackers.[67]

Future potential

Self-trackers collect vital signs including sleep measurements on a continuous basis for millions of individuals in an unprecedented way. Although the devices are designed for personal monitoring, people are sharing their measurements with their physicians. Given proper guidelines and standards in addition to insightful visualization and meaningful statistical summaries of the various measurements, they could interpret events and trends in the data and use it in decision support for the patient's care. Properly validated wearables and non-wearables will allow us to assess changes in sleep patterns over extended periods, including seasonal effects on sleep[84] and sleep apnoea severity.[85] Changes with age and weight increase, alcohol and caffeine intake, exercise etc. can then also be studied in much more detail than current measures allow. If this is done in conjunction with sleep diaries and questionnaires, very valuable data on sleep can be collected from any given individual. When using wearables and non-wearables, the sensors generate low-resolution data.[86] Consequently, because the data from these sensors does not capture all the vital signs that a PSG does, the data gathering needs to include an extended period of time.

TOOLS FOR MAXIMIZING FUTURE SLEEP MEASUREMENTS

Data management platform – for improved clinical care and research

The term "digital platform" indicates that the platform is a piece of software relying on resilient hardware, while it is also an intermediary that connects needs with resources; sellers with customers, users with service providers or patients with healthcare professionals.[87,88] A digital platform is an organizational, technical and regulatory construct that facilitates value exchange and value creation. Such constructs are especially interesting in a healthcare context to facilitate data exchange and data sharing. The types of digital platforms, where the main objective is data sharing, are termed data management platforms.

For data management platforms to be utilized in sleep research, four main aspects are of importance. These are scalability, layered-modular architecture, security and level of access. Regarding scalability, the era of big data is pushing the limits of size. Big data is today not merely an impressive, exciting concept of the future, instead we are already there, gathering big data on patients and research participants for healthcare purposes.[89] To handle this scalability, the literature suggests designing and developing the infrastructure of data management platforms through a layered-modular architectural approach.[90,91] Such architecture entails dividing the platform into a content layer, service layer, network layer and device layer. To date, digital platforms have mostly been designed, developed, and used for two main purposes: i) development platforms and ii) transaction platforms and the architecture of them shares similar trades, such as the layers outlined above. Development platforms are designed for the purpose of facilitating app development whereas transaction platforms, are designed for the purpose of facilitating various types of transactions,[92] where one such transaction can be data. Less focus has been on the design and development of digital platforms, meant for healthcare settings and designed for the purpose of sharing data specifically. However, the literature on digital platforms where the main transaction is data, is growing.[88]

Limitations

In healthcare, a large area of improvement is needed in relation to security and level of access while sharing data in general and while sharing the various types of data such as specified in this paper which varies in size, granularity and type. In addition to that, the data discussed herein originates from various types of sensors and devices, which makes the sharing of data, to a shared location, increasingly complex and that is where security and level of access comes in. Additionally, sharing data between end-user groups with diverse needs is also a known limitation of most healthcare systems. E.g., the electronic patient record is designed to serve a documenting and administrating purpose for healthcare professionals but has later been further extended to include sharing options for patients accessing data, at least in some countries. This divergence in design versus use is not optimal. Instead, it is truly important to define the end-users into stakeholder groups upfront and to structure the architectural choices accordingly. Few attempts have been successfully executed

where data can be securely shared between researchers, healthcare professionals and patients through the same digital platform within healthcare, where the level of access is controlled carefully, and that is where the potential of data management platforms comes in.

Future potential

Future potentials include designing and developing secure data management platforms for sharing data on sleep between three main types of end-users: i) researchers, ii) healthcare professionals and iii) patients and research participants. These end-user groups then have access to a subset of the data, depending on their level of access. This type of architecture can be used to support research through connecting various kinds of data, collected by healthcare professionals and patients into one common place. That is why we would like to argue that data management platforms are at the nexus of where the future of sleep research and sleep measurements will be, but securely through layered-modular architecture and tight access control. Furthermore, data management platforms can be used to display a variety of data, gathered through different types of devices and methods, in the same platform. Also, the same data presentation for devices from different manufacturers can then be visualized to facilitate their easier use for healthcare professionals, focusing on the information needs of this end-user type.

Importance of co-design and co-care

There is a history of failed implementations of large-scale digital infrastructures in healthcare settings. These digital infrastructures have been pushed down the organizations in top-down manner and one such example is the electronic patient records, which suffer from usability issues in most countries.[93-95] Based on that, when designing and developing digital technology, such as the data management platforms elaborated on above, with the purpose of supporting everyday life and work, the design processes can differ substantially. The design process can on the one hand be done with detachment from the end-users (i.e., the designated users of the digital technology), of which the electronic patient record is a famous example, or on the other hand the design process can be conducted through engagement with the end-users.[88]

There has been a longstanding focus on end-user participation and engagement in the design process in research,[96] whereas in practice, the detachment paradigm is sometimes considered the most effective way forward in costly processes of large-scale digital innovation. However, we would like to argue that when designing and developing digital technology to support a group of heterogenous end-users, the most effective way in coping with the complexity of the needs of the various stakeholder groups, is through end-user engagement.[97] When involving the end-users, there must be a dual focus on empowerment of the end-users and on generating focused ideas which can be handed over to the developers. This can be done through co-design.

Co-design refers to a collaborative creative activity where end-users, who are not trained in design work, engage with designers on ideas in order to further the design process and thus incorporate the specific needs of the end-users early on.[98] More specifically, in co-design the design process is regarded as a specific aspect of co-creation where the end-users are engaged as co-designers (i.e., as collaborative agents or actors in the design process) where it is essential that the end-users are seen as a valuable resource to further the design process.[98] Involving the end-users in design is not new. In participatory design it has been the guiding philosophy for half a decennium. An important building block in the move towards participatory design was written already in 1972.[99] The fundaments of co-design as an approach thereby entail the end-users having a voice in the design processes that ultimately impact their lives.[100,101] In healthcare settings, where patients, healthcare professionals and researchers intersect and use the same digital infrastructure, in different ways, there is the need to involve representatives from each of those end-user groups, and not just one.

When selecting end-users as co-designers in the care for frail patients' context-related issues can be created, such as related to identity. Accordingly, in a successful co-design process, where patients are involved, it is vital that the end-users can identify themselves as the future end-users of the digital technology that is being designed,[102,103] and that the selection of representatives (i.e., co-designers) is diverse.[97] Additionally, the co-design process can set the scene for co-care, where the patients are participants in their own care to a larger extent because the digital

infrastructure was designed by them, for them.[87,97] Studies also show that healthcare professionals who regard the same system as their own, designed by and for them, are more likely to accept the system and the change it entails.[97]

Limitations

The main limitation of involving various types of stakeholders, is time. It takes time to meet with the representatives of the end-users repeatedly and show them new developments in the design, get their feedback and iterate that feedback into the design. The iterations take time, time which pays off in the long run, but still slows down the process in the beginning.[101] Another known limitation is leveraging various needs and knowing when to favour e.g., the healthcare professionals over the patients or vice versa when they have conflicting views on the design.[97]

Future potentials

The way the design process is orchestrated and organized, is essential to the outcome and acceptance of the digital infrastructure in the long run. Facilitating the collaboration between the patients and their healthcare professionals early on can function as a basis for understanding the future use situation and can foster acceptance for the system, and for the change. Conclusively, design approaches, such as co-design, where genuine user participation is key in the design process, also have to consider the frail end-users and their specific needs in a co-design process, while catering to the needs of the healthcare professionals and ultimately researchers. This makes the complexity high, but a change is needed e.g., in the development of subjective measurements of sleep and the results will likely improve greatly the tools we have for both clinical and research needs, in a process such as shown in Figure 3.

Improved analysis of subjective and objective data

The different types of sleep data available as described above vary greatly e.g., in terms of subjective vs. objective nature, granularity, frequency, timespan, amount of noise and data quality. To extract knowledge and insights from the data, there are several methods, processes, algorithms and systems which fall under the interdisciplinary data science field, also known as data mining and analytics. These include pre-processing the data by removing noise and errors, extracting and

engineering useful and informative variables from structured and unstructured data, finding correlations between different measurements and events, summarizing main aspects of the data to give an overview, visualizing the data in insightful ways and inspecting trends.[104] The cross-industry standard process model for data mining, or CRISP-DM is the most widely used model for these tasks and applies to any domain, including analysis of sleep data. It breaks the cyclic process into six phases which the practitioner is typically required to move back and forward between.[105] The process, seen in Figure 4, usually begins with understanding the needs of the domain where the techniques will be deployed, together with an investigation of the available data. Once the relevant data has been selected, it must be prepared, which involves pre-processing and transforming it to obtain clean target data and having the relevant features extracted. Next comes the actual modelling phase, where statistical and machine learning techniques are applied to discover insightful patterns. The resulting models are then evaluated to measure their performance and interpret the results. When the model is ready, it can be deployed and the insights that were gained from the process used to enhance the domain understanding.

A major component of data science is the modelling phase which usually consists of machine learning, i.e., algorithms that improve automatically through experience or by using large amounts of data to learn from. The main approaches in machine learning are unsupervised learning, supervised learning and reinforcement learning.[104,106] *Unsupervised learning* is used to find similar patterns that can be used to describe common profiles of the observations in the data. Unsupervised learning is also used to detect anomalous observations that can be indications of errors in the data collection, or abnormal behaviour that need to be investigated further.[106] In sleep analysis, unsupervised learning can be used to find sleep patterns in the long-term data of one person or phenotypes in larger datasets well as to e.g., identify sleep stages.[25,107-111] In contrast, *supervised learning* requires labelled data, such as events and class labels of the observations.[82] This approach is used to detect events during sleep, e.g., in OSA to discover intricate relationships between vital signs, events, sleep stages and phenotypes, and predict future occurrences regarding health and well-being.[112-114]

Despite the vast amounts of available data of all types, the sleep research community has only recently started exploiting machine learning and first results indicate their huge potential.[115] With the drastic advancement of technology and increase in computational power, a type of machine learning methods called deep learning have made a lasting mark in areas where complex, fine grained, sequential data is in abundance, and their prevalence continues to grow.[116,117] Deep learning is a synonym for a variety of deep neural network architectures that are made up of several layers of artificial neurons, that pass features of the input data between them while optimizing the importance of the links connecting the neurons in order to learn the structure in the data, either in an unsupervised or supervised manner.[118] In sleep research, deep neural networks have been trained on the various signals from PSG to classify sleep stages, detect OSA, find distinction between people with and without specific symptoms and predict sleep quality, to name a few.[119-121] Deep learning methods require massive datasets during training, but their unprecedented ability to discover patterns in highly complex data make them the prime contender to automatically analyse PSG data.

Limitations

A major limitation when performing supervised machine learning in traditional sleep studies is the need for labelled data. This labelling requires manual scoring by an expert somnologist or sleep technologist, where a sleep measurement is partitioned into sleep stages and various events are detected. It typically takes a trained sleep specialist 1.5-2 hours to score one overnight polysomnography and 30 minutes for a type 3 sleep study.[46] Although efforts have been made to automate such scoring of sleep measurements,[122] the sleep specialist must still go over the whole recordings afterwards to edit the automatic scoring.[39,46]

Another limitation is the need for computational power, especially when working with deep neural networks. However, cloud solutions where models can be trained in reasonable time are becoming more readily available. This then may raise issues of data security, as the data is often quite sensitive. Finally, building deep neural networks with complex data requires expert knowledge and sufficient training, both on the technical side and in the domain where the data comes from. In the absence of such unicorns, it will be necessary for sleep specialists to master the proper computer

science knowledge to work with neural networks or for computer scientists to dive into the domain of sleep analysis. Collaboration between experts in the two fields is also one way forward, which in addition fosters multidisciplinary research.

Future potentials

There is a long tradition of using PSG to diagnose sleep. However, PSG are invasive and time consuming to analyse, as described above. While PSG data is multivariate, complex and in high frequency, it only provides a one-night snapshot of the individual's sleep characteristics. In contrast, data from wearables offers fewer dimensions (HR, steps, sleep stages, etc.), the measurements are less frequent, but they span a much longer time. These are gaining popularity as they offer the possibility to track and analyse sleep and sleeping behaviour for an extended period of time, at an individual level. This can be seen as a single subject (n=1) clinical trial, where an individual patient is viewed as the sole unit of observation in the study. The goal is then to use data-driven methods to determine the optimal intervention for the individual.[123] As the individual uses the wearables for a long time, it is possible to use the data collected to learn the individual's baseline behaviour and then automatically detect patterns or observations that are out of the ordinary and could be signs of illness or cause for an intervention, i.e., with a physician. The widespread use of wearables furthermore enables research on sleeping patterns and behaviours on larger and more heterogeneous cohorts of people than ever before. Recent studies confirm age-related changes in sleep as well as gender differences in sleep patterns.[124]

As mentioned above, there remains much research and development of deep learning methods to analyse sleep data. Current deep learning applications for sleep measurements build on either convolutional neural networks and/or recurrent neural networks and have been shown to work well.[116] However, they are still limited by the scarcity of labelled data, for which active learning and transfer learning could offer a solution. In addition, reinforcement learning could be used in specific cases.

Active learning is an interactive machine learning scheme where the algorithm uses the available labelled data to train the model and then asks the expert using the system to label specific observations to enhance the learning and improve the model's performance.[125] This would be especially useful for scoring sleep measurements,

since the system could repeatedly query the expert for input, which leads to more labelled data, and perhaps provides the necessary distinction where the algorithm struggles, and it can thus focus its learning on the difficult grey areas. In this way, the algorithm actively learns from the expert.

Transfer learning is an approach where a deep learning model that is trained on a data set in a specific domain is applied to another dataset in another domain. In the domain of sleep studies, there may be different devices that measure the same signals in slightly different ways.[126] As mentioned above, deep learning models require a lot of annotated and labelled data, which may not always be available. When much labelled data exists for one device, but not for another, it is possible to train a model to detect events using the richly annotated data, and then apply it to the other dataset although it may be slightly different. This is called transfer learning.

Reinforcement learning is a machine learning approach that learns optimal behaviour by training an algorithm for a given task using trial and error to maximize the long-term reward.[114] It has not been applied in the sleep research community, to the best of our knowledge, but there is potential for it for example for personalization. As data on sleep patterns in continuously collected through wearables, the algorithm that computes quality of sleep can be personalized using deep reinforcement learning. The approach would then accommodate for individual patterns and behaviours and over time get better at assessing different sleep features at a personal level.

Data science offers much potential for the future of clinical practice, and for the future of sleep research, from gathering a wider variety of data to sophisticated modelling with deep learning architectures and insightful visualizations. Herein we go through and suggest active learning, transfer learning and reinforcement learning as concrete paths to take in machine learning with sleep data but also want to forward the importance in developing new types of methods through sleep research as well as focus on longitudinal data collected through wearables.

SUMMARY

In this paper, we have argued for various ways of collecting data to enlighten the sleep research of the future. First, we described the various types of subjective and objective data and discussed their limitations and future potential. Secondly, we highlighted the prospects of digital management platforms to store and present the data, the importance of co-design when designing and developing such platforms and the opportunities data science opens for the analysis of the data. Thirdly, we have demonstrated the importance of considering data from different sources in the care of a patient, since it provides a means of seeing a fuller picture. Most of the data discussed in this paper can be visualized for various stakeholder groups, through clever and insightful visualizations from which the healthcare professionals can draw informed decisions, and researchers can see trends from various sources, together in one visualization.

As we see it, the future of sleep research, and the future of informed clinical practice is through the combination of various types of data. By introducing and combining new approaches from various research domains on the one hand and alongside patients and healthcare professionals on the other hand, we will be able to understand new widths of the impact of sleep and lack of sleep that will impact and inform both research and clinical work for future improvements.

Figures

Name of patient: Andrea	Date: 26/03/2021
1. What time did you get into bed?	22:35
2. What time did you try to go to sleep	23:00
3. How long did it take you to fall asleep?	15 min.
4. How many times did you wake up, not counting your final awakening?	2 times
5. In total, how long did these awakenings last?	40 min.
6. What time was your final awakening?	07:00
7. What time did you get out of bed for the day?	07:15

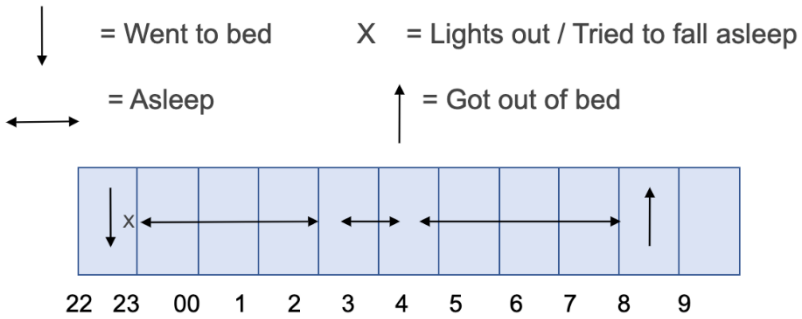

Figure 1: Examples of two different types of sleep diary, one with a quantitative approach where the subject needs to fill in specific times in hours and minutes and a qualitative one where the subjects fill in approximate times in a visual sleep diary.

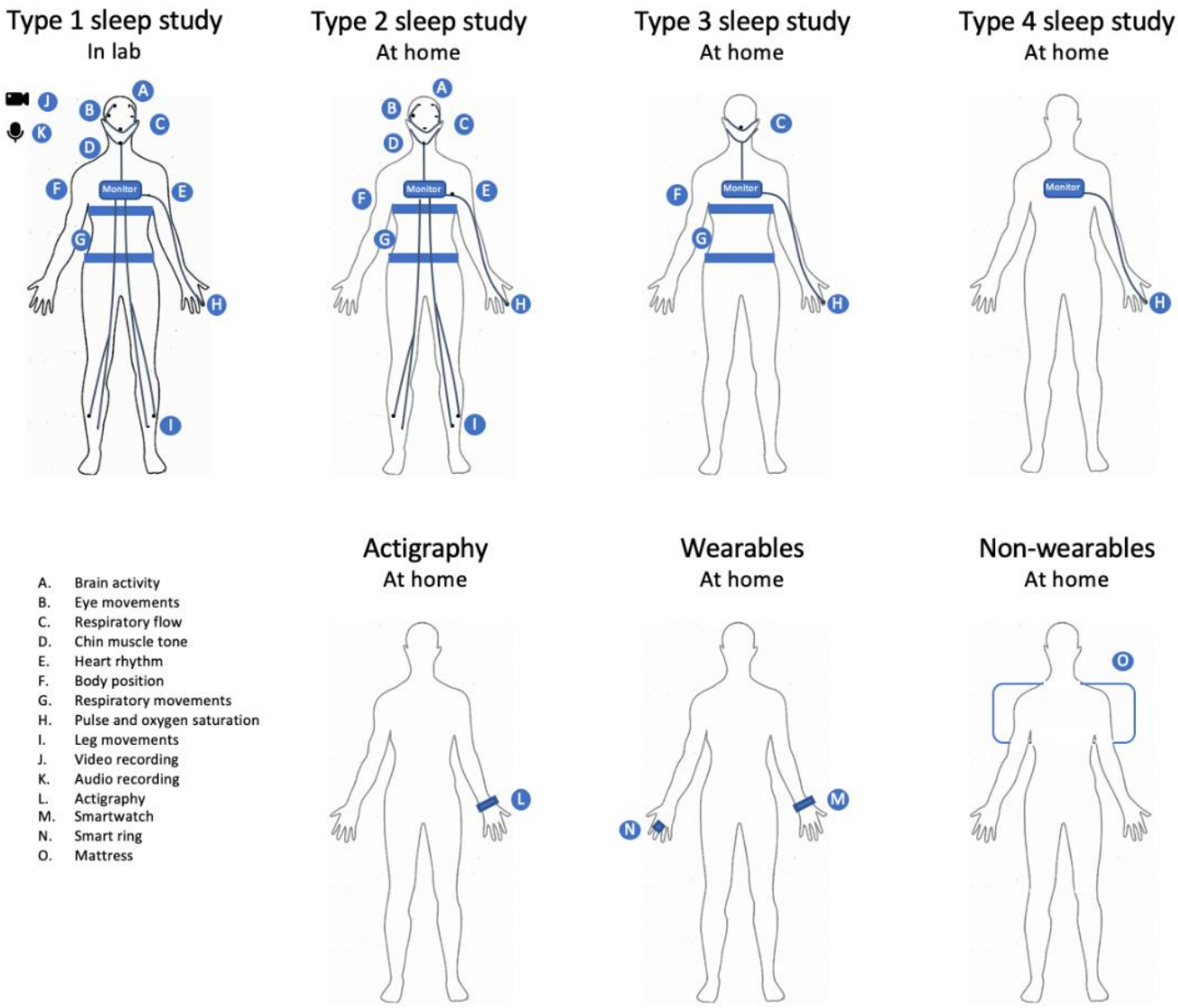

Figure 2: Different types of objective sleep measurements.

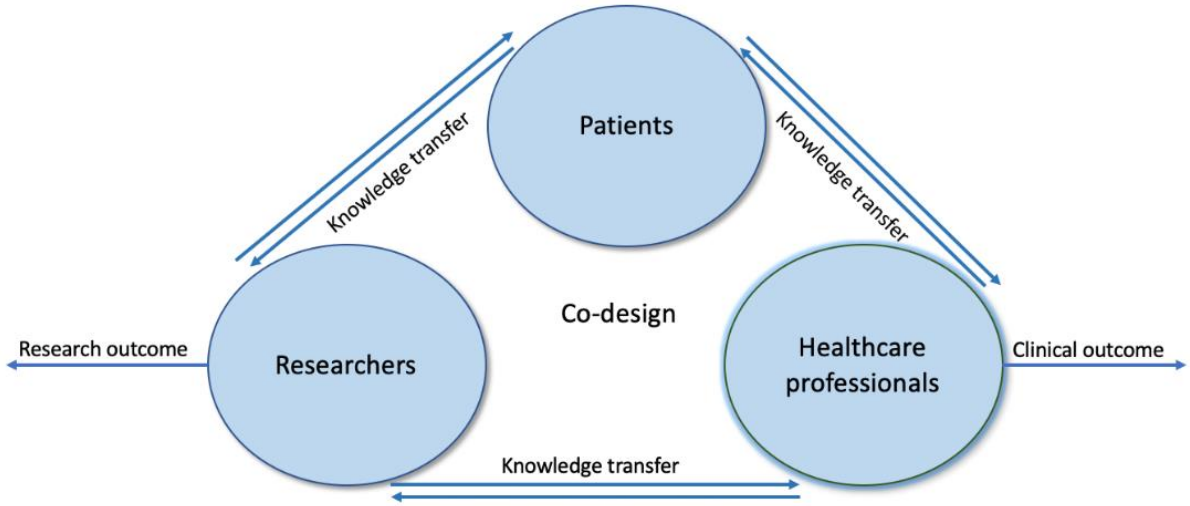

Figure 3: Bedside-to-bench and Bench-to-bedside: the power of knowledge transfer between researchers, healthcare professionals and patients to improve both clinical and research outcomes.

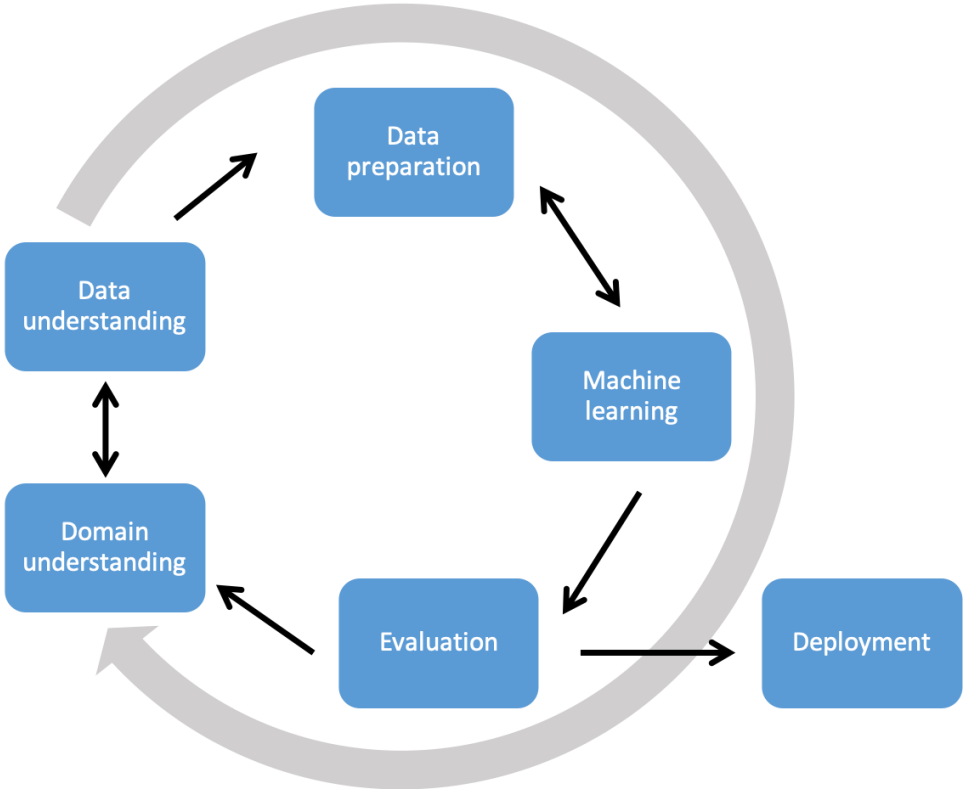

Figure 4: The pathway to improve automatic analysis of sleep measurements with machine learning.

Acknowledgements

We thank Kristin Anna Ólafsdóttir, expert sleep technologist and Dr. Dirk Pevernagie for reviewing the paper and providing excellent suggestions for improvements.

References

1. Óskarsdóttir M, Bravo C, Sarraute C, Vanthienen J, Baesens B. The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. Applied Soft Computing. 2019; 74: 26-39.
2. Douglas JA, Chai-Coetzer CL, McEvoy D, et al. Guidelines for sleep studies in adults - a position statement of the Australasian Sleep Association. Sleep Med. 2017; 36 Suppl 1: S2-S22.
3. *International classification of sleep disorders, 3rd ed.* Darien, IL: American Academy of Sleep Medicine; 2014.
4. Shahid A, Wilkinson K, Marcu S, Shapiro CM. *STOP, THAT and one hundred other sleep scales.* Springer Science & Business Media; 2012.
5. Boynton G, Vahabzadeh A, Hammoud S, Ruzicka DL, Chervin RD. Validation of the STOP-BANG Questionnaire among Patients Referred for Suspected Obstructive Sleep Apnea. J Sleep Disord Treat Care. 2013; 2 (4).
6. Silva GE, Vana KD, Goodwin JL, Sherrill DL, Quan SF. Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth Sleepiness Scales. J Clin Sleep Med. 2011; 7 (5): 467-472.
7. Prasad KT, Sehgal IS, Agarwal R, Nath Aggarwal A, Behera D, Dhooria S. Assessing the likelihood of obstructive sleep apnea: a comparison of nine screening questionnaires. Sleep Breath. 2017; 21 (4): 909-917.
8. Hwang M, Zhang K, Nagappa M, Saripella A, Englesakis M, Chung F. Validation of the STOP-Bang questionnaire as a screening tool for obstructive sleep apnoea in patients with cardiovascular risk factors: a systematic review and meta-analysis. BMJ Open Respir Res. 2021; 8 (1).
9. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med. 1999; 131 (7): 485-491.
10. Marti-Soler H, Hirotsu C, Marques-Vidal P, et al. The NoSAS score for screening of sleep-disordered breathing: a derivation and validation study. Lancet Respir Med. 2016; 4 (9): 742-748.
11. Le Grande MR, Jackson AC, Beauchamp A, Kerr D, Driscoll A. Diagnostic accuracy and suitability of instruments that screen for obstructive sleep apnoea, insomnia and sleep quality in cardiac patients: a meta-analysis. Sleep Med. 2021.
12. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 1991; 14 (6): 540-545.
13. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep. 1992; 15 (4): 376-381.
14. Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth Sleepiness Scale. Chest. 1993; 103 (1): 30-36.
15. Kaida K, Takahashi M, Akerstedt T, et al. Validation of the Karolinska sleepiness scale against performance and EEG variables. Clin Neurophysiol. 2006; 117 (7): 1574-1581.
16. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. Psychophysiology. 1973; 10 (4): 431-436.
17. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989; 28 (2): 193-213.
18. Fedson AC, Pack AI, Gislason T. Frequently used sleep questionnaires in epidemiological and genetic research for obstructive sleep apnea: A review. Sleep Med Rev. 2012; 16 (6): 529-537.
19. Baldwin CM, Kapur VK, Holberg CJ, Rosen C, Nieto FJ, Sleep Heart Health Study G. Associations between gender and measures of daytime somnolence in the Sleep Heart Health Study. Sleep. 2004; 27 (2): 305-311.
20. Nigro CA, Dibur E, Borsini E, et al. The influence of gender on symptoms associated with obstructive sleep apnea. Sleep Breath. 2018; 22 (3): 683-693.
21. Bauters FA, Loof S, Hertegonne KB, Chirinos JA, De Buyzere ML, Rietzschel ER. Sex-specific sleep apnea screening questionnaires: closing the performance gap in women. Sleep Med. 2020; 67: 91-98.
22. Westbrook L, Saperstein A. New categories are not enough: Rethinking the measurement of sex and gender in social surveys. Gender & Society. 2015; 29 (4): 534-560.

23. McNicholas WT. Screening for sleep-disordered breathing: the continuing search for a reliable predictive questionnaire. Lancet Respir Med. 2016; 4 (9): 683-685.
24. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J Anaesth. 2017; 11 (Suppl 1): S80-S89.
25. Keenan BT, Kim J, Singh B, et al. Recognizable clinical subtypes of obstructive sleep apnea across international sleep centers: a cluster analysis. Sleep. 2018; 41 (3).
26. 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.
27. Grote L, Puertas FJ. ssessment of sleep disorders and diagnostic procedures. 2. The clinical interview and clinical examination. In: Bassetti C, Dogas Z, Peigneux P, eds. Sleep Medicine Textbook. European Sleep Research Society. : 2014: 111-123.
28. https://aasm.org/resources/medsleep/(harding)questions.pdf.
29. Merikangas KR, Zhang J, Emsellem H, et al. The structured diagnostic interview for sleep patterns and disorders: rationale and initial evaluation. Sleep Med. 2014; 15 (5): 530-535.
30. Taylor DJ, Wilkerson AK, Pruiksma KE, et al. Reliability of the Structured Clinical Interview for DSM-5 Sleep Disorders Module. J Clin Sleep Med. 2018; 14 (3): 459-464.
31. Boyatzis RE. *Transforming qualitative information: Thematic analysis and code development.* sage; 1998.
32. Guest G, MacQueen KM, Namey EE. *Applied thematic analysis.* sage publications; 2011.
33. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? International journal of qualitative studies on health and well-being. 2014; 9.
34. Buysse DJ, Ancoli-Israel S, Edinger JD, Lichstein KL, Morin CM. Recommendations for a standard research assessment of insomnia. Sleep. 2006; 29 (9): 1155-1173.
35. Morin CM. Measuring outcomes in randomized clinical trials of insomnia treatments. Sleep medicine reviews. 2003; 7 (3): 263-279.
36. Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep. 2012; 35 (2): 287-302.
37. Riemann D, Baglioni C, Bassetti C, et al. European guideline for the diagnosis and treatment of insomnia. Journal of sleep research. 2017; 26 (6): 675-700.
38. Palagini L, Manni R, Aguglia E, et al. Expert opinions and consensus recommendations for the evaluation and management of insomnia in clinical practice: joint statements of five Italian scientific societies. Frontiers in Psychiatry. 2020; 11.
39. Berry RB, Brooks R, Gamaldo CE, et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.6.* Darien, Illinois.: American Academy of Sleep Medicine.; 2020.
40. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. Journal of Clinical Sleep Medicine. 2017; 13 (3): 479-504.
41. *International classification of sleep disorders, 3rd ed.* Darien, IL: American Academy of Sleep Medicine; 2014.
42. Pevernagie DA, Gnidovec-Strazisar B, Grote L, et al. On the rise and fall of the apnea-hypopnea index: A historical review and critical appraisal. J Sleep Res. 2020: e13066.
43. 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.
44. Chung F, Yang Y, Brown R, Liao P. Alternative scoring models of STOP-bang questionnaire improve specificity to detect undiagnosed obstructive sleep apnea. Journal of Clinical Sleep Medicine. 2014; 10 (9): 951-958.
45. Goldstein CA, Karnib H, Williams K, Virk Z, Shamim-Uzzaman A. The utility of home sleep apnea tests in patients with low versus high pre-test probability for moderate to severe OSA. Sleep and Breathing. 2018; 22 (3): 641-651.
46. 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.
47. Roeder M, Bradicich M, Schwarz EI, et al. Night-to-night variability of respiratory events in obstructive sleep apnoea: a systematic review and meta-analysis. Thorax. 2020; 75 (12): 1095-1102.

48. Oksenberg A, Arons E, Nasser K, Vander T, Radwan H. REM-related obstructive sleep apnea: the effect of body position. Journal of Clinical Sleep Medicine. 2010; 6 (4): 343-348.
49. Gabryelska A, Białasiewicz P. Association between excessive daytime sleepiness, REM phenotype and severity of obstructive sleep apnea. Scientific reports. 2020; 10 (1): 1-6.
50. Srijithesh P, Aghoram R, Goel A, Dhanya J. Positional therapy for obstructive sleep apnoea. Cochrane Database of Systematic Reviews. 2019; (5).
51. Andrade L, Paiva T. Ambulatory versus laboratory polysomnography in obstructive sleep apnea: comparative assessment of quality, clinical efficacy, treatment compliance, and quality of life. Journal of Clinical Sleep Medicine. 2018; 14 (8): 1323-1331.
52. Zhang Z, Sowho M, Otvos T, et al. A comparison of automated and manual sleep staging and respiratory event recognition in a portable sleep diagnostic device with in-lab sleep study. Journal of Clinical Sleep Medicine. 2020; 16 (4): 563-573.
53. Dietz-Terjung S, Martin AR, Finnsson E, et al. Proof of principle study: diagnostic accuracy of a novel algorithm for the estimation of sleep stages and disease severity in patients with sleep-disordered breathing based on actigraphy and respiratory inductance plethysmography. Sleep and Breathing. 2021: 1-8.
54. Yalamanchali S, Farajian V, Hamilton C, Pott TR, Samuelson CG, Friedman M. Diagnosis of obstructive sleep apnea by peripheral arterial tonometry: meta-analysis. JAMA Otolaryngology–Head & Neck Surgery. 2013; 139 (12): 1343-1350.
55. Maslow AH. The psychology of science. 1966.
56. 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 Health Inform. 2018; 22 (4): 1124-1132.
57. Miettinen T, Myllymaa K, Muraja-Murro A, et al. Screen-printed ambulatory electrode set enables accurate diagnostics of sleep bruxism. J Sleep Res. 2018; 27 (1): 103-112.
58. 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).
59. Grote L, McNicholas WT, Hedner J, collaborators E. Sleep apnoea management in Europe during the COVID-19 pandemic: data from the European Sleep Apnoea Database (ESADA). Eur Respir J. 2020.
60. Korkalainen H, Leppanen T, Duce B, et al. Detailed assessment of sleep architecture with deep learning and shorter epoch-to-epoch duration reveals sleep fragmentation of patients with obstructive sleep apnea. IEEE journal of biomedical and health informatics. 2020.
61. Randerath W, Bassetti CL, Bonsignore MR, et al. Challenges and perspectives in obstructive sleep apnoea: report by an ad hoc working group of the Sleep Disordered Breathing Group of the European Respiratory Society and the European Sleep Research Society. European respiratory journal. 2018; 52 (3).
62. Ancoli-Israel S, Martin JL, Blackwell T, et al. The SBSM guide to actigraphy monitoring: clinical and research applications. Behavioral sleep medicine. 2015; 13 (sup1): S4-S38.
63. Fekedulegn D, Andrew ME, Shi M, Violanti JM, Knox S, Innes KE. Actigraphy-based assessment of sleep parameters. Annals of work exposures and health. 2020; 64 (4): 350-367.
64. Rognvaldsdottir V, Gudmundsdottir SL, Brychta RJ, et al. Sleep deficiency on school days in Icelandic youth, as assessed by wrist accelerometry. Sleep medicine. 2017; 33: 103-108.
65. Marino M, Li Y, Rueschman MN, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep. 2013; 36 (11): 1747-1755.
66. Sadeh A, Acebo C. The role of actigraphy in sleep medicine. Sleep medicine reviews. 2002; 6 (2): 113-124.
67. de Zambotti M, Godino JG, Baker FC, Cheung J, Patrick K, Colrain IM. The Boom in Wearable Technology: Cause for Alarm or Just What is Needed to Better Understand Sleep? Sleep. 2016; 39 (9): 1761-1762.
68. Swan M. The quantified self: Fundamental disruption in big data science and biological discovery. Big data. 2013; 1 (2): 85-99.
69. Mettler T, Wulf J. Physiolytics at the workplace: Affordances and constraints of wearables use from an employee's perspective. Information Systems Journal. 2019; 29 (1): 245-273.

70. Fox G, Connolly R. Mobile health technology adoption across generations: Narrowing the digital divide. Information Systems Journal. 2018; 28 (6): 995-1019.
71. Depner CM, Cheng PC, Devine JK, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. sleep. 2020; 43 (2): zsz254.
72. Shelgikar AV, Anderson PF, Stephens MR. Sleep tracking, wearable technology, and opportunities for research and clinical care. Chest. 2016; 150 (3): 732-743.
73. Ko P-RT, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. Journal of clinical sleep medicine. 2015; 11 (12): 1455-1461.
74. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2 compared with polysomnography in adults. Chronobiol Int. 2018; 35 (4): 465-476.
75. Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. Sleep Breath. 2012; 16 (3): 913-917.
76. Gruwez A, Libert W, Ameye L, Bruyneel M. Reliability of commercially available sleep and activity trackers with manual switch-to-sleep mode activation in free-living healthy individuals. Int J Med Inform. 2017; 102: 87-92.
77. Scott H, Lack L, Lovato N. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. Sleep Medicine Reviews. 2020; 49: 101227.
78. Baron KG, Duffecy J, Berendsen MA, Mason IC, Lattie EG, Manalo NC. Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep. Sleep medicine reviews. 2018; 40: 151-159.
79. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. International Journal of Behavioral Nutrition and Physical Activity. 2015; 12 (1): 1-22.
80. Kubala AG, Barone Gibbs B, Buysse DJ, Patel SR, Hall MH, Kline CE. Field-based measurement of sleep: Agreement between six commercial activity monitors and a validated accelerometer. Behavioral sleep medicine. 2020; 18 (5): 637-652.
81. Norström L, Islind AS, Lundh Snis UM. Algorithmic work: the impact of algorithms on work with social media. 2020.
82. Baron KG, Abbott S, Jao N, Manalo N, Mullen R. Orthosomnia: are some patients taking the quantified self too far? Journal of Clinical Sleep Medicine. 2017; 13 (2): 351-354.
83. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. J Sports Sci. 2018; 36 (15): 1734-1741.
84. Li W, Bertisch SM, Mostofsky E, Vgontzas A, Mittleman MA. Associations of daily weather and ambient air pollution with objectively assessed sleep duration and fragmentation: a prospective cohort study. Sleep Med. 2020; 75: 181-187.
85. Cassol CM, Martinez D, da Silva FABS, Fischer MK, Lenz MDCS, Bós Â. Is sleep apnea a winter disease?: meteorologic and sleep laboratory evidence collected over 1 decade. Chest. 2012; 142 (6): 1499-1507.
86. Boe AJ, Koch LLM, O'Brien MK, et al. Automating sleep stage classification using wireless, wearable sensors. NPJ digital medicine. 2019; 2 (1): 1-9.
87. Islind AS, Lindroth T, Lundin J, Steineck G. Co-Designing a Digital Platform with Boundary Objects: Bringing Together Heterogeneous Users in Healthcare. Health & Technology. 2019.
88. Islind AS. *Platformization: Co-Designing Digital Platforms in Practice*, University West; 2018.
89. Bragazzi NL, Guglielmi O, Garbarino S. SleepOMICS: how big data can revolutionize sleep science. International journal of environmental research and public health. 2019; 16 (2): 291.
90. Yoo Y, Henfridsson O, Lyytinen K. Research commentary—the new organizing logic of digital innovation: an agenda for information systems research. Information systems research. 2010; 21 (4): 724-735.
91. Hylving L, Schultze U. Accomplishing the layered modular architecture in digital innovation: The case of the car's driver information module. The Journal of Strategic Information Systems. 2020; 29 (3): 101621.
92. Gawer A. *Platforms, markets and innovation.* Edward Elgar Publishing; 2011.
93. Ellingsen G, Monteiro E. Electronic patient record development in Norway: The case for an evolutionary strategy. Health Policy and Technology. 2012; 1 (1): 16-21.

94. Fitzgerald G, Russo NL. The turnaround of the London ambulance service computer-aided despatch system (LASCAD). European Journal of Information Systems. 2005; 14 (3): 244-257.
95. Monteiro E, Pollock N, Hanseth O, Williams R. From artefacts to infrastructures. Comput Supported Coop Work. 2013; 22 (4-6): 575-607.
96. Bødker S, Ehn P, Sjögren D, Sundblad Y. Co-operative Design—perspectives on 20 years with 'the Scandinavian IT Design Model'. In: proceedings from the proceedings of NordiCHI; 2000.
97. Islind AS, Lundh Snis U. From Co-Design to Co-Care: Designing a Collaborative Practice in Care. Systems, Signs & Actions. 2018; 11 (1): 1-24.
98. Sanders EB-N, Stappers PJ. Co-creation and the new landscapes of design. Co-design. 2008; 4 (1): 5-18.
99. Cross N. *Design Research Society Conference. Design participation.* Academy Editions; 1972.
100. Kensing F, Greenbaum J. Heritage: Having a say. In: Routledge international handbook of participatory design. Routledge; 2013: 21-36.
101. Joshi SG, Bratteteig T. Designing for prolonged mastery. On involving old people in participatory design. Scandinavian Journal of Information Systems. 2016; 28 (1).
102. Woll A. Use of Welfare Technology in Elderly Care. 2017.
103. Malmborg L, Binder T, Brandt E. Co-designing senior interaction: inspiration stories for participatory design with health and social care institutions. In: proceedings from the Workshop, PDC; 2010.
104. Baesens B. *Analytics in a big data world: The essential guide to data science and its applications.* John Wiley & Sons; 2014.
105. Shearer C. The CRISP-DM model: the new blueprint for data mining. Journal of data warehousing. 2000; 5 (4): 13-22.
106. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media; 2009.
107. Zinchuk AV, Jeon S, Koo BB, et al. Polysomnographic phenotypes and their cardiovascular implications in obstructive sleep apnoea. Thorax. 2018; 73 (5): 472-480.
108. Ma EY, Kim JW, Lee Y, Cho SW, Kim H, Kim JK. Combined unsupervised-supervised machine learning for phenotyping complex diseases with its application to obstructive sleep apnea. Sci Rep. 2021; 11 (1): 4457.
109. El-Manzalawy Y, Buxton O, Honavar V. Sleep/wake state prediction and sleep parameter estimation using unsupervised classification via clustering. In: proceedings from the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2017.
110. Pien GW, Ye L, Keenan BT, et al. Changing Faces of Obstructive Sleep Apnea: Treatment Effects by Cluster Designation in the Icelandic Sleep Apnea Cohort. Sleep. 2018; 41 (3).
111. Ye L, Pien GW, Ratcliffe SJ, et al. The different clinical faces of obstructive sleep apnoea: a cluster analysis. Eur Respir J. 2014; 44 (6): 1600-1607.
112. Patti CR, Shahrbabaki SS, Dissanayaka C, Cvetkovic D. Application of random forest classifier for automatic sleep spindle detection. In: proceedings from the 2015 IEEE Biomedical Circuits and Systems Conference (BioCAS); 2015.
113. Mendez MO, Ruini DD, Villantieri OP, et al. Detection of sleep apnea from surface ECG based on features extracted by an autoregressive model. In: proceedings from the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2007.
114. Sutton RS, Barto AG. *Reinforcement learning: An introduction.* MIT press; 2018.
115. Mostafa SS, Mendonça F, G Ravelo-García A, Morgado-Dias F. A systematic review of detecting sleep apnea using deep learning. Sensors. 2019; 19 (22): 4934.
116. Fiorillo L, Puiatti A, Papandrea M, et al. Automated sleep scoring: A review of the latest approaches. Sleep medicine reviews. 2019; 48: 101204.
117. Gunnarsson BR, vanden Broucke S, Baesens B, Óskarsdóttir M, Lemahieu W. Deep Learning for Credit Scoring: Do or Don't? European Journal of Operational Research. 2021.
118. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning, vol. 1. In: MIT press Cambridge; 2016.
119. Zhai B, Perez-Pozuelo I, Clifton EA, Palotti J, Guan Y. Making sense of sleep: Multimodal sleep stage classification in a large, diverse population using movement and cardiac sensing.

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2020; 4 (2): 1-33.
120. Korkalainen H, Leppanen T, Aakko J, et al. Accurate Deep Learning-Based Sleep Staging in a Clinical Population with Suspected Obstructive Sleep Apnea. IEEE J Biomed Health Inform. 2019.
121. 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).
122. Penzel T, Conradt R. Computer based sleep recording and analysis. Sleep medicine reviews. 2000; 4 (2): 131-148.
123. Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? Personalized medicine. 2011; 8 (2): 161-173.
124. Jonasdottir SS, Minor K, Lehmann S. Gender differences in nighttime sleep patterns and variability across the adult lifespan: a global-scale wearables study. Sleep. 2021; 44 (2): zsaa169.
125. Settles B. Active learning. Synthesis lectures on artificial intelligence and machine learning. 2012; 6 (1): 1-114.
126. Torrey L, Shavlik J. Transfer learning. In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI global; 2010: 242-264.