

SPECIAL ARTICLES

## Artificial intelligence in sleep medicine: background and implications for clinicians

Cathy A. Goldstein, MD<sup>1</sup>; Richard B. Berry, MD<sup>2</sup>; David T. Kent, MD<sup>3</sup>; David A. Kristo, MD<sup>4</sup>; Azizi A. Seixas, PhD<sup>5</sup>; Susan Redline, MD<sup>6</sup>; M. Brandon Westover, MD, PhD<sup>7</sup>

<sup>1</sup>Sleep Disorders Center, Department of Neurology, University of Michigan, Ann Arbor, Michigan; <sup>2</sup>Division of Pulmonary, Critical Care and Sleep Medicine, University of Florida, Gainesville, Florida; <sup>3</sup>Department of Otolaryngology-Head and Neck Surgery, Vanderbilt University Medical Center, Nashville, Tennessee; <sup>4</sup>University of Pittsburgh, Pittsburgh, Pennsylvania; <sup>5</sup>Department of Population Health, Department of Psychiatry, NYU Langone Health, New York, New York; <sup>6</sup>Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts; <sup>7</sup>Neurology Department, Massachusetts General Hospital, Boston, Massachusetts

Polysomnography remains the cornerstone of objective testing in sleep medicine and results in massive amounts of electrophysiological data, which is well-suited for analysis with artificial intelligence (AI)-based tools. Combined with other sources of health data, AI is expected to provide new insights to inform the clinical care of sleep disorders and advance our understanding of the integral role sleep plays in human health. Additionally, AI has the potential to streamline day-to-day operations and therefore optimize direct patient care by the sleep disorders team. However, clinicians, scientists, and other stakeholders must develop best practices to integrate this rapidly evolving technology into our daily work while maintaining the highest degree of quality and transparency in health care and research. Ultimately, when harnessed appropriately in conjunction with human expertise, AI will improve the practice of sleep medicine and further sleep science for the health and well-being of our patients.

**Citation:** Goldstein CA, Berry RB, Kent DT, et al. Artificial intelligence in sleep medicine: background and implications for clinicians. *J Clin Sleep Med.* 2020;16(4):609–618.

### INTRODUCTION

#### Purpose

The evaluation and treatment of sleep disorders hinges in large part on the use of polysomnography (PSG), which results in the creation of large amounts of *labeled* electrophysiological data. Therefore, sleep medicine is well positioned to benefit from advances that use “big data” to create artificially intelligent computer programs that may lead to: (1) more accurate classification and diagnosis of diseases and disorders, (2) prediction of disease and treatment prognosis, (3) characterization of disease subtypes, (4) precise and automated instrumentation through sleep scoring, and (5) optimization and personalization of treatments, such as positive airway pressure (PAP), all of which will promote patient-centered care.

Until recently, most automated pattern recognition tasks (eg, sleep staging) have relied on rule-based computer programs, which are vulnerable to human error and bias. Computational advances now enable computers to recognize patterns within data without requiring explicitly programmed rules.

Artificial intelligence (AI) refers to the capability of computer systems to perform tasks conventionally considered to require human intelligence, such as speech recognition, decision making, and visual recognition of patterns and objects. In recent years machine learning (ML) has come to dominate AI, such that the terms AI and ML are often used interchangeably, a convention we adopt in this paper. ML algorithms and programs learn patterns by adjusting parameters to improve performance on tasks, such as prediction, classification, dimension reduction,

or clustering. Therefore, they provide powerful tools for understanding relationships within datasets. When datasets are appropriately large, diverse, and representative, the derived models can generalize to other populations.

The large amount of electrophysiological data generated in PSG recordings is an obvious substrate for AI applications. Combined with demographics, genetic information, and behavioral, psychosocial, lifestyle and other biological data, AI approaches hold promise to provide new insights to inform diagnosis and clinical care of sleep disorders.

A second area of sleep medicine primed to benefit from AI is population health. AI has the potential to advance our understanding of the integral roles that sleep and circadian biology play in human health on a large scale. Additionally, the rich, longitudinal, self-generated data collected during the sleep period (eg, PAP download data and wearable heart rate and motion data) are well suited for AI applications, to (1) distill this data into actionable knowledge to improve the practice of sleep medicine for better patient care and (2) effectively analyze this unprecedented amount of signal to inform precision health.

This paper will briefly review AI/ML concepts, discuss current applications of AI in sleep medicine, present potential use cases, and discuss advantages and disadvantages.

#### Artificial intelligence and machine learning

A comprehensive description of AI/ML is beyond the scope of this paper; however, the following discussion of basic principles will help explain the relevance of this technology for sleep medicine.

ML algorithms are computer programs that improve with experience and prior data, without intervention from direct programming commands. Most ML tasks can be divided into supervised learning (learning to map an input  $x$  to an output  $y$ , based on a set of input-output examples [eg, predicting sleep stages from PSG signals]), unsupervised learning (finding patterns or clusters in a set of inputs, with no labeled output variables provided), or reinforcement learning (algorithms learn based on interacting with the environment and receiving penalties and rewards).

Algorithms are developed using a training dataset and tested against a previously unseen or “held-out” test dataset. The use of a held-out test set is required to avoid biased (usually inflated) estimates of how well a model performs, which may happen when the model is overfit to the training data. The performance of ML algorithms relies on representative training datasets and appropriately chosen testing metrics. For example, a ML algorithm designed to evaluate PSG data for obstructive sleep apnea (OSA) would likely perform poorly if trained only on patients with central sleep apnea. Similarly, a ML algorithm trained on a clinic sample of predominantly men with mostly severe OSA would likely perform poorly in a population-based dataset of men and women with a wide range of OSA severity and subtypes. Additionally, ground truth error (“annotation noise”) may yield inaccurate algorithms. For example, inaccurately scored respiratory events in PSG training data may degrade performance of a ML algorithm to detect OSA when applied to *de novo* PSG tests. Performance metrics for an algorithm must be considered against a current gold standard. An algorithm designer may desire perfect accuracy, but this may not be possible depending on the testing metric being evaluated. For example, human PSG scorers rarely achieve 100% agreement. Therefore, when comparing the performance of an automated ML sleep staging algorithm to technologist scored PSG as a “gold standard,” the most appropriate goal for the ML algorithm may be to achieve algorithm-expert agreement comparable to the level of expert-expert agreement.

### Applications in health care and medicine thus far

Over the last two decades, ML algorithms have demonstrated immense potential to unearth clinically relevant patterns in medical datasets. Diagnostic capabilities are most apparent in image interpretation. Numerous radiological,<sup>1,2</sup> retinal,<sup>3,4</sup> digitized pathology<sup>1,5</sup> and skin lesion images<sup>6</sup> have acted as substrates for supervised ML predictions. The massive volume of signals derived from an electrocardiogram and electroencephalogram (EEG) is also well suited for ML analysis for arrhythmia and seizure detection.<sup>7,8</sup> Even home video footage has provided a data stream for ML predictions in autism.<sup>9</sup> Innovative applications of unsupervised learning may provide alternatives to our current diagnostic classification systems by demonstrating previously uncharacterized subtypes within known conditions.<sup>10–12</sup>

Other ML applications focus on outcomes, prevention, and treatment, and have included predictive models of surgical complications and mortality,<sup>13</sup> sepsis development<sup>14</sup> readmission risk,<sup>15</sup> cancer survival,<sup>16,17</sup> and predictors of treatment response. ML driven devices can physically assist in delivering

treatment such as continuous glucose monitors that interact with insulin pumps in a closed loop to better control blood sugar in patients with type I diabetes mellitus.<sup>18</sup> Day-to-day clinical operations can be improved by ML, evident by medication error alert software<sup>19</sup> and an ongoing collaboration between Stanford and Google that may eventually allow for a clinic room computer to act as a scribe during the visit.<sup>20</sup>

These examples barely scratch the surface of possibilities, and patients, providers, insurers, and other stakeholders are all likely to benefit from ML tools; however, despite their potential value, most of these techniques have not yet been adopted in clinical practice. The FDA approved the first AI diagnostic device meant for use without further physician interpretation on April 11, 2018.<sup>21</sup> The software analyzes retinal photographs to determine the presence of diabetic retinopathy and helps extend the reach of a previously underutilized screening recommendation critical to the eye health of diabetic patients.

## ARTIFICIAL INTELLIGENCE IN SLEEP MEDICINE

### Work to date

#### Sleep staging

Sleep staging is a labor intensive procedure due to the required manual inspection, in 30-second epochs, of EEG, electrooculogram, and electromyogram derivations by sleep technologists.<sup>22</sup> Human performance of sleep staging is evaluated with interrater reliability among sleep technologists as quantified by kappa ( $\kappa$ ) values that reflect epoch-by-epoch agreement above chance. The cited  $\kappa$  value range of approximately 0.68–0.76<sup>23</sup> is the benchmark against which human-computer algorithm agreement is measured. Currently, the most researched application of AI in sleep medicine is the classification of sleep stages in PSG data. Additionally, automated PSG scoring has been evaluated as a measure to augment sleep technologist performance. For example, the use of certain computer-derived PSG characteristics (automated sleep spindle detection, sleep depth, and delta duration) improved sleep technologist interrater reliability when coupled with manual review,<sup>24</sup> and novel measures of sleep depth have demonstrated the ability to predict arousal in subsequent epochs.<sup>25</sup>

Since the late 1990s, multiple groups have investigated the accuracy of automated sleep staging and have demonstrated human-computer algorithm agreement comparable to human interrater reliability.<sup>26–34</sup> However, due to small sample sizes that limited the ability to evaluate algorithm performance on a testing set of unique patients, these studies may not be generalizable. Additionally, prior limitations in computational power, and lack of available software, have prevented adoption of automated sleep staging.<sup>35</sup>

More recent studies have overcome these limitations and highlight the potential for generalizability with preserved, robust performance of ML algorithms across large, heterogeneous patient populations. ML algorithms that employ various classifiers and different methods of feature determination have been trained on datasets containing thousands of study participants.<sup>35–37</sup> During testing on novel datasets, ML algorithms have demonstrated

sleep staging accuracy similar to interrater reliability among scorers with reported  $\kappa$  values up to 0.80.<sup>36</sup>

Additional signal outputs not typically used to stage sleep, such as heart rate variability and respiratory dynamics, may also contribute to the ability of ML algorithms to determine arousals and sleep stage transitions.<sup>38</sup> Future ML efforts may redefine how PSG tests are scored as new associations between complex signal patterns and various outcomes are discovered.

### Scoring of respiratory events and movements

Like sleep staging, manual scoring of respiratory events by visual inspection of airflow and respiratory effort channels is a resource-intensive practice vulnerable to interscorer variability. Automated scoring of apneas and hypopneas has demonstrated accuracy comparable to manual scorers.<sup>32,33,39</sup> Recently, deep neural network evaluation of 10,000 PSG tests<sup>36</sup> automatically calculated apnea-hypopnea indices strongly correlating with expert scoring ( $r^2 = 0.85$ ). Automated limb movement index scoring in the same study also correlated strongly with expert scoring ( $r^2 = 0.79$ ). ML has been applied to more limited channel assessments including those based on pulse oximetry in children<sup>40,41</sup> with promising accuracy for out-of-center tools.

### Wearable sensor analysis

Consumer sleep technology (CST) collects multiple physiological signals, such as heart rate (from optical photoplethysmography) and acceleration. Although the role of CST in the practice of sleep medicine remains unclear,<sup>42</sup> ML is already used by some commercial manufacturers to estimate sleep from the massive amount of information collected by wearable devices.<sup>43</sup> A recent investigation demonstrated that ML predictions from heart rate variability features derived from wrist photoplethysmography, and motion from triaxial accelerometry, allowed for sleep-wake differentiation and sleep stage prediction with  $\kappa$  values of 0.55 and 0.42 respectively.<sup>44</sup> Deep neural network analyses have been applied to motion and heart rate signal from the Apple Watch to estimate sleep stage<sup>45</sup> and evaluate for sleep-disordered breathing.<sup>46</sup>

Although a systematic review of published works that use ML for sleep-related applications is beyond the purview of this paper, these examples provide a small snapshot of the growing body of peer-reviewed literature. At the time of this publication, a PubMed search for “machine learning sleep” revealed nearly 300 citations, the majority published in the last 2–3 years.

### Use cases

In addition to the automated staging of sleep and scoring of respiratory events, ML techniques have the potential to reveal patterns within PSG signals not captured by traditional summary measures. This information, particularly when combined with other clinical and demographic characteristics, may provide more in-depth subtyping and precise diagnostic information to improve the clinical care of sleep disorders. Additionally, like the broader applications in medicine such as AI scribe software, ML is expected to provide tools to improve the efficiency of our day-to-day operations. The following cases are examples of potential applications in the field of sleep medicine.

## Improved diagnosis and subtyping of disorders

**Excessive daytime sleepiness:** Excessive daytime sleepiness is present in approximately 20% of the general population<sup>47</sup> and is often the presenting symptom in a sleep disorders center. OSA is a common, treatable cause of sleepiness, and it accounts for almost 70% of individuals assessed in sleep centers.<sup>48</sup> However, metrics traditionally used to quantify severity, such as the apnea-hypopnea index (AHI) and oxygen desaturation nadir, do not correlate strongly with sleepiness.<sup>49–53</sup> A key to understanding sleepiness in OSA may lie in PSG features that are not routinely summarized in sleep study reports. Previous work demonstrated that novel PSG parameters, such as respiratory event-linked EEG changes, better predict sleepiness than AHI<sup>54</sup> and have been shown to decrease with PAP therapy.<sup>55</sup>

In individuals with sleepiness not explained by sleep-disordered breathing or other readily apparent cause, a Multiple Sleep Latency Test (MSLT) is often pursued as part of the evaluation for a central disorder of hypersomnolence. Unfortunately, the mean sleep latency derived from MSLT, which is used to categorize individuals as hypersomnolent or normal, is sensitive to medications and substances, circadian phase,<sup>56</sup> and sleep loss. Additionally, the MSLT has poor test-retest reliability.<sup>57</sup> PSG evaluation with linear dimension analysis<sup>58</sup> and deep learning<sup>59</sup> has demonstrated the capability of ML models to objectively confirm narcolepsy with accuracy similar to traditional PSG followed by MSLT. ML tools might also be applied to signals acquired during MSLT to assist with diagnosis and reveal previously unseen subgroups within the central disorders of hypersomnolence.

**Sleep-disordered breathing:** It is increasingly recognized that OSA is caused by multiple pathophysiologic mechanisms, and that the AHI fails to capture the breadth of physiological heterogeneity. Traits that contribute to sleep-disordered breathing include upper airway neuromuscular collapsibility, arousal threshold, loop gain, and circulatory time; combinations of these factors define unique groups of patients with OSA<sup>60</sup> and importantly, they associate with comorbidity and response to treatment with PAP.<sup>60,61</sup> ML may be a useful technique to further identify unique subtypes.<sup>62</sup>

Additionally, our current measurements may oversimplify the distinction between obstructive and central sleep apnea. Cardiopulmonary coupling evaluation identified a distinct subtype, marked by non-rapid eye movement predominant sleep apnea characterized by an increased propensity for central sleep-disordered breathing and periodicity.<sup>63</sup> Because these individuals are more likely to respond suboptimally to PAP due to treatment-emergent central apneas,<sup>63</sup> automated measures like cardiopulmonary coupling could provide anticipatory management in devising treatment strategies.

The appropriate use of non-PAP treatment modalities may also be informed by harnessing the rich information within the PSG. Initial work has developed predictive models for determining response to supplemental oxygen (which blunts chemoreflex-related ventilatory control),<sup>64</sup> hypnotics (which raise the arousal threshold),<sup>65–67</sup> and oral appliances (which increase airway caliber).<sup>68</sup> While these approaches have utilized

hypothesis-driven physiological approaches, much is unknown regarding the physiological underpinnings of sleep-disordered breathing subtypes. Therefore, application of data-driven ML approaches may empirically identify improved metrics of treatment response.

The potential for novel ML algorithms to identify individuals with distinct subtypes not captured by the AHI is clear. However, the overall utility of this approach in clinical practice requires additional study. Because the physiological signals derived from routine PSG are highly underutilized by our current summary metrics, ML-based tools provide the opportunity for predictive subtyping to improve the precision of sleep-disordered breathing classification and treatment.

**REM sleep behavior disorder:** Appropriate recognition of REM sleep behavior disorder (RBD) is crucial given its association with incident alpha-synucleinopathy neurodegenerative disorders in at least 80% of individuals.<sup>69,70</sup> RBD is confirmed by the PSG finding of REM sleep without atonia (RSWA), which requires detailed visual inspection of chin and limb electromyogram tone. The clinically significant levels of the two types of RSWA, excessive transient and sustained, has not been specified. The ability of antidepressant medication to increase muscle tone during stage R sleep further complicates the assessment of RSWA.<sup>71</sup>

Given the subjectivity in the accurate visual identification of RSWA, computer-assisted algorithms have been developed but are not used widely in clinical care.<sup>72–76</sup> ML (supervised and reinforcement learning) models may improve the assessment of RSWA and may even specify when RSWA is related to antidepressant use.

A particularly important question is whether PSG features of patients with RBD, combined with clinical information, can reveal which individuals will subsequently develop alpha-synucleinopathy. When neuroprotective agents become available, sleep clinicians must have the capacity to identify electrophysiological signatures and clinical characteristics predictive of individuals who may benefit from disease modifying agents; therefore, use of ML may further a clinical imperative to identify preclinical presentation of neurodegenerative disease.

**Patient-generated data:** ML applications will transcend the sleep laboratory given the widespread, longitudinal collection of sleep-related metrics in the ambulatory environment, including CST such as wearables and “nearables”<sup>42,77</sup> and PAP download data.

Wrist actigraphy is an option for clinical use in the diagnosis and evaluation of patients with sleep disorders<sup>78</sup>; however, existing scoring methods do not capitalize on the longitudinal, ambulatory measurements that are obtained with continuous actigraphy monitoring. Currently, actigraphy is used to quantify sleep parameters (typically total sleep time) or visually appraise sleep and wake patterns (for example, in suspected circadian rhythm sleep-wake disorders); however, data-driven algorithms applied to motion measured by actigraphy may provide additional information about sleep-wake behaviors beyond that of traditional techniques, potentially playing a role in augmenting physician diagnosis of sleep disorders.

In addition, clinical grade actigraphy has historically only monitored movement. Newer wearable devices, primarily those marketed toward consumers, now capture additional metrics, such as heart rate.<sup>42</sup> Improved sleep prediction by wearable device data analyzed with ML would allow for inexpensive, objective assessments of sleep duration and quality over days to weeks, potentially with improved classification compared with existing algorithms. However, significant development and validation is required, which ideally would include publication of underlying algorithms.

Obvious uses of validated wearable devices would include the evaluation and management of central disorders of hypersomnolence and circadian rhythm sleep-wake disorders.<sup>79</sup> Beyond that, sleep duration could be measured longitudinally in individuals with OSA treated with PAP and compared to adherence data to determine clinical effectiveness of PAP therapy,<sup>77</sup> which may reveal the source of persistent symptoms in OSA patients despite “adequate” PAP adherence.

### **Delivery of care in sleep medicine**

As noted in other specialties, particularly radiology and pathology,<sup>1</sup> ML tools likely will transform how we practice sleep medicine and, with appropriate engagement of the sleep community, will augment current care by achieving improved efficiencies and reproducibility, in some cases leading to new disease classifiers and predictors. The most obvious application of ML in sleep medicine to improve clinical efficiency is automated PSG sleep staging, but many other manual clinical activities may benefit from some degree of automation, such as review and documentation of PAP-generated adherence and efficacy data.

ML analysis of the continually collected, massive amount of PAP-generated data could alert the provider to deterioration in patient adherence or control of sleep-disordered breathing and trigger an intervention. PAP-generated data may be particularly valuable when combined with real-time patient report from mobile health applications. ML analysis of this data could then in turn provide automated feedback to the patient for self-management.

High-throughput automated systems will determine risk and optimize interventions by combining multiple sources of sleep data, such as PSG, PAP devices, CST, the electronic health record, and patient-generated, real-time self-reports.

### **Population health**

In addition to moving sleep medicine toward precision diagnosis and personalized treatment in the clinic, improved subtyping of sleep disorders may also improve our understanding of the role of sleep in health and disease.

Inconsistencies regarding the influence of OSA on other health outcomes is a problem amenable to the use of ML. For example, after the identification of distinct subtypes in OSA, Zinck and colleagues demonstrated that, unlike traditional AHI categorizations, these previously undefined groups were predictive of cardiovascular events.<sup>62</sup> It can be anticipated that routine extraction of these and other quantitative subtypes from PSG signals may better define the contribution of sleep-disordered breathing to multiple disease states.



Sleepiness is highly variable across the population, and traditional PSG measurements only explain a small portion of the variance in objective or self-reported measures of sleepiness.<sup>49,80</sup> ML analysis is likely to capture critical aspects of sleep that reflect the homeostatic processes that underlie sleepiness and recovery from insufficient sleep, such as sleep microarchitecture, the complex dynamics of sleep across the sleep period, or the interplay between sleep with other physiological processes.

For more than 20 years, numerous investigations have demonstrated a U-shaped relationship between sleep duration and multiple diseases as well as mortality.<sup>81–83</sup> However, the causative mechanisms remain unclear. The knowledge gap may result from the quantification of only one aspect of sleep.<sup>84</sup> Sleep is a highly faceted state with distinct and quantifiable features including not only duration, but also continuity, timing, alertness, quality, regularity, and rhythmicity.<sup>84,85</sup> ML analyses of dozens of potential sleep predictors identified decreased sleep-wake rhythmicity and reduced sleep continuity as risk factors for increased mortality.<sup>84</sup> Additionally, wearable data, which provides the opportunity to collect a wealth of information in the free-living environment over an extensive duration, across heterogeneous populations, could reveal characteristics predictive of vulnerability to homeostatic and circadian challenges. In sum, ML analyses of large datasets are likely to reveal new relationships between sleep and health outcomes and identify more specific targets for intervention on a population level.

### Advantages of machine learning in sleep medicine

The previous discussion highlights only a few of the opportunities ML could provide the field of sleep medicine.

PSG captures quantitative data from multiple physiological systems and across time, which provides the opportunity to assess temporal features across multiple scales of measurement. Advanced analytic techniques may provide additional value of the PSG beyond confirming a diagnosis of sleep-disordered breathing, titrating PAP, and evaluating movements and behaviors. The generation of new quantitative metrics from the PSG, particularly when integrated with other sources of data—such as clinical history, biochemical markers, behavioral information, and genomic data—could create a robust knowledge base to define patient subgroups whose sleep disorders differ mechanistically and respond differently to treatments, thus supporting personalization of sleep medicine. Prediction capabilities could inform prevention strategies at a public health level. Additionally, ML can identify characteristics that predict progression from prodromal symptoms to full disease expression and therefore guide study participant selection to improve power of clinical trials.<sup>86</sup>

The use of AI to stage sleep and score respiratory and movement events could reduce the time sleep technologists are required to devote to PSG scoring, enabling them to provide greater assistance with patient needs including titration of PAP and troubleshooting patient complaints about their PAP interface. Additionally, ML techniques may allow the sleep medicine provider to efficiently distill vast amounts of data from multiple sources, including laboratory and imaging results, assessments by physicians in other specialties, PAP adherence

and efficacy metrics, and patient-generated data from wearable and mobile devices.

### Disadvantages of machine learning in sleep medicine

Analysis of data derived from PSG in conjunction with clinical, demographic, biochemical and genomic variables, should be guided by a thoughtful conceptual framework that considers the plausibility of linking specific types of data, the reliability of those data elements and metrics, and the feasibility of their measurement in large numbers of individuals.

The enthusiasm of the sleep community to utilize these approaches and tools is exemplified by the increasing proportion of data requests made of the National Sleep Research Resource (<https://www.sleepdata.org>), a sleep data repository, funded by the National Institutes of Health, containing over 30,000 sleep studies, including raw physiological records, annotated and summary files, as well as variable clinical outcome and covariate data. More than 40 terabytes of data have been shared with over 400 users across the globe. Nearly half of the requests involve use of ML for purposes such as improved sleep staging and better outcome prediction.

While exciting, analysis by AI of large datasets and diverse data types is not without its challenges. The depth and complexity of the PSG provides an almost limitless number of quantitative metrics that can be extracted. However, many of these will have little physiological meaning or utility. Conversely, many potentially informative features may be unknown. Some ML approaches may be sensitive to variations in the underlying data, due to differences in how data were collected or saved and the presence of artifact. Differences in data quality and content across datasets may not be obvious, underscoring the need for clinicians and investigators to understand the nature of the data they analyze, and for all such research and clinical applications to include manual review of data being input and output.

Other limitations of AI algorithms and the products of such analyses relate to lack of transparency, especially when algorithms are generated for proprietary purposes. Algorithms that are “black boxes,” as is common in commercially available, wearable technology and in many proprietary sleep scoring programs, preclude independent verification and validation and limit the community’s understanding of the specific elements that may be relevant for disease or diagnosis classification, preventing independent extension of previous research. In addition, proprietary algorithms often are validated on datasets that are not publicly available. This lack of transparency directly opposes national imperatives to improve the rigor and reproducibility of scientific products, and it reduces enthusiasm for clinical use.

Analysis of large numbers of candidate predictors requires large data samples, with independent samples for training and testing. Validating initial findings requires care due to heterogeneity in the samples, as some associations may vary by age, sex and underlying disease distributions, as well as a variety of contextual influences. Prediction also can be limited due to the lack of gold standard or “ground truth” for some physiological or clinical outcomes. For example, use of novel metrics to predict the AHI or sleep stages is somewhat circular given

that these latter variables, generated by humans, are imperfect. Further limitations arise when needing to choose among “bronze” standards that have changed over time (eg, AHI defined using variable definitions).

There is limited availability of well-annotated and large datasets with the diversity and outcomes needed to fully support the many opportunities that big data and AI hold for the sleep community. Clinical sleep laboratories produce large amounts of data daily. However, there currently is no infrastructure or support to collect and archive clinical data in ways to advance big data analytics. A major barrier is the lack of a single tool to export sleep signal data in its raw form. While European Data Format files are commonly used, they are not completely standardized; other formats exist, and no such tool is routinely used in clinical laboratories for data archival. Another challenge is lack of a standardized vocabulary for annotating key sleep events. To harness the technology that has the potential to truly advance sleep health, the sleep medicine community would vastly benefit from a strategic implementation of a national data-sharing initiative, possibly linking existing efforts and leveraging the National Sleep Research Resource, that includes a framework for standardizing sleep meta-data and for archiving, displaying and providing access to large amounts of physiological signals and annotation files.

Integration of ML tools into the routine clinical practice of sleep medicine will also present significant challenges. The following approach describes logistical considerations as ML-generated data are unlikely to stand alone and will require clinician review and knowledge (ie, augmented intelligence). Therefore, if appropriate measures are not put into place, integration of AI may increase clinician workload, technology resource utilization at the health care system level, security risk, and liability.

## INITIAL IMPLEMENTATION CONSIDERATIONS

AI applications to sleep medicine pose significant opportunities and challenges. We should adopt processes to improve our understanding of the mechanisms underlying disturbed sleep and resultant outcomes, to more accurately subtype sleep disorders for greater precision in care, and to enhance efficiency of day-to-day operations in the sleep laboratory and clinic. The appropriate role of AI in sleep medicine is to augment, not replace, the clinical decision-making and care delivery of the sleep medicine team.

### Regulation

The speed of ML technology development poses significant challenges to regulatory institutions. In response to the complexity and rapid evolution of digital health tools, the FDA has issued the Digital Health Innovation Action Plan to foster technological advances and ensure patient safety. The most relevant initiative for ML tools is the Digital Health Software Precertification (Pre-Cert) program, which applies to technologies that fall under the designation of software as a medical device. A working model of the Pre-Cert program was released in April 2018, proposing regulation that targeted software manufacturers instead of the products.<sup>87</sup> Manufacturers are

expected to display principles of organizational excellence that include a commitment to product quality, patient safety, clinical responsibility, cybersecurity, and a proactive culture. The Pre-Cert program mechanism will allow products from precertified manufacturers to enter the market without delay, and the FDA currently seeks public comment on how to capture the 30 software-related review elements to approve products from companies with the Pre-Cert designation while decreasing burden. Additionally, postmarket monitoring will continually assess real-world performance analytics of health outcomes, user experience, and product performance. Measures of product performance will assess accuracy, reliability, and security of a software as a medical device product.

A specific challenge posed in the certification of ML tools is that different ML models have different minimum requirements for training datasets; therefore, recommended training set properties for one model may not be applicable to another type of model. Efforts should instead focus on setting minimum performance benchmarks for models and specific product applications, akin to standardized testing benchmarks that are set for technicians and clinicians. By setting forth requirements for transparency, validation, quality and utility, leading medical associations can shape the marketplace such that ML developers work hand-in-hand with physicians and scientists to advance the understanding of health and disease while improving patient care.

A key performance benchmark used by AI algorithms in diagnostic applications is the ability to perform a task (eg, apnea detection and classification) at a level comparable to a human expert (eg, a well-trained, experienced sleep technologist). At this juncture, algorithm-expert agreement equivalent to cited interscorer reliability is suggested as the required minimal level of performance for clinical use. However, the achievement of this level of performance in a testing set does not ensure generalizability to the user’s sleep laboratory. Before incorporating AI independent of sleep technologists, sleep laboratories should work with the product manufacturer to trial the software of interest in their own laboratory to determine real-life performance. Of note, with internal testing, the laboratory may find that some, but not all PSG metrics, are reliably classified by the algorithm (eg, algorithm accurately stages sleep but does not detect respiratory events). Unchanged from current practice, full physician review is expected after a study is scored by a sleep technologist with assistance from AI-based software and, if indicated, manual scoring of the entire record may remain necessary for certain studies. Sleep laboratory disclosure of software performance in their own laboratory will serve to advance transparency in the use of AI clinical applications. Dissemination and synthesis of aggregate results of sleep laboratory validation testing will benefit patients, clinicians and scientists.

At this time, only software analyzing PSG data should be considered for implementation. AI products designed with or for data obtained from acquisition methods that are not considered gold-standard (eg, consumer wearables that use photoplethysmography and accelerometry to estimate sleep, headband-embedded dry electrode EEG sensors, noncontact sleep and respiratory assessment devices) have not been verified

for clinical use. However, as these sensors achieve clinical validity, algorithms applied to these signals may reveal that these devices can provide benefit to the clinical evaluation and management of sleep disorders.

### Logistical considerations

A training mechanism must be in place to assist care providers and health systems with integration of AI software into care pathways. Clinicians are already faced with multiple information streams that they are expected to monitor in their daily work; therefore, to reduce further burden, any newly introduced software is expected to smoothly interface with currently used electronic health records and PSG software.

Additionally, implementation of AI software should be guided by lessons learned from adoption of the electronic health record. For example, one of the unintended consequences of the computerization of medicine is “alert fatigue,” the desensitization of providers to electronically generated, automated warnings due to the sheer number of these notifications received each day.<sup>88</sup> The goal of AI software product interfaces should be improvement of patient care without increasing data burden or contributing to alert fatigue. Therefore, input of sleep medicine physicians, advanced practice providers, sleep technologists, behavioral sleep medicine clinicians, and sleep and circadian scientists is vital to the development of new digital AI tools.

The initial application of AI in sleep medicine will likely be the automated staging of sleep and scoring of respiratory and movement parameters, which easily integrates into the current sleep center workflow. However, innovation in reimbursement mechanisms for sleep professionals’ time and effort will be required for future development of clinical decision support tools, which may include novel measures of sleep quality from the PSG signal as well as health indicators based on demographic, clinical, and/or PAP download data. Furthermore, new clinical tools may draw attention to previously unidentified abnormalities of uncertain significance or discovery of new pathologies that justify medical assessment and increase patient contact. Given the current shortage of sleep specialists,<sup>89</sup> clinic infrastructure changes to meet these needs may be required.

Data repositories associated with AI products should have stringent security measures in place to protect patient privacy. Data storage methods must be compliant with the health insurance portability and accountability act. Well-defined methods will be needed to govern access to data within these repositories.

Currently available sources of objective sleep data, which include the PSG, home sleep apnea test, actigraphy, and PAP download, can inform the evaluation and treatment of sleep disorders. However, the relevance of this ancillary information is highly dependent on the provider’s interaction with data. Manual review and careful clinical correlation are required for appropriate interpretation of objective findings. AI tools are not exempt, and output should not be taken at face value but should be considered as a component of the comprehensive clinical assessment required for the diagnosis and management of sleep disorders.

### Ethics

Use of ML in both research and clinical practice poses significant ethical dilemmas. Health care disparities that already

exist, such as sex discrepancies in the evaluation of sleep-disordered breathing, could be magnified with the implementation of ML.

The data from which algorithms originate, the particular goals of the developers, and the settings in which these tools are deployed are all sources of bias.<sup>90</sup> Because model development and training are most likely to occur when massive amounts of easily accessible data are available, bias may lie in the characteristics of individuals who are more likely to act as a data source.<sup>91</sup> For example, data may be more likely to originate from patients who present to a specialty clinic, use academic as opposed to community health centers, consent to participate in research, and have means that allow for self-generated data. Therefore, models trained on this data may not generalize to individuals that do not fit this profile. AI developers and clinicians must be vigilant to ensure that ML tools do not promote bias by excluding individuals with decreased resources and reduced access to health care. Rigorous testing and certification standards will need to be developed to clearly communicate to clinicians the populations that are appropriate for any given AI model, and inequities should be noted and mitigated.

AI will allow for the collection of massive amounts of patient-generated data and storage in repositories accessible to the health system. However, collection, transmission, and storage of data does not guarantee that a trained health care professional has appropriately reviewed and acted on these data. For example, signs of patient deterioration (eg, increasing PAP-derived AHI or nonadherence to therapy) may be recorded but not viewed or addressed. Additionally, if tools are developed that provide automated recommendations to patients based on these data without a medical provider’s evaluation, problems that previously would have been brought to the provider’s attention may be neglected. The gap between data acquisition and appropriate data utilization may lie in provider education as well as development and implementation of new quality improvement tools relevant to new workflows. Just as appropriate licensure is required to perform and interpret certain diagnostic tests, AI tools should not be implemented for clinical purposes until all potential users demonstrate competence interacting with the software, displaying an understanding of how to integrate these workflows with appropriate quality control.

Ultimately, the clinician is responsible for patient outcomes; therefore, these potential pitfalls highlight that AI tools should augment, but not substitute, the clinical judgement of the sleep medicine provider.

## CONCLUSIONS

The most immediate application of AI in sleep medicine, analysis of multiple physiological signals acquired during PSG, is expected to deepen our understanding of the architecture of normal and disturbed sleep, improve disease subtyping, and increase efficiency of sleep laboratory operations to improve patient care. However, no singular objective assessment can replace comprehensive clinical appraisal. AI analyses must be used in conjunction with careful assessment of patient signs and symptoms, demographics and comorbidities, and reassessment

over the course of the chronic conditions we treat. Like any diagnostic tool, AI will be dependent on the clinician's aptitude and the context in which it is incorporated in order to achieve clinical utility.

AI is likely to transform medicine; however, development of appropriate tools that can be effectively translated from research applications to patient care will require significant support. The American Academy of Sleep Medicine Foundation has already identified AI as one of the specific research domains targeted by its 2020 Strategic Research Award program. With continued collaboration across the sleep disorders care team, researchers, and product developers, AI will deepen our understanding of disturbed sleep and its contributions to health, facilitating the care of all patients with sleep medicine needs.

## ABBREVIATIONS

AHI, apnea-hypopnea index  
 AI, artificial intelligence  
 CST, consumer sleep technology  
 EEG, electroencephalogram  
 FDA, US Food and Drug Administration  
 ML, machine learning  
 MSLT, Multiple Sleep Latency Test  
 OSA, obstructive sleep apnea  
 PAP, positive airway pressure  
 PSG, polysomnography  
 RBD, REM sleep behavior disorder  
 REM, rapid eye movement  
 RSWA, REM sleep without atonia

## REFERENCES

- Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. *JAMA*. 2016;316(22):2353–2354.
- Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med*. 2018;15(11):e1002686.
- De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24(9):1342–1350.
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–2410.
- Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. 2017;318(22):2199–2210.
- Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–118.
- Acharya UR, Hagiwara Y, Adeli H. Automated seizure prediction. *Epilepsy Behav*. 2018;88:251–261.
- Rajpurkar P, Hannun AY, Haghighpanahi M, Bourm C, Ng AY. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. <https://arxiv.org/abs/1707.01836>. Published July 6, 2017. Accessed February 19, 2020.
- Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: A development and prospective validation study. *PLoS Med*. 2018;15(11):e1002705.
- Fontanella S, Frainay C, Murray CS, Simpson A, Custovic A. Machine learning to identify pairwise interactions between specific IgE antibodies and their association with asthma: a cross-sectional analysis within a population-based birth cohort. *PLoS Med*. 2018;15(11):e1002691.
- Saria S, Goldenberg A. Subtyping: what it is and its role in precision medicine. *IEEE Intell Syst*. 2015;30(4):70–75.
- Simpson A, Tan VY, Winn J, et al. Beyond atopy: multiple patterns of sensitization in relation to asthma in a birth cohort study. *Am J Respir Crit Care Med*. 2010;181(11):1200–1206.
- Corey KM, Kashyap S, Lorenzi E, et al. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data (Pythia): a retrospective, single-site study. *PLoS Med*. 2018;15(11):e1002701.
- Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med*. 2020 Jan 21. [Epub ahead of print].
- Muoio D. IT leaders highlight big success with machine learning models. <https://www.healthcareitnews.com/news/it-leaders-highlight-big-success-machine-learning-models>. Published March 5, 2018. Accessed February 14, 2020.
- Beck AH, Sangoi AR, Leung S, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med*. 2011;3(108):108ra113.
- Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med*. 2002;8(1):68–74.
- Peters AL, Ahmann AJ, Hirsch IB, Raymond JK. Advances in glucose monitoring and automated insulin delivery: supplement to Endocrine Society clinical practice guidelines. *J Endocr Soc*. 2018;2(11):1214–1225.
- Schiff GD, Volk LA, Volodarskaya M, et al. Screening for medication errors using an outlier detection system. *J Am Med Inform Assoc*. 2017;24(2):281–287.
- Chiu C-C, Tripathi A, Chou K, et al. Speech recognition for medical conversations. [https://www.isca-speech.org/archive/Interspeech\\_2018/abstracts/0040.html](https://www.isca-speech.org/archive/Interspeech_2018/abstracts/0040.html). Published 2018. Accessed February 19, 2020.
- US Food & Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye>. April 11, 2018. Accessed February 14, 2020.
- Berry RB, Brooks R, Gamaldo CE, et al; for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Version 2.4. Darien, IL: American Academy of Sleep Medicine; 2017.
- Danker-Hopfe H, Anderer P, Zeithofer J, et al. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. *J Sleep Res*. 2009;18(1):74–84.
- Younes M, Hanly PJ. Minimizing interrater variability in staging sleep by use of computer-derived features. *J Clin Sleep Med*. 2016;12(10):1347–1356.
- Younes M, Ostrowski M, Soiferman M, et al. Odds ratio product of sleep EEG as a continuous measure of sleep state. *Sleep*. 2015;38(4):641–654.
- Anderer P, Gruber G, Parapatics S, et al. An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24 x 7 utilizing the Siesta database. *Neuropsychobiology*. 2005;51(3):115–133.
- Anderer P, Moreau A, Woertz M, et al. Computer-assisted sleep classification according to the standard of the American Academy of Sleep Medicine: validation study of the AASM version of the Somnolyzer 24 x 7. *Neuropsychobiology*. 2010;62(4):250–264.
- Fraivan L, Lweesy K, Khasawneh N, Fraivan M, Wenz H, Dickhaus H. Classification of sleep stages using multi-wavelet time frequency entropy and LDA. *Methods Inf Med*. 2010;49(3):230–237.
- Hassan AR, Bhuiyan MI. A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. *J Neurosci Methods*. 2016;271:107–118.



30. Lajnef T, Chaibi S, Ruby P, et al. Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. *J Neurosci Methods*. 2015;250:94–105.
31. Liang SF, Kuo CE, Hu YH, Cheng YS. A rule-based automatic sleep staging method. *J Neurosci Methods*. 2012;205(1):169–176.
32. Malhotra A, Younes M, Kuna ST, et al. Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep*. 2013;36(4):573–582.
33. Punjabi NM, Shifa N, Dorffner G, Patil S, Pien G, Aurora RN. Computer-assisted automated scoring of polysomnograms using the Somnolyzer system. *Sleep*. 2015;38(10):1555–1566.
34. Schaltenbrand N, Lengelle R, Toussaint M, et al. Sleep stage scoring using the neural network model: comparison between visual and automatic analysis in normal subjects and patients. *Sleep*. 1996;19(1):26–35.
35. Patanaik A, Ong JL, Gooley JJ, Ancoli-Israel S, Chee MWL. An end-to-end framework for real-time automatic sleep stage classification. *Sleep*. 2018;41(5).
36. Biswal S, Sun H, Goparaju B, Westover MB, Sun J, Bianchi MT. Expert-level sleep scoring with deep neural networks. *J Am Med Inform Assoc*. 2018;25(12):1643–1650.
37. Sun H, Jia J, Goparaju B, et al. Large-scale automated sleep staging. *Sleep*. 2017;40(10).
38. Citi L, Bianchi MT, Klerman EB, Barbieri R. Instantaneous monitoring of sleep fragmentation by point process heart rate variability and respiratory dynamics. *Conf Proc IEEE Eng Med Biol Soc*. 2011;2011:7735–7738.
39. Pittman SD, MacDonald MM, Fogel RB, et al. Assessment of automated scoring of polysomnographic recordings in a population with suspected sleep-disordered breathing. *Sleep*. 2004;27(7):1394–1403.
40. Crespo A, Alvarez D, Kheirandish-Gozal L, et al. Assessment of oximetry-based statistical classifiers as simplified screening tools in the management of childhood obstructive sleep apnea. *Sleep Breath*. 2018;22(4):1063–1073.
41. Hornero R, Kheirandish-Gozal L, Gutierrez-Tobal GC, et al. Nocturnal oximetry-based evaluation of habitually snoring children. *Am J Respir Crit Care Med*. 2017;196(12):1591–1598.
42. Khosla S, Deak MC, Gault D, et al. Consumer sleep technology: an American Academy of Sleep Medicine position statement. *J Clin Sleep Med*. 2018;14(5):877–880.
43. de Zambotti M, Rosas L, Colrain IM, Baker FC. The sleep of the ring: comparison of the OURA sleep tracker against polysomnography. *Behav Sleep Med*. 2017;17(2):124–136.
44. 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).
45. Walch O, Huang Y, Forger D, Goldstein C. Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. *Sleep*. 2019;42(12).
46. Tison GH, Singh AC, Ohashi DA, et al. Cardiovascular risk stratification using off-the-shelf wearables and a multi-task deep learning algorithm. *Circulation*. 2017;136(Suppl 1):A21042.
47. Young TB. Epidemiology of daytime sleepiness: definitions, symptomatology, and prevalence. *J Clin Psychiatry*. 2004;65(Suppl 16):12–16.
48. Punjabi NM, Welch D, Strohl K. Sleep disorders in regional sleep centers: a national cooperative study. Coleman II Study Investigators. *Sleep*. 2000;23(4):471–480.
49. Gottlieb DJ, Whitney CW, Bonekat WH, et al. Relation of sleepiness to respiratory disturbance index: the Sleep Heart Health Study. *Am J Respir Crit Care Med*. 1999;159(2):502–507.
50. Guilleminault C, Partinen M, Quera-Salva MA, Hayes B, Dement WC, Nino-Murcia G. Determinants of daytime sleepiness in obstructive sleep apnea. *Chest*. 1988;94(1):32–37.
51. Olson LG, Cole MF, Ambrogetti A. Correlations among Epworth Sleepiness Scale scores, multiple sleep latency tests and psychological symptoms. *J Sleep Res*. 1998;7(4):248–253.
52. Roue N, Gomez S, Mediano O, et al. Daytime sleepiness and polysomnography in obstructive sleep apnea patients. *Sleep Med*. 2008;9(7):727–731.
53. Sharkey KM, Orff HJ, Tosi C, Harrington D, Roye GD, Millman RP. Subjective sleepiness and daytime functioning in bariatric patients with obstructive sleep apnea. *Sleep Breath*. 2013;17(1):267–274.
54. Chervin RD, Burns JW, Ruzicka DL. Electroencephalographic changes during respiratory cycles predict sleepiness in sleep apnea. *Am J Respir Crit Care Med*. 2005;171(6):652–658.
55. Chervin RD, Shelgikar AV, Burns JW. Respiratory cycle-related EEG changes: response to CPAP. *Sleep*. 2012;35(2):203–209.
56. Carskadon MA, Wolfson AR, Acebo C, Tzischinsky O, Seifer R. Adolescent sleep patterns, circadian timing, and sleepiness at a transition to early school days. *Sleep*. 1998;21(8):871–881.
57. Trotti LM, Staab BA, Rye DB. Test-retest reliability of the multiple sleep latency test in narcolepsy without cataplexy and idiopathic hypersomnia. *J Clin Sleep Med*. 2013;9(8):789–795.
58. Olsen AV, Stephansen J, Leary E, et al. Diagnostic value of sleep stage dissociation as visualized on a 2-dimensional sleep state space in human narcolepsy. *J Neurosci Methods*. 2017;282:9–19.
59. Stephansen JB, Olesen AN, Olsen M, et al. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. *Nat Commun*. 2018;9(1):5229.
60. Eckert DJ, White DP, Jordan AS, Malhotra A, Wellman A. Defining phenotypic causes of obstructive sleep apnea. Identification of novel therapeutic targets. *Am J Respir Crit Care Med*. 2013;188(8):996–1004.
61. Wellman A, Edwards BA, Sands SA, et al. A simplified method for determining phenotypic traits in patients with obstructive sleep apnea. *J Appl Physiol*. 2013;114(7):911–922.
62. Zinчук AV, Jeon S, Koo BB, et al. Polysomnographic phenotypes and their cardiovascular implications in obstructive sleep apnoea. *Thorax*. 2018;73(5):472–480.
63. Thomas RJ, Mietus JE, Peng CK, et al. Differentiating obstructive from central and complex sleep apnea using an automated electrocardiogram-based method. *Sleep*. 2007;30(12):1756–1769.
64. Sands SA, Edwards BA, Terrill PI, et al. Identifying obstructive sleep apnoea patients responsive to supplemental oxygen therapy. *Eur Respir J*. 2018;52(3):1800674.
65. Carter SG, Berger MS, Carberry JC, et al. Zopiclone increases the arousal threshold without impairing genioglossus activity in obstructive sleep apnea. *Sleep*. 2016;39(4):757–766.
66. Eckert DJ, Malhotra A, Wellman A, White DP. Trazodone increases the respiratory arousal threshold in patients with obstructive sleep apnea and a low arousal threshold. *Sleep*. 2014;37(4):811–819.
67. Eckert DJ, Owens RL, Kehlmann GB, et al. Eszopiclone increases the respiratory arousal threshold and lowers the apnoea/hypopnoea index in obstructive sleep apnoea patients with a low arousal threshold. *Clin Sci (Lond)*. 2011;120(12):505–514.
68. Edwards BA, Andara C, Landry S, et al. Upper-airway collapsibility and loop gain predict the response to oral appliance therapy in patients with obstructive sleep apnea. *Am J Respir Crit Care Med*. 2016;194(11):1413–1422.
69. Iranzo A, Fernandez-Arcos A, Tolosa E, et al. Neurodegenerative disorder risk in idiopathic REM sleep behavior disorder: study in 174 patients. *PLoS One*. 2014;9(2):e89741.
70. Schenck CH, Boeve BF, Mahowald MW. Delayed emergence of a parkinsonian disorder or dementia in 81% of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder: a 16-year update on a previously reported series. *Sleep Med*. 2013;14(8):744–748.
71. McCarter SJ, St Louis EK, Sandness DJ, et al. Antidepressants Increase REM sleep muscle tone in patients with and without REM sleep behavior disorder. *Sleep*. 2015;38(6):907–917.
72. Ferri R, Gagnon JF, Postuma RB, Rundo F, Montplaisir JY. Comparison between an automatic and a visual scoring method of the chin muscle tone during rapid eye movement sleep. *Sleep Med*. 2014;15(6):661–665.
73. Ferri R, Manconi M, Plazzi G, et al. A quantitative statistical analysis of the submental muscle EMG amplitude during sleep in normal controls and patients with REM sleep behavior disorder. *J Sleep Res*. 2008;17(1):89–100.

74. Frauscher B, Gabelia D, Biermayr M, et al. Validation of an integrated software for the detection of rapid eye movement sleep behavior disorder. *Sleep*. 2014;37(10):1663–1671.
75. Mayer G, Kesper K, Ploch T, et al. Quantification of tonic and phasic muscle activity in REM sleep behavior disorder. *J Clin Neurophysiol*. 2008;25(1):48–55.
76. McCarter SJ, St Louis EK, Duwell EJ, et al. Diagnostic thresholds for quantitative REM sleep phasic burst duration, phasic and tonic muscle activity, and REM atonia index in REM sleep behavior disorder with and without comorbid obstructive sleep apnea. *Sleep*. 2014;37(10):1649–1662.
77. Thomas RJ, Bianchi MT. Urgent need to improve PAP management: the devil is in two (fixable) details. *J Clin Sleep Med*. 2017;13(5):657–664.
78. Smith MT, McCrae CS, Cheung J, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine clinical practice guideline. *J Clin Sleep Med*. 2018;14(7):1231–1237.
79. American Academy of Sleep Medicine. *International Classification of Sleep Disorders*. 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014.
80. Briones B, Adams N, Strauss M, et al. Relationship between sleepiness and general health status. *Sleep*. 1996;19(7):583–588.
81. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep*. 2010;33(5):585–592.
82. Seixas AA, Henclewood DA, Langford AT, McFarlane SI, Zizi F, Jean-Louis G. Differential and combined effects of physical activity profiles and prohealth behaviors on diabetes prevalence among blacks and whites in the US population: a novel Bayesian belief network machine learning analysis. *J Diabetes Res*. 2017;2017:5906034.
83. Seixas AA, Henclewood DA, Williams SK, et al. Sleep duration and physical activity profiles associated with self-reported stroke in the united states: application of bayesian belief network modeling techniques. *Front Neurol*. 2018;9:534.
84. Wallace ML, Stone K, Smagula SF, et al. Which sleep health characteristics predict all-cause mortality in older men? an application of flexible multivariable approaches. *Sleep*. 2018;41(1).
85. Buysse DJ. Sleep health: can we define it? Does it matter? *Sleep*. 2014;37(1):9–17.
86. Searles Nielsen S, Warden MN, Camacho-Soto A, Willis AW, Wright BA, Racette BA. A predictive model to identify Parkinson disease from administrative claims data. *Neurology*. 2017;89(14):1448–1456.
87. US Food & Drug Administration. Developing a Software Precertification Program: A Working Model. <https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM605685.pdf>. Published April 2018. Accessed February 14, 2020.
88. Agency for Healthcare Research and Quality. Alert Fatigue. <https://psnet.ahrq.gov/primer/alert-fatigue>. Updated September 2019. Accessed February 14, 2020.
89. Watson NF, Rosen IM, Chervin RD. The past is prologue: the future of sleep medicine. *J Clin Sleep Med*. 2017;13(1):127–135.
90. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. *N Engl J Med*. 2018;378(11):981–983.
91. Chen JH, Asch SM. Machine learning and prediction in medicine - beyond the peak of inflated expectations. *N Engl J Med*. 2017;376(26):2507–2509.

## ACKNOWLEDGMENTS

The authors thank the AASM staff members who assisted with the development of this article. The authors also are grateful for the feedback provided by the new members of the AASM Artificial Intelligence in Sleep Medicine Committee: Arun Badi, MD, PhD; Hao Cheng, MD; Daniel V. Fabbri, PhD; Thomas Gustafson, MD; and Octavian Ioachimescu, MD, PhD, MBA.

## SUBMISSION & CORRESPONDENCE INFORMATION

**Submitted for publication February 14, 2020**

**Submitted in final revised form February 14, 2020**

**Accepted for publication February 14, 2020**

Address correspondence to: Cathy A. Goldstein MD, University of Michigan Sleep Disorders Center, C728 Med Inn Building, 1500 E. Medical Center Drive, SPC5845, Ann Arbor, Michigan, 48109-5845; Email: [cathygo@med.umich.edu](mailto:cathygo@med.umich.edu)

## DISCLOSURE STATEMENT

The authors constitute the 2018–2019 AASM Artificial Intelligence in Sleep Medicine Committee.