

REVIEW

Open Access



The case for using digital EEG analysis in clinical sleep medicine

Magdy Younes^{1,2}

Abstract

Evaluation of sleep in clinical polysomnograms continues to rely almost exclusively on visual scoring that implements rules proposed by Rechtschaffen and Kales nearly 50 years ago. Apart from its cost and time-consuming nature, visual scoring has limitations including: A) Sleep depth, which is a continuous variable, is treated as if it changes in a stepwise fashion from light (stage 1), to intermediate (stage 2) to deep (stage 3). B) Even with this limited scale, there is considerable inter-scorer variability, particularly in scoring stages 1 and 3 of non-REM sleep, thereby adding uncertainty to %time spent in these stages as a reliable metric for evaluating sleep depth. C) Limitation in scoring some of EEG features, including 1) arousal intensity, 2) extent of Alpha intrusion and 3) frequency, and characteristics of sleep spindles and K complexes. Digital analysis can solve these problems but producing a reliable system has been a challenge. In this review I begin with recent advances in digital scoring of sleep according to the Rechtschaffen and Kales rules and conclude that this technology has progressed enough to make it possible to obtain reliable, reproducible scoring, comparable in accuracy to scoring by highly experienced technologists, with minimal editing. This is followed by description of several new metrics that can be obtained if digital scoring systems were to be used routinely in clinical studies. The scientific evidence supporting the potential of these metrics to positively impact sleep medicine practice and the wide range of such metrics in patients studied in the sleep laboratory are highlighted.

Keywords: Sleep depth, Arousal intensity, Alpha intrusion, Sleep spindles, K Complexes, Odds ratio product, ORP, Michele sleep scoring

Background

The polysomnogram (PSG) is the cornerstone of investigations in clinical sleep medicine. Interpretation of sleep in these studies is based primarily on the scoring rules introduced by Rechtschaffen and Kales (R&K) almost 50 years ago (Rechtschaffen and Kales 1968). In 1996, Kubicki et al. lamented the fact that, up to then, digital analysis of PSGs was focused on implementing the R&K rules more efficiently and not on exploring the micro-structure of sleep which, they argued, could provide clinically important information (Kubicki and Herrmann 1996). In the intervening 20 years, advances in digital technology have revolutionized almost every aspect of our lives. Yet, the only benefit to sleep studies from this digital revolution has been conversion of data format

from ink on paper to fancy digital displays. It is true that we now need much less space to store data, less time to retrieve patient information and the ability to store data on digital media, and change the montage and filters after data collection. But, the medically-helpful information we get has hardly changed. We still divide non-REM sleep into three distinct stages, consider arousals as simply present or absent, and we see all kinds of differences between patients' EEGs that might well explain the patient's problems, but we have no idea what they mean. R&K rules were introduced when visual scoring was the only way to make any sense of the massive data generated from sleep studies. It was not feasible then to propose visual criteria for defining more than a few sleep stages, to ask technologists to count spindles or alpha bursts or characterize their durations and intensity...etc. We have learned from basic science that differences in these features may mean something, but because we don't get this information in clinical studies we can't

Correspondence: mkyounes@shaw.ca

¹Sleep Disorders Centre, Misericordia Health Centre, Winnipeg, Canada

²YRT Ltd, Winnipeg, MB, Canada



determine the clinical utility of measuring them. And, since we don't have proof that measuring these differences will impact patient care, why should we change how we score clinical studies? A catch 22!

Over the past four decades several dozen systems were proposed for digital sleep scoring and some of these have been validated for clinical use. There is, however, extreme resistance to the use of such systems in clinical practice. The usual excuse is that they are not reliable enough and require much human editing, thereby defeating their primary purpose of economy, speed and consistency. In this review I am hoping to make a strong case for using digital analysis of the EEG routinely in clinical studies. This case is based on two main arguments:

- 1) Criticism of digital systems' ability to reproduce R&K staging is no longer justified.
- 2) Even if full manual editing is still insisted upon, and the economy achieved by digital scoring according to R&K is not large, including digital scoring routinely in clinical studies would make it possible to easily obtain potentially valuable information that is not possible to obtain with visual scoring.

Although the evidence provided below in support of these arguments is primarily from my own work (Malhotra et al. 2013; Azarbarzin et al. 2014; Younes et al. 2015a; Azarbarzin et al. 2015; Younes et al. 2015b; Younes and Hanly 2016; Younes et al. 2015c; Meza et al. 2016; Younes et al. 2016; Younes and Hanly PJ 2016; Amatory et al. In Press; Younes) (I found no other relevant work), such evidence can be produced by anyone who is involved in digital scoring, and the additional information I generated in my own system can be

generated by other systems. My intention is to simply encourage the use of digital scoring systems in clinical practice, regardless of which system is used. Availability of digital EEG analysis in every PSG system would greatly facilitate the introduction, testing and utilization of specialized information on microstructure. My own results are simply used here to illustrate some of what might be achieved with digital analysis.

Comparison between manual and automatic scoring of sleep according to Rechtschaffen and Kales

The inconsistent results and time-consuming, expensive nature of manual sleep staging are well recognized. For this reason, numerous attempts have been made to automate this process (See Penzel et al. (Penzel et al. 2007) and Lajnef et al. (Lajnef et al. 2015a) for a listing of various automatic systems proposed). Of the several dozen systems tried so far three have shown enough promise to be used in clinical studies and are commercially available (Malhotra et al. 2013; Younes et al. 2015b; Pittman et al. 2004; Anderer et al. 2005; Punjabi et al. 2015). Agreement between these systems and expert scorers for sleep variables is similar to agreement found between two scorers (Malhotra et al. 2013; Younes et al. 2015b; Younes et al. 2015c; Younes et al. 2016; Pittman et al. 2004; Anderer et al. 2005; Punjabi et al. 2015). One of these systems (Michele Sleep Scoring, MSS, YRT Ltd, Winnipeg, Canada) has received the most evaluation and its results (Younes et al. 2015b; Younes et al. 2015c; Younes et al. 2016) are shown in Table 1 in juxtaposition to reported agreement between two expert technologists. The data reported by

Table 1 Agreement between MSS and manual scoring compared to agreement between two scorers

Variable	ICCs MSS vs. Manual scoring			ICCs Between two scorers		
	Malhotra et al. ^a	Younes et al. ^b	Younes et al. ^c	Within site ^a Mean (range)	Between sites ^a Mean (95% CI)	From other sources ^d
Total sleep time	0.87	0.96	0.92	0.89(0.78–0.98)	0.87 (0.85–0.89)	
Sleep Efficiency	0.74	0.84	0.76	0.80 (0.69–0.96)	0.77 (0.73–0.80)	
Non-REM stage 1	0.56	0.65	0.63	0.62(0.39–0.80)	0.44 (0.39–0.49)	
Non-REM stage 2	0.84	0.85	0.80	0.75(0.49–0.90)	0.61 (0.57–0.66)	
Non-REM stage 3	0.47	0.65	0.74	0.56(0.27–0.83)	0.40 (0.35–0.45)	
REM Sleep	0.64	0.85	0.84	0.78(0.64–0.92)	0.69 (0.64–0.72)	
Sleep onset latency		0.72	0.76			0.30–1.00
REM onset latency	0.55	0.77	0.65	0.67(0.32–0.90)	0.55 (0.50–0.59)	
#Arousals-REM sleep	0.39			0.55(0.28–0.88)	0.52 (0.47–0.57)	
#Arousals-non REM sleep	0.83			0.59(0.24–0.75)	0.58 (0.53–0.62)	
Arousal/awakening Index		0.75	0.72			0.09–0.85

ICC intraclass correlation coefficient, CI confidence interval, MSS Michele sleep scoring, REM rapid-eye-movement sleep; ^aFrom (Malhotra et al. 2013); ^bFrom (Younes et al. 2015b); ^cFrom (Younes et al. 2015c); ^dCompiled from (Danker-Hopfe et al. 2004), (Pittman et al. 2004), (Magalang et al. 2013), and (Zhang et al. 2015)

Malhotra et al. (Malhotra et al. 2013) is used as the primary source for this information because each PSG was scored by 10 technologists from five academic centers and because the same scoring guidelines were used by MSS and technologists. Furthermore, their results, reported in Table 1, are representative of earlier reports (Pittman et al. 2004; Anderer et al. 2005; Ferri et al. 1989; Norman et al. 2000; Collop 2002; Danker-Hopfe et al. 2004; Magalang et al. 2013). Where information was missing in Malhotra's study, reference values were compiled from other sources (Pittman et al. 2004; Danker-Hopfe et al. 2004; Magalang et al. 2013; Zhang et al. 2015).

Table 1 shows that agreement between *unedited* MSS and the average of 10 scorers (leftmost column) is well within the range of ICCs observed for comparisons between two scorers in the same institution (fourth column) or between scorers in different institutions (fifth column) and exceeds the average between-site ICCs in several sleep variables. In two subsequent validation studies (second and third columns, Table 1), utilizing a newer version of MSS, agreement of unedited MSS scores with manual scoring (one scorer) was also within the range of agreement reported by Malhotra et al. (Malhotra et al. 2013) for comparison between two scorers (Table 1). In another study in which 5-stage epoch-by-epoch comparisons were made between MSS scores and the scores of two academic technologists, MSS scores agreed with the scoring of one or both scorers in 87% of epochs (Younes et al. 2016). The comparable figures in the literature range from 68.0% to 82.6% (Pittman et al. 2004; Anderer et al. 2005; Norman et al. 2000; Danker-Hopfe et al. 2004; Rosenberg and Van Hout 2013). In the same study (Younes et al. 2016) each PSG was scored seven times, three times by each of two technologists and once by a third technologist. A true scoring error was defined as one that was not assigned in any of the other sessions. The number of errors made by any of the technologists in a single session averaged 13 epochs/PSG. The corresponding number for the unedited Auto score was 23 epochs/PSG, a clinically insignificant difference of 10 epochs/PSG (<2% of epochs), even without editing.

Notwithstanding these good results, there has been tremendous resistance to using any of the three validated systems in clinical laboratories. As judged from my experience with our own system (MSS) the main reason for this resistance is that Auto-scoring differs from manual scoring by local technologists in many epochs. This necessitates editing. Because the location of epochs with scoring differences cannot be predicted, all epochs need to be reviewed. There is little saving in time or expense.

In a recent informal evaluation in our laboratory, we found that technologists spend between 30 and 120 min

editing files scored by MSS. How can it be necessary to edit so much when it has been proven through rigorous studies (Malhotra et al. 2013; Younes et al. 2015b; Younes et al. 2015c; Younes et al. 2016) that, *without editing*, the summary results are not that different from manual scoring? To investigate this issue, we introduced an algorithm in MSS that recorded all editing actions taken and the time spent on editing. Technologists fully edited the automatic scoring of 42 PSGs (Younes et al. 2015b). Intraclass correlation coefficients (ICCs) for agreement between manual and auto-scoring *before editing* were 0.94 for TST, 0.76 for SE, 0.87 for stage W, 0.63 for N1, 0.81 for N2, 0.55 for N3, and 0.86 for REM sleep. Notwithstanding the fact that these ICCs were well within the range seen between expert scorers (Pittman et al. 2004; Anderer et al. 2005; Ferri et al. 1989; Norman et al. 2000; Collop 2002; Danker-Hopfe et al. 2004; Magalang et al. 2013; Zhang et al. 2015) the technologists performed an average 90 ± 47 changes/file to the automatic sleep stage (Younes et al. 2015b). These changes were often in opposite directions and involved many types of changes such that the net effect on duration of any sleep stage was clinically insignificant except in a few patients. Having found that the vast majority of editing changes are of little clinical consequence (e.g. changing TST from 360 to 350 min, or REM time from 35 to 45 min), we introduced a feature in MSS (Editing Helper) that scans the summary results looking for potential errors that, in our judgement, may influence clinical management (Younes et al. 2015b). These potentially significant errors included very early sleep or REM onset, too much awake time, N3 time or REM time, too little REM time...etc. The technologists were then asked to edit the automatic scoring of 102 full PSGs, once doing a full edit and once following only the suggestions of the Editing Helper. This group included 49 patients with sleep apnea, 12 patients with periodic limb movements $>15 \text{ hr}^{-1}$, 14 patients with insomnia and 27 patients with no pathology. The Helper issued an average of 2.5 ± 1.2 suggestions per file (Younes et al. 2015b). Editing time was reduced from 59 ± 26 to 6 ± 7 min while the ICCs for comparisons between manual and the abbreviated editing were not different from the ICCs for manual vs. full edit comparisons (0.87 ± 0.08 vs. 0.89 ± 0.09) (Younes et al. 2015b).

In a more recent study (Younes and Hanly PJ 2016) epoch by epoch agreement in 5-stage sleep scoring between two senior technologists was 78.9% (kappa statistic = 71.1%). When the scoring of each technologist was independently edited based on features calculated within MSS (odds-ratio-product (Younes et al. 2015a), spindles, K complexes, delta wave duration), % agreement increased to 96.5% (kappa statistic = 95.1%). ICCs for comparisons between the edited and original manually

determined times in different stages were excellent and well within the accepted range for agreement between two expert scorers (Younes and Hanly PJ 2016). Thus, using the features generated by MSS to edit the manual scores essentially eliminated inter-observer variability while the edited score was in acceptable agreement with the original scoring of both technologists. This shows that the features extracted by MSS to stage sleep are a good compromise between the scoring of these features by two expert technologists.

It is not clear what more evidence is needed to convince decision makers that digital systems currently exist that can save a lot of time and money while producing consistent reliable results. A paradox exists in this regard. As illustrated by the numerous studies reporting on inter-observer scoring variability (Pittman et al. 2004; Anderer et al. 2005; Norman et al. 2000; Danker-Hopfe et al. 2004; Rosenberg and Van Hout 2013) the most one can expect between two experts scoring sleep in the same PSGs is agreement in 85% of epochs. This means that it is acceptable to have 120 disagreements between two humans in a typical 800-epoch PSG. Yet, any difference between the automatic score and the scoring of the local physician/technologist is unacceptable! This paradox has two fundamental underpinnings. First, numerous very poor digital systems were previously introduced and failed miserably. This resulted in a general mistrust of automatic systems. Second, decisions to implement, or not, a new digital scoring system are made by a local expert, or by an administrator who relies on local experts. When evaluating the new system, the local expert will inevitably find differences between his/her scoring and the Auto-score. Human nature causes one to trust his/her own scoring over that of another system, regardless of how many validation studies were published, and the decision will almost invariably be against the digital system. It follows that use of automatic scoring will only become commonplace if payers or regulatory bodies encourage/promote its use, with obvious stipulations as to what is an acceptable automatic system and the editing required. The problem with regulatory bodies is that they also rely on experts, most of whom have, for the reasons mentioned above, already decided that automatic scoring is inaccurate. This impasse will remain as the main barrier to moving sleep staging into the 21st century.

Enhancements to conventional sleep scoring

Assessment of sleep depth and quality

Insomnia and non-restorative sleep, are very common in the general population (Ohayon and Reynolds 2009; Ohayon 2005) and in patients with cognitive and psychiatric disorders (Nissen et al. 2006; Riemann and Voderholzer 2003; Baglioni et al. 2010). Sleep studies in such patients

sometimes reveal organic disorders such as sleep apnea or a movement disorder. However, in many cases, sleep studies provide no explanation for the patient's complaint. Management of such patients is problematic. This is not a trivial issue since there is increasing evidence that poor sleep is a risk factor for cognitive impairment (Nissen et al. 2006; Altena et al. 2008), mood disorders (Riemann and Voderholzer 2003; Baglioni et al. 2010), weight gain (Patel and Hu 2008; Spiegel et al. 2009), diabetes (Nilsson et al. 2004; Mallon et al. 2005), and increased overall mortality (Gallicchio and Kalesan 2009; Cappuccio et al. 2010).

A normal sleep study in the face of sleep complaints may indicate either that the complaint represents a perception problem or that the criteria currently used to evaluate sleep quality are not sensitive enough to identify poor sleep. There are reasons to believe the latter proposition and that, as suggested by Jackson et al. (Jackson and Bruck 2012), analysis of sleep microstructure may provide a fruitful alternative for uncovering differences during sleep in these individuals:

A) *Sleep Depth is Not Adequately Described by the Conventional R&K Stages*: Figure 1 shows 6 epochs representing progression of EEG (C3/A2) from full wakefulness (Panel A), to deep sleep (stage N3, panel F). In panel B the dense beta activity seen in panel A disappeared and a sleep pattern appeared in the middle of the epoch (horizontal bar). Despite the marked difference in appearance between epochs A and B, epoch B continued to be staged awake because the alpha pattern occupied > 15 s (Berry et al. 2012). A little later, the sleep pattern extended for 18 s (panel C). This epoch is now staged N1 even though it is much closer in appearance to panel B than panel B is to panel A. The same pattern continued in the next epoch but a spindle appeared (panel D). The stage is now N2, even though the EEG looks very similar to that in panel C (staged N1) and panel B (staged awake). In panel E, the EEG is substantially different from that of panel D and much closer to the EEG in panel F (staged N3). Yet, it is still staged N2 because delta wave duration did not reach 6 s (Berry et al. 2012). This figure illustrates that: a) unlike the stepwise progression of R&K, sleep progresses gradually from full wakefulness to deep sleep, and b) the same stage may include a wide range of sleep depths. Clearly, 4 h of N2 with a D pattern, cannot be equated with 4 h of N2 consisting primarily of pattern E.

Apart from the above consideration, scoring of stages N1 and N3 is subject to much inter-rater variability (Malhotra et al. 2013; Anderer et al. 2005; Punjabi et al. 2015; Ferri et al. 1989; Norman et al. 2000; Collop 2002; Danker-Hopfe et al. 2004; Magalang et al. 2013; Zhang et al. 2015; Rosenberg and Van Hout 2013). In a recent study in which each epoch was scored manually three

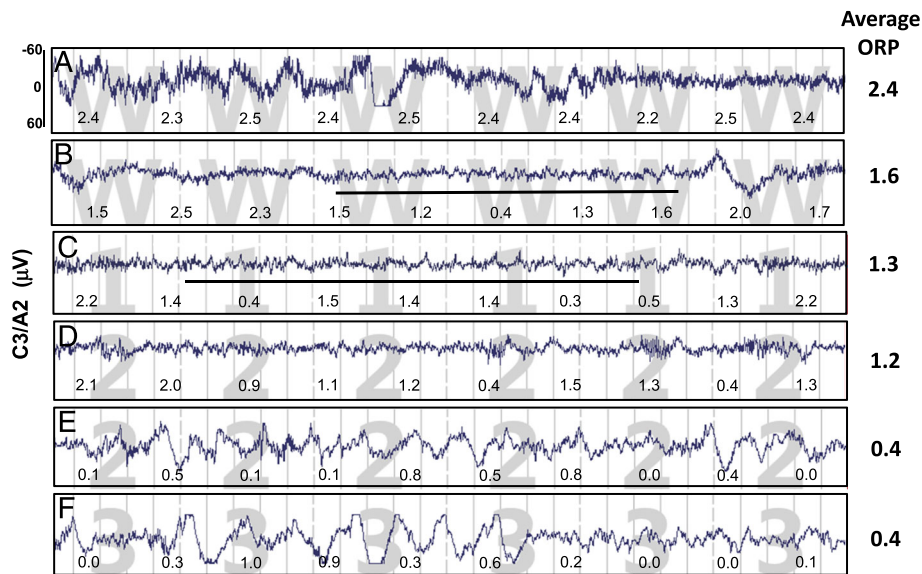


Fig. 1 Tracings showing progression from full wakefulness to deep sleep. Both tracings **a** and **b** were staged awake despite their substantially different visual appearance and the presence of a 12-second period of sleep in **b** (horizontal line; 15 s of sleep are required to score sleep). Although its pattern is substantially similar to that of tracing **b**, tracing **c** was scored asleep (stage 1) because the sleep pattern lasted 18 s (horizontal line). Tracings **d** and **e** were both scored stage 2 even though the pattern in **d** is very similar to stage 1 (**c**), but a spindle appeared, while the pattern in **e** is very similar to delta sleep (**f**) but the duration of delta waves was just shy of 6 s. The numbers within each tracing are the 3-second odds-ratio-product (ORP) values and the number to the right of each tracing is the 30-second ORP average. Note the marked difference in ORP distribution within the two awake tracings and the transient reduction in ORP during the brief sleep in tracing **b**. ORP values are quite different between the two stage 2 epochs with ORP in tracing **d** being close to that in tracing **c** while ORP in tracing **e** being similar to that of tracing **f**. C3/A2, electroencephalogram. ORP reflects the visual appearance of the EEG. Adapted from (Younes et al. 2015a)

times by two senior technologists and once by another senior technologist (i.e. 7 scores for each epoch) the likelihood of an epoch scored N1 by any technologist being scored N1 in all other six scores was only 9.7% (Younes and Hanly PJ 2016). The corresponding likelihood for an epoch scored N3 by at least one scorer was 24%. Given this uncertainty about the scores of two of the three non-REM stages, it is difficult to have much confidence in using the fractions of time spent in each of these stages as indices of sleep quality/depth.

B) Conventional Indices of Sleep Quality are Difficult to Interpret: Sleep quality is conventionally evaluated by a number of variables that include sleep onset latency (SOL), sleep efficiency (SE), times spent in different non-REM stages, and extent of sleep disruption (e.g. arousal and awakening index (A/AW index) and wake after sleep onset (WASO)). The normal range for each of these variables is very wide so that it is difficult to conclude that a person's sleep is poor unless one of these variables is grossly abnormal. Furthermore, in comparing two groups of patients, or the same patients in two conditions, the differences are often contradictory, for example SE may be better (higher) but N1% and/or A/AW index may be worse. Since the units of measurement of these indices are different (% , rate, minutes) the contradictory changes cannot be integrated into a unitary

index that describes the net difference in sleep quality between the two groups/interventions.

It is clearly impractical to visually divide sleep into many more stages than the current ones or to visually assign a number to overall sleep quality. Digital analysis is required. The first step in staging sleep by MSS is to generate an index called odds-ratio-product (ORP). ORP is the probability of 3-second EEG segments falling in epochs staged awake by a consensus of expert technologists (Younes et al. 2015a). Each 3-second segment is subjected to fast Fourier analysis to generate the power in four frequency bands; delta (0.33–2.33 Hz), theta (2.66–6.33 Hz), alpha/sigma (7.33–14 Hz), and beta (14.33 to 35.0 Hz). The power in each band is assigned a rank (0 to 9) depending on where it lies within the entire range of powers (in the relevant band) observed in >400,000 artifact-free 3-second segments collected from clinical PSGs (many with severe sleep fragmentation). Each 3-second segment is then assigned a four-digit number (bin number), consisting of the four ranks in succession, resulting in 10,000 possible spectral patterns. For example bin number 9549 describes a segment with very high delta power, average theta power, average alpha power and very high beta power. Figure 2 shows several examples of 3-second EEG tracings with their bin numbers.

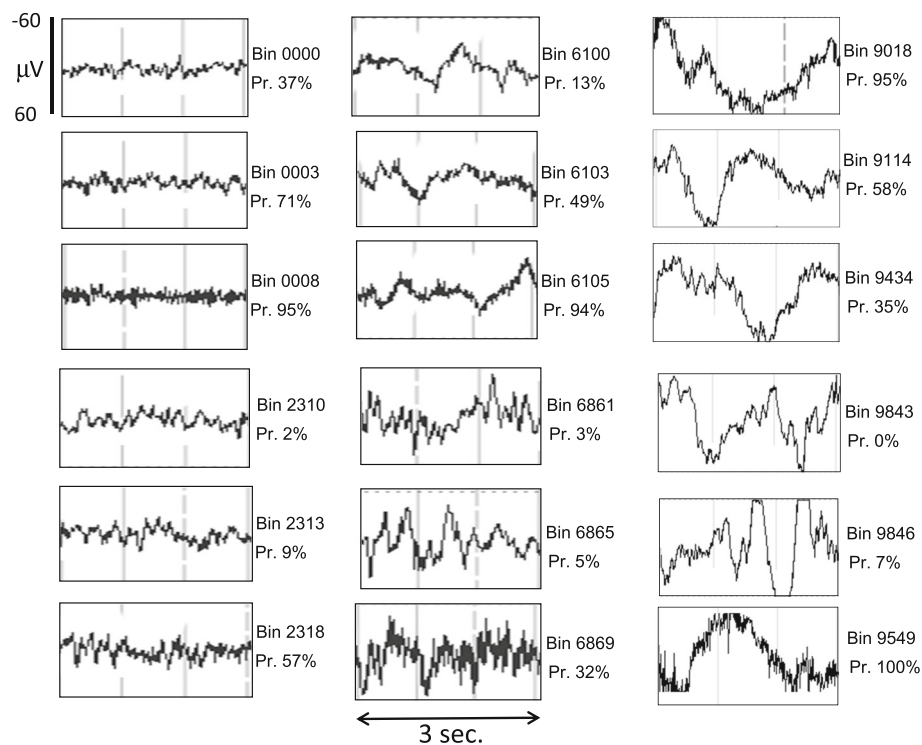


Fig. 2 Three-second EEG tracings (C3/A2) showing a range of patterns, their bin numbers and probability (Pr.) of the pattern occurring during periods scored awake or as arousals by a consensus of expert scorers. The four digits in the bin number represent the normalized powers in (from left to right) delta, theta, alpha/sigma and beta frequency ranges. Note that a variety of patterns can share the same probability. The probability is determined by the relation of the 4 powers to each other. It increases as alpha and beta powers (last two digits) increase and as theta power (second digit) decreases. High delta power may increase or decrease the probability depending on the other powers (5)

A look-up table is consulted to determine the probability of each of the 10,000 patterns occurring in 30-second epochs scored awake or during an event scored as an arousal. This table was constructed from the results of manual scoring, by three very senior technologists, of the same PSGs containing the 3-second segments used to generate the 10,000 bin numbers. This table indicates that a segment with, for example, a bin number of 9549 was never seen except during arousals or in epochs staged awake (i.e. probability is 100%, Fig. 2). On the other hand, the table indicates that only 7% of segments with bin number 9846 are seen during wakefulness or arousals; the probability is 7% (Fig. 2). The 0 to 100 probability values were all divided by 40 (% awake time in the development PSGs) to generate the ORP. Thus, an ORP of 2.5 indicates a pattern that only occurs during wakefulness or in arousals, an ORP of 1.25 indicates a pattern with an equal probability of occurring during wakefulness or sleep and an ORP of 0 never occurs during wakefulness or in arousals.

The average of the 10 ORP values in each 30s epoch is the primary variable used in MSS to determine whether the patient is almost certainly awake (ORP >2.0), almost certainly asleep (ORP <1.0) or is in an intermediate state

(ORP 1.0–2.0) (Younes et al. 2015a). Epochs with intermediate values are staged awake or asleep based on a number of ancillary features. As indicated earlier, this staging system has proved quite accurate (Malhotra et al. 2013; Younes et al. 2015b; Younes et al. 2015c; Younes et al. 2016; Younes and Hanly PJ 2016). However, other than its utility in distinguishing wakefulness from sleep, ORP proved to be a good continuous measure of sleep depth, as indicated by the following observations (Younes et al. 2015a):

(A) *Relation between ORP and conventional sleep stages:* Figure 1 shows that average ORP in the 30s epochs decreased progressively as stage moved from full wakefulness (panel A) to deep sleep (panel F). Reflecting what the eye sees, there is little change in average ORP as stage moved from a “dozing” awake state (panel B) through stage 1 (Panel C) and early stage 2 (panel D) while there was a large drop in ORP within stage 2 as the pattern changed from one that resembles stage 1 (panel D) to a pattern that resembles stage 3 (panel E). Spindles are a traditional marker of stage 2 (Rechtschaffen and Kales 1968; Berry et al. 2012). However, spindles are present throughout stage 2, regardless of whether the background EEG pattern is close to that of stage 1 or stage 3, and there

is no evidence that the first appearance of a spindle (or K complex) on an EEG background of stage 1 represents a major shift in sleep depth.

Figure 3 shows average ORP in all epochs scored in each of the conventional stages in individual patients. The data in this table are from 317 patients used in five different studies in two sleep centers ((Younes et al. 2015a; Younes et al. 2015c; Meza et al. 2016; Younes) and one internal study). These studies included males and females with a very wide range of age and body habitus. Studies 1, 2 and 4 included some patients with no sleep pathology and with insomnia but the majority had different degrees of sleep apnea and, to a lesser extent, PLMs (Younes et al. 2015a; Younes et al. 2015c; Younes). The internal Study (study 3) included exclusively patients with moderate/severe obstructive sleep apnea. Study 5 (Meza et al. 2016) included 30 patients with excessive daytime somnolence (Epworth scale 15 ± 3) and no sleep pathology during the nocturnal PSG. It can be seen that in all five studies average ORP decreased progressively as stage progressed from awake to N3 but that within each stage, except N3, values in individual patients varied over a wide range. The lowest values were in stage N3 where average ORP was <0.4 in all studies and it exceeded 0.6 in only 12 of 252 patients who developed N3 sleep.

That ORP is consistently close to zero in N3 regardless of age, gender, body habitus, sleep pathology, equipment used or the sleep center where the study was performed (Fig. 3), is of particular importance since it indicates that when the patient is in the deepest sleep by current criteria, ORP is always near zero. This supports the notion that differences in ORP within other stages reflect

differences in sleep depth. Had these differences been related to individual or technical differences ORP in stage N3 would have shown the same degree of variability.

(B) *Relation between ORP and Arousability*: By its very name, sleep depth reflects the ease with which the brain arouses in response to arousal stimuli (arousability); the lighter sleep is, the easier it is to arouse from a given stimulus (Philip et al. 1994; Roehrs et al. 1994; Berry et al. 1998). Arousal occurs when the intensity of a spontaneous or induced stimulus exceeds the intensity required to cause arousal; the arousal threshold. Thus, a high arousal/awakening index could indicate a generally low arousal threshold (light sleep), a high frequency of intense arousal stimuli, or both. Distinction between these two mechanisms is of obvious importance in determining the cause of sleep fragmentation.

Since ORP is lowest in N3 (Fig. 2) and arousability is lowest (Philip et al. 1994; Roehrs et al. 1994; Berry et al. 1998), it can be assumed that ORP correlates with conventionally measured arousal threshold. However, we wanted to determine if differences in ORP within lighter stages (Figs. 1 and 2) also reflect different degrees of arousability. Measurement of arousal threshold is typically performed by applying stimuli (e.g. sound) with different intensities and determining the minimum stimulus intensity that causes EEG arousal (Philip et al. 1994; Roehrs et al. 1994; Martin et al. 1997). Alternatively, if the arousal stimulus is known, it is determined by measuring the intensity of the known stimulus just before arousal (e.g. determining arousal threshold in OSA by measuring pharyngeal pressure just before arousal (Berry and Gleeson 1997)). Neither approach can be used to determine arousal threshold for spontaneous unknown stimuli. Furthermore,

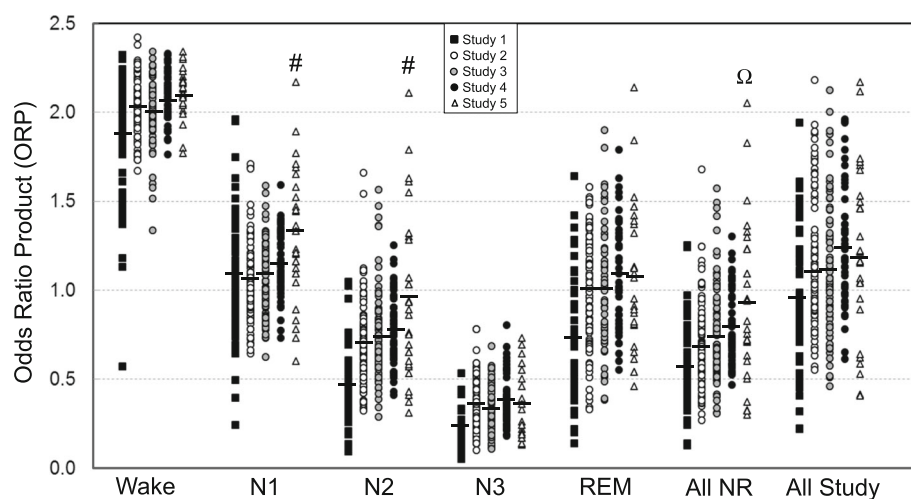


Fig. 3 Odds-ratio-product (ORP) in different conventional sleep stages compiled from five different studies. Each dot is the average ORP in all epochs staged as a given stage in one patient. N1-N3, stages 1–3 of non-rapid-eye-movement (NR) sleep. #, significantly higher than in all other studies; Ω , significantly higher than studies 1–3. Study 1, from (Younes et al. 2015a). Study 2, from (Younes et al. 2015c). Study 3, internally acquired validation data in 79 patients with moderate/severe obstructive sleep apnea. Study 4, from (Younes). Study 5, from (Meza et al. 2016)

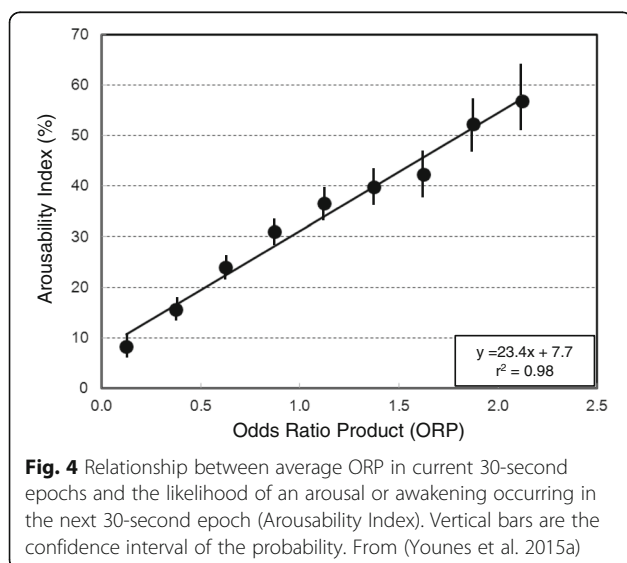
with the first approach, multiple stimulus intensities need to be applied to determine the minimum intensity at which arousal occurs and the sequence needs to be repeated several times during stable sleep to allow for the spontaneous changes in threshold (Philip et al. 1994; Roehrs et al. 1994). Thus, much time of relatively stable sleep is required to determine a single arousal threshold value that represents the average threshold over the period of testing. Accordingly, current approaches cannot be used to determine the instantaneous arousal threshold or the arousal threshold for stimuli responsible for spontaneous arousals.

We used a completely different approach to determine the relation between instantaneous ORP and arousability (Younes et al. 2015a). The approach is based on the reasonable expectations that: a) the brain receives constantly sensory information from all parts of the body, b) the peripheral sensory information is independent of sleep depth, and c) these peripheral inputs include stimuli of different intensities. When the brain is more arousable, more of these spontaneous inputs will exceed the threshold for arousal, resulting in a higher probability for arousals to occur. All 30-second epochs in non-REM sleep were sorted by their ORP value and the total ORP range (0–2.5) was divided into 10 equal mini-ranges, 0.25 each. For each mini-range we determined the likelihood of an arousal/awakening to occur in the following 30-second epoch (arousability index). We pooled the results of 58 patients with assorted sleep disorders in order to filter out inter-individual differences in the range of intrinsic stimulus intensities. There was a near perfect correlation between current ORP and the arousability index (Fig. 4) (Younes et al. 2015a). This figure indicates that when ORP is, for example, 0.5 and the spontaneous sensory stimuli of the patient are comparable in frequency and

intensity to the average of all 58 patients used, the patient is expected to have an arousal every ≈ 5 epochs (20% probability), or an arousal index of 24 hr^{-1} . On the other hand, at an ORP of 1.5 the probability of developing an arousal in the next epoch is $\approx 44\%$, corresponding to an arousal index of 53 hr^{-1} . Clearly, if spontaneous arousal stimuli in a patient are weaker or less frequent than the average in the patients studied here the arousal index is expected to be less than the value predicted from this relation, and vice versa. This can be used to determine whether excessive sleep fragmentation is the result of increased arousability (high ORP) or excessive spontaneous stimuli (see *Potential Clinical Utility of ORP*, below).

The relation shown in Fig. 4 confirms that current ORP reflects the ease with which the brain can be aroused, and can therefore be used as a measure of arousability. It has advantages over conventional arousal threshold determination in that: a) No intervention is required. Therefore, it can be determined in clinical studies. b) Arousability can be determined on an epoch-by-epoch basis making it possible to examine dynamic changes in this important variable.

As may be expected from the data of Fig. 4, there was a strong correlation between average ORP in different sleep stages and the arousal/awakening index in different patients (Younes et al. 2015a). However, it was not clear whether a high average ORP was responsible for the high A/AW index or the converse; average ORP is high because there is more sleep fragmentation. Taking advantage of the newfound ability to measure instantaneous arousability using ORP, in a subsequent study we determined the dynamics of sleep recovery following arousals (Younes and Hanly 2016). We found that patients differ markedly in how quickly sleep deepens following an arousal (Fig. 5a). In all patients ORP increases during arousal and there is a step decrease in ORP at the end of arousal. However, patients differ in the ORP level at the end of this step decrease (about 9 s after the end of arousal) (Fig. 5a). Average ORP at 9 s following arousal (ORP-9) ranged from 0.23 (very deep sleep) to 1.74 (mean \pm SD: 0.70 ± 0.32). When ORP-9 was high (e.g. patient X, Fig. 5a) ORP decreased gradually over a few minutes and, given time without arousal, deep sleep could be reached. There was a very strong correlation between ORP-9 and average ORP in non-REM sleep (Fig. 6). In multiple regression analysis to determine the main correlates with average non-REM ORP (ORP_{NR}), ORP-9 and A/AW index emerged as the only significant correlates, accounting for 83% of the variability in ORP_{NR} (Younes and Hanly 2016). The mechanism by which A/AW index influences average ORP_{NR} is illustrated in Fig. 5b. Thus, the more frequent the arousals, the less time is available for ORP to decrease before the next arousal resets ORP back to ORP-9.



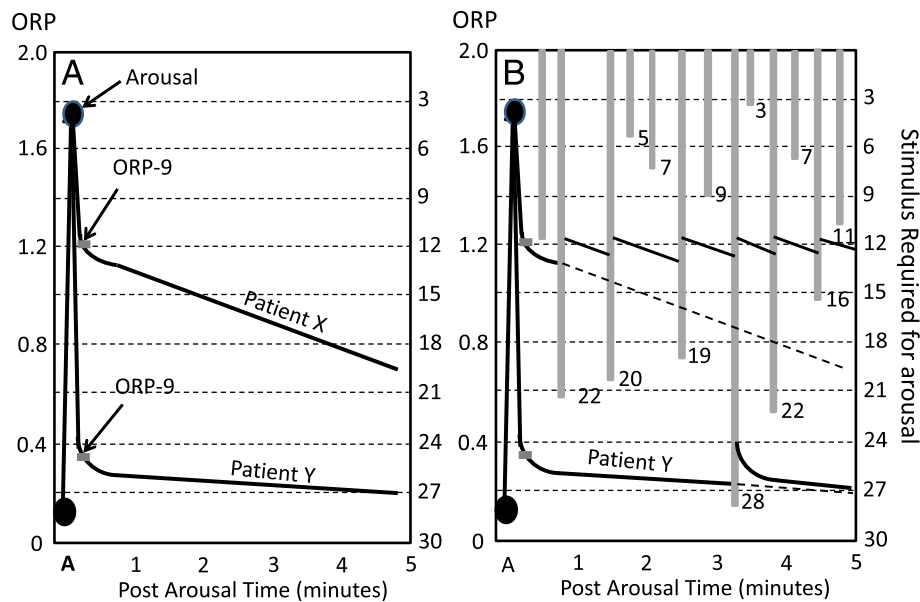


Fig. 5 a Dynamics of sleep depth (odds-ratio-product; ORP) following arousal in two patients. In both cases ORP increases during the arousal and there is a step decrease in ORP immediately (within ≈ 9 s) after the arousal. In patient Y deep sleep returns very quickly after the arousal (ORP at 9 s (ORP-9) is 0.35). In patient X ORP-9 is high (1.2) and sleep deepens gradually over several minutes but only if there are no subsequent arousals. With a high ORP-9, the patient remains in a highly arousable state for several minutes following an arousal. Adapted from (Younes and Hanly 2016). **b** Impact of post-arousal sleep dynamics on sleep continuity and average sleep depth. In the presence of random arousal stimuli of various intensities (grey columns) patient X is more likely to develop another arousal soon after each arousal with the result that sleep can be highly fragmented and average ORP remains high. By contrast, Patient Y is relatively immune to further arousals and average ORP is low throughout

It is interesting to note that ORP-9 was independent of age, gender, body habitus or the sleep disorder (Younes and Hanly 2016). Furthermore, ORP-9 was not affected when patients with severe OSA were placed on CPAP despite reductions in A/AW index and in average ORP. Thus, ORP-9 is likely a trait.

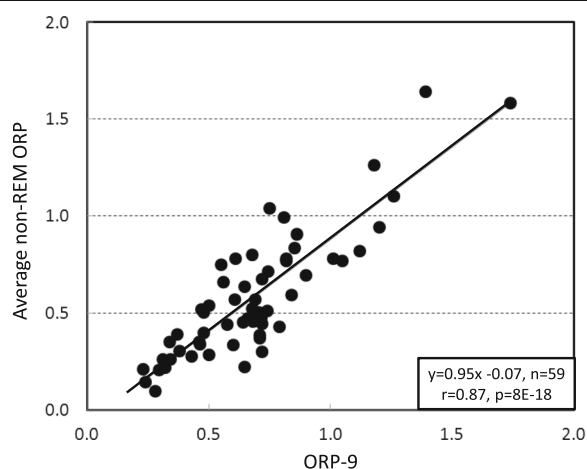


Fig. 6 Relationship between immediate post arousal odds-ratio-product (ORP-9) and average non-REM ORP in patients with sleep apnea. Each dot represents a separate patient. Adapted from (Younes and Hanly 2016)

In summary, average ORP during non-REM sleep is largely determined by the speed with which sleep deepens following arousals. When these dynamics are fast (low ORP-9), the patient returns to deep sleep very quickly and is resistant to all but very intense stimuli. Arousals will be infrequent and the patient remains in deep sleep (patient Y, Fig. 5b). However, when ORP-9 is high, the patient may or not progress to deep sleep depending on the frequency and intensity of arousal stimuli. In the presence of relatively frequent/strong stimuli these patients are more likely to sustain severe sleep fragmentation and to remain in light sleep throughout (patient X, Fig. 5b). It is well known that some healthy subjects are more prone to sleep disruption in the face of stress than others (Bonnet and Arand 2003; Drake et al. 2004) and that this trait may be inherited (Drake et al. 2008; Bonnet and Arand 2010; Heath et al. 1990). Post-arousal sleep dynamics may provide the basis for these inter-individual differences.

Potential Clinical Utility of ORP: Most of the potential applications discussed below are based on what we currently know about what ORP and ORP-9 indicate, and simply represent ideas for future research that can be facilitated by inclusion of ORP in clinical sleep reports:

- a) Investigation of Primary Insomnia: Availability of ORP during periods staged awake may help distinguish

patients who are wide awake during these periods (Fig. 1a) from those who develop frequent mini-sleep periods but fail to progress to sustained sleep (e.g. Fig. 1b). Furthermore, it is now well established that a hyperarousal state exists in some patients with primary insomnia (Bonnet and Arand 2010; Riemann et al. 2010). The EEG representation of the hyperarousal state is an increase in power in the beta frequency range during sleep (Freedman 1986; Krystal et al. 2002; Perlis et al. 2001; Buysse et al. 2008). ORP is extremely sensitive to relative EEG beta power (Younes et al. 2015a). Accordingly, ORP during non-REM sleep should be elevated in patients with the hyperarousal state. It is possible that the etiology in patients who are wide awake during awake periods and develop high quality sleep when they sleep is related to lifestyle issues and these patients may be better candidates for cognitive behavioral therapy, whereas in those who fail to progress to deep sleep the problem is in central sleep control or excessive arousal stimuli from some source. The last two possibilities can be distinguished by ORP-9.

- b) Investigation of paradoxical insomnia: Previous studies identified microstructure abnormalities in patients with paradoxical insomnia (Krystal et al. 2002; Perlis et al. 2001; Parrino et al. 2009). All reported abnormalities (increased beta activity (Perlis et al. 2001), decreased delta and increased alpha/beta activity (Buysse et al. 2008), and cyclic alternating pattern (Parrino et al. 2009)) result in a higher ORP during non-REM sleep (Younes et al. 2015a). Thus, having ORP_{NR} available in clinical studies may help identify patients with abnormal sleep quality from those in whom the problem is in perception of sleep. Furthermore, in a recent study (Meza et al. 2016) in which we measured ORP during multiple sleep latency tests (MSLTs) and correlated ORP values with probability of sleep being perceived after each nap we identified patients who failed to perceive sleep despite reaching ORP levels that were followed by sleep perception in the vast majority of patients. Thus, measurement of ORP and post-nap sleep perception during MSLTs may help identify patients with a central perception abnormality.
- c) Investigation of patients with idiopathic hypersomnia and non-restorative sleep: These symptoms may be related to poor quality sleep or a need to sleep longer than the patient is sleeping. As indicated earlier, sleep may be of poor quality (e.g. Fig. 1d) even though the patient spends a normal amount of time in stage N2. Knowledge of ORP in stage N2 may help identify those in whom the problem is poor sleep quality. In a recent study (Meza et al. 2016), we found that patients with excessive somnolence (Epworth scale 15.2 ± 3.0) and

normal nocturnal sleep had, on average, significantly higher ORP values during non-REM sleep (Study #5, Fig. 2).

- d) Evaluation of the impact of an intervention on sleep quality: A number of interventions are commonly used to improve sleep quality in patients with sleep disorders including mechanical devices (CPAP, mandibular devices) for respiratory sleep disorders and medications for insomnia, depression and nonrestorative sleep. As indicated above, conventional indices of sleep quality are not sensitive enough for this purpose and are often difficult to interpret. ORP may be suited for this purpose.
- e) Enhancements to the multiple sleep latency test (MSLT) (Meza et al. 2016): Despite the fact that MSLTs are resource intensive, expensive and inconvenient to the patient the clinical utility of this test is very limited. Except for the occasional confirmation of narcolepsy, the only information gained is average sleep onset latency (SOL). A short SOL does not necessarily confirm pathologic somnolence since the range is extremely wide in asymptomatic subjects (Levine et al. 1988; Drake et al. 2010). Furthermore, a short SOL simply means that the patient managed to develop 15 s of light sleep (Berry et al. 2012) in no more than one epoch under conditions that are highly conducive to sleep. In many patients sleep does not progress beyond this extremely light phase (Meza et al. 2016). In a recent examination of 150 naps in 30 patients with excessive somnolence SOL was <5 min in 47 naps (21 patients). Of these, ORP decreased below 1.0 in only 13 naps (9 patients) and below 0.5 in only two naps in one patient (Meza et al. 2016). This indicates that within a given SOL there are gradations of objective sleepiness. Although, on average, patients with a SOL <5 min are at greater risk of motor vehicle accidents than those with SOL >10 min (Drake et al. 2010), the risk in an individual patient is difficult to assess from SOL; the difference in accident risk between the two groups was barely significant despite the fact that the study involved >600 patients followed for 10 years (Drake et al. 2010). It is possible that including the times at which different ORP values are reached may provide a better assessment of accident risk in individual patients (Meza et al. 2016).
- f) Real time applications: ORP can be calculated very quickly and a monitor that outputs ORP in real time (every 3 s and/or as a moving average) has become available (Younes et al. 2016). Such information may be useful in monitoring depth of sedation in intensive care units. As well, ORP may be useful in detecting periods of intermittent dozing during activities that require vigilance (e.g. compare Fig. 1a and b).

Assessment of arousal intensity

The visual scoring of arousals involves a binary decision; present or not (Berry et al. 2012). Yet, visually, arousals vary greatly in intensity and duration (Fig. 7). These differences may have clinical implications. In one study in patients with OSA there was a strong correlation between intensity and the magnitude of post-event ventilatory overshoot, suggesting that arousal intensity promotes recurrence of the obstructive events (Younes 2004). It is clearly not practical to expect technologists to assign a visual scale to every arousal, and such a scale would be highly subjective. More recently, a method was developed for digital assessment of arousal intensity on a scale of 1 to 9 (Azarbarzin et al. 2014). Figure 7 shows examples of arousals with different digital scales. There was a linear relation in all patients between arousal scale and heart rate response to arousals (Azarbarzin et al. 2014). Interestingly, both the average arousal intensity and the slope of the relation between intensity and the increase in heart rate varied considerably among subjects (Azarbarzin et al. 2014). Arousal scale was also found to correlate with the magnitude of pharyngeal dilator muscles' response to arousal (Amatoury et al. 2016).

It is possible that differences in arousal intensity between subjects may explain why some patients develop more daytime symptoms than others for the same degree of sleep fragmentation. It is also possible that the heart rate response to arousal may be predictive for

development of cardiovascular complications in patients with sleep fragmentation. Availability of this information in clinical sleep studies would facilitate investigation of these possibilities.

Assessment of alpha intrusion (alpha-delta sleep)

Intrusion of alpha activity in non-REM sleep is frequently present in patients with fibromyalgia, chronic fatigue and non-restorative sleep (Anch et al. 1991; Branco et al. 1994; Moldofsky et al. 1975; Olsen et al. 2013), arthritis (Mahowald et al. 1989), insomnia (Martinez et al. 2010; Riedner et al. 2016) and depression (Hauri and Hawkins 1973; Jaimcharyatam et al. 2011). Although it may be seen in asymptomatic subjects (Scheuler et al. 1983; Horne and Shackell 1991), its prevalence and extent are clearly greater in these disorders such that a clinical association is well established. Whether alpha intrusion causes the patient's complaints, is a consequence of disruptive stimuli, or both is not clear (Riedner et al. 2016; Pivik and Harman 1995; Stone et al. 2008). Large studies are needed to identify its clinical significance.

Visual quantification of the alpha intrusions in routine sleep studies is impractical. Utilization of this potentially useful marker in clinical practice has been hampered by the need for digital analysis and lack of quantitative guidelines for its identification and quantification. In order to generate ORP, our clinical sleep scoring program performs spectral analysis on consecutive 3-second

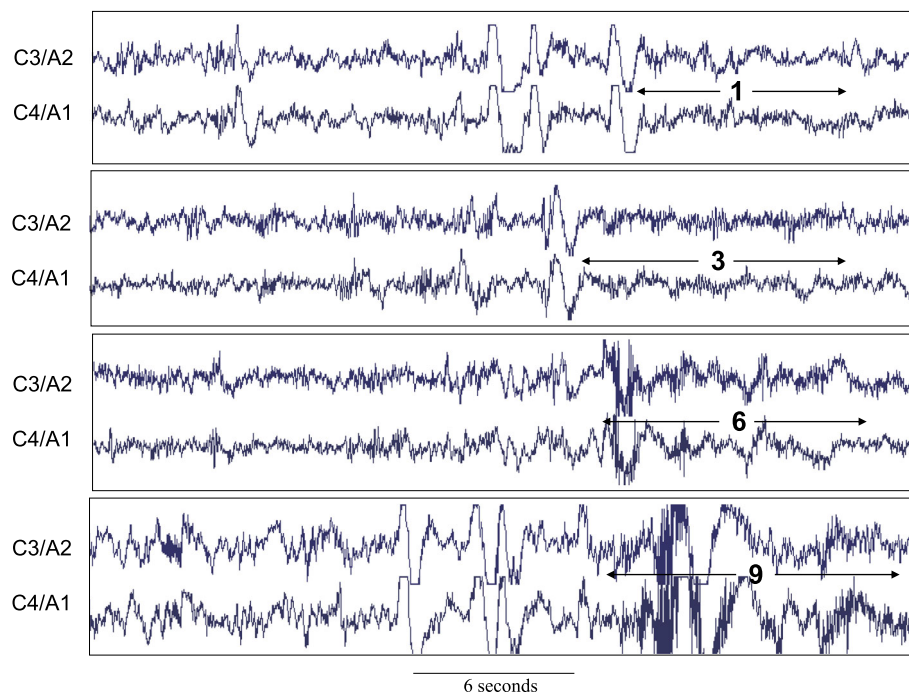


Fig. 7 Examples of arousal with different intensity scales in the same patient. C3/A2 and C4/A1 are central electroencephalograms. From (Azarbarzin et al. 2014)

epochs throughout the PSG and calculates power in different EEG frequency ranges including the alpha range (Younes et al. 2015a). By comparing the 3-second alpha power with the corresponding visual appearance of the EEG we established a threshold of $30 \mu V^2$ for what one can confidently score as alpha intrusion. The fraction of 3-second epochs with ORP <1.0 (i.e. patient is clearly asleep) that meet this threshold is reported as the alpha intrusion index. The alpha intrusion index was evaluated in 448 PSGs scored by MSS. In 60% of patients the index was $<1\%$ while it was $>5\%$ in 15% of patients and $>20\%$ in 4% of patients. This information is presented not to suggest that the threshold for scoring alpha intrusion should be the same as used here. The criteria can obviously be changed with experience or by consensus. Rather, it is presented to show that if clinical PSGs are routinely subjected to digital analysis such an index, regardless of what criteria are used to measure it, can be painlessly obtained in a limitless number of studies, making it possible to identify its clinical significance easily and inexpensively.

Assessment of sleep spindles and K complexes

Sleep spindles have been extensively investigated in research studies (recently reviewed by Clawson et al. (Clawson et al. 2016)). It is clear that spindles are involved in learning and memory (Clemens et al. 2006; Cox et al. 2014; Yotsumoto et al. 2009) and their characteristics are correlated with intelligence and cognitive ability (Geiger et al. 2011; Fogel et al. 2007; Schabus et al. 2006). They are greatly reduced in developmental abnormalities in children (Ellingson and Peters 1980; Selvittella et al. 2009; Godbout et al. 2000; Limoges et al. 2005), in adult schizophrenia (Ferrarelli et al. 2007; Ferrarelli et al. 2010; Manoach et al. 2010; Manoach et al. 2014) and in patients with Parkinson's disease with dementia (Latreille et al. 2015). Spindle density, duration and amplitude decrease with age but the rate of decline is not the same in all subjects (Nicolas et al. 2001; Crowley et al. 2002; Guazzelli et al. 1986; Wei et al. 1999; Principe and Smith 1982). There is a correlation between age-related decline in spindle activity and decline in cognitive functioning (Peters et al. 2008; Mander et al. 2014). Sleep disruption by sleep-related disorders (e.g. sleep apnea,) is frequently associated with cognitive impairment. It is not known whether cognitive impairment resulting from sleep fragmentation, per se, is also associated with reduced spindle activity. Availability of information about spindle characteristics during clinical studies before and after correction of the sleep fragmentation would make it possible to easily address this issue. Furthermore, should it become evident that cognitive impairment due to sleep disruption, per se, is not associated with reduced spindle activity, or is not reversible if

present, it will be clear that reduction of spindle activity is at least a marker for the presence of a neurodegenerative process.

Likewise, K complexes (Rechtschaffen and Kales 1968; Loomis et al. 1939) have been extensively investigated (see (Halasz 2005) (Halasz 2005) and Halasz (Halász 2015), for reviews). K complexes can be routinely evoked by experimental stimuli (e.g. noise or airway obstruction) but also occur spontaneously during stages N2 and N3 of non-REM sleep. Because of the similarity between K complexes and some of the delta waves encountered in deep sleep it is possible that some spontaneous complexes represent preliminary appearance of deeper sleep (De Gennaro et al. 2000). However, because the delta waves of deep sleep have no consistent pattern, it is reasonable to assume that a high frequency of slow waves with a consistent K complex appearance reflects a high frequency of naturally occurring noxious stimuli. Knowledge of the frequency of K complexes may therefore be helpful in identifying the presence of excessive sub-threshold arousal stimuli that may contribute to the patient's complaints. Excessive K complexes in the absence of a clear cause in the PSG (e.g. sleep disordered breathing, PLMs) may prompt a search for, and correction of, other somatic or environmental sources of arousal stimuli.

In the current visual scoring of clinical studies spindles and K complexes are used exclusively to identify stage N2 (Berry et al. 2012). There is no consideration of spindle characteristics or frequency of K complexes. Accordingly, such potentially useful information is not captured. A number of digital methods for identifying and characterizing spindles (Ferrarelli et al. 2007; Martin et al. 2013; Wamsley et al. 2012; Mölle et al. 2002; Bódizs et al. 2009; Wendt et al. 2012; Devuyst et al. 2010; Lajnef et al. 2015b; Ray et al. 2015) and K complexes (Bremer et al. 1970; Krohne et al. 2014; Richard and Lengelle 1998; Bankman et al. 1992; Parekh et al. 2015) have been proposed and used in research studies but none has been adapted to clinical studies. Although there is agreement on the visual appearance of these events (Berry et al. 2012), there is no agreement on how to define them in quantitative terms. This and the considerable inter-scorer variability in visual scoring of these events (Warby et al. 2014; Wendt et al. 2015) have made it difficult to arrive at the "optimal" digital approach. In the opinion of this writer, it is preferable to begin using any reasonable approach, and refine it later if necessary, than wait for a consensus that has been elusive for over 50 years. Once such an approach becomes routinely available, its results in normal people can be established, making it possible to determine when a patient's value is abnormal.

As indicated earlier, there are three validated commercial systems for digital sleep scoring (Morpheus,

somnolyzer, MSS). Since all three are capable of breaking down non-REM sleep into its three stages, all include algorithms for spindle and K complex detection. There is no information regarding these algorithms in Morpheus or Somnolyzer. The algorithms in MSS have proven robust enough such that when spindles and K complexes identified by this system were used to edit (overrule) the manual scoring of non-REM sleep N1/N2 discrepancies between scorers decreased from 42 ± 36 epochs/PSG to 10 ± 10 epochs/PSG (Younes and Hanly PJ. Minimizing Interrater Variability in Staging Sleep by Use of Computer-Derived Features. *J Clin Sleep Med* 2016). The algorithms used by MSS to identify spindles and K complexes have been described before (Younes and Hanly PJ 2016). In addition to calculating the frequency of the two events in stage N2, the software also determines average spindle duration and power. In 316 PSGs scored by MSS, average spindle density over all N2 time in each central derivation (C3/M2, C4/M1) ranged 0.2 to 8.9 min^{-1} in different subjects and K complex density ranged from 0.09 to 0.84 min^{-1} . Average spindle duration ranged from 1.05 to 1.7 s and average spindle power ranged from 1.9 to 117.0 μV^2 . These values are presented here to show that there is a wide range in all these characteristics within the clinical sleep population, which likely reflects clinical/pathological differences between patients, and also to show the ease with which a large amount of such information can be obtained if digital identification of these characteristics becomes a routine component of EEG analysis.

Conclusion

It is argued that the technology of digital sleep scoring has progressed enough to make it possible to obtain reliable, reproducible scoring, comparable in accuracy to that of highly trained technologists, with minimal editing. At the same time, it has become clear that inter-rater variability in scoring sleep is sufficiently serious as to raise doubt about the validity of R&K's N1 to N3 stages as a measure of sleep depth. Apart from elimination of variability in conventional sleep assessment between scorers, and reduction in cost, inclusion of digital scoring in the routine analysis of clinical PSGs would make it possible to obtain information that is not possible to obtain with visual scoring. These include providing in each patient a continuous index of sleep depth, and quantitative estimates of arousal intensity, alpha intrusion, and characteristics of spindles and K complexes. Although the clinical utility of these indices has not yet been proven, there is sufficient information from scientific investigations that they might well explain some complaints/disorders that are currently not explained after conventional scoring. Widespread availability of this additional information at no cost and with no intervention would greatly facilitate research into the clinical utility of these indices.

Acknowledgements

The author acknowledges the contributions of all YRT employees, without whose work none of the new technologies could have been developed, and of the many collaborators in the research studies which evaluated these technologies (see references (Malhotra et al. 2013; Azarbarzin et al. 2014; Younes et al. 2015a; Azarbarzin et al. 2015; Younes et al. 2015b; Younes and Hanly 2016; Younes et al. 2015c; Meza et al. 2016; Younes et al. 2016; Younes and Hanly PJ 2016; Amatoury et al. 2016; Younes et al. *J Clin Sleep Med*. In Press.

Funding

Not Applicable.

Availability of data and materials

No datasets were generated or analyzed anew for this review.

Authors' contributions

The author was the sole contributor to this review.

Authors' information

MY (MD, FRCPC, PhD) is distinguished professor emeritus at the University of Manitoba.

Competing interests

MY is the owner of YRT Ltd, the company that developed the scoring system (MSS) and associated algorithms. A patent is pending for the ORP depth of sleep index.

Consent for publication

Not applicable.

Ethics approval and consent to participate

No new experimental recordings were generated for this review. All data provided here that were generated in the author's laboratory was: a) already published (see relevant publications for ethics approval), b) in submitted manuscripts (reference (Younes)) describing studies approved by the Research Ethics Board of the University of Manitoba, c) obtained from previously recorded PSGs used in published studies, or d) clinical PSGs obtained at the Misericordia Sleep Center and scored by MSS (MSS is an approved scoring system by FDA and Health Canada). Patients undergoing clinical sleep studies at our center sign/decline to sign a consent form allowing the use of their anonymized results in research. Only data from patients who agreed to the use of their data was included here.

Received: 23 August 2016 Accepted: 22 December 2016

Published online: 06 February 2017

References

- Altena A, van der Werf YD, Strijers RL, Van Someren EJ. Sleep loss affects vigilance: effects of chronic insomnia and sleep therapy. *J Sleep Res*. 2008;17:335–43.
- Amatoury J, Azarbarzin A, Younes M, Jordan AS, Wellman A, Eckert DJ. Arousal Intensity is a Distinct Pathophysiological Trait in Obstructive Sleep Apnea. *Sleep*. 2016;39:2091–100.
- Anch AM, Lue FA, MacLean AW, Moldofsky H. Sleep physiology and psychological aspects of the fibrositis (fibromyalgia) syndrome. *Can J Psychol*. 1991;45:179–84.
- Anderer P, Gruber G, Parapatics S, Woertz M, Miazhyńska T, Klosch G, 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;51:115–33.
- Azarbarzin A, Ostrowski M, Hanly P, Younes M. Relationship between arousal intensity and heart rate response to arousal. *Sleep*. 2014;37:645–53.
- Azarbarzin A, Ostrowski M, Keenan B, Younes M, Kuna S. Arousal responses during overnight polysomnography and their reproducibility in healthy young adults. *Sleep*. 2015;38:1313–21.
- Baglioni C, Spiegelhalter K, Lombardo C, Riemann D. Sleep and emotions: focus on insomnia. *Sleep Med Rev*. 2010;14:227–38.
- Bankman IN, Sigillito VG, Wise RA, Smith PL. Feature-based detection of the K-complex wave in the human electroencephalogram using neural networks. *IEEE Trans Biomed Eng*. 1992;39:1305–10.

- Berry RB, Gleeson K. Respiratory arousal from sleep: mechanisms and significance. *Sleep*. 1997;20:654–75.
- Berry RB, Asyali MA, McNellis MI, Khoo MC. Within-night variation in respiratory effort preceding apnea termination and EEG delta power in sleep apnea. *J Appl Physiol* (1985). 1998;85:1434–41.
- Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. *J Clin Sleep Med*. 2012;8:597–619.
- Bódizs R, Körmendi J, Rigó P, Lázár AS. The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm. *J Neurosci Methods*. 2009;178:205–13.
- Bonnet MH, Arand DL. Situational insomnia: consistency, predictors, and outcomes. *Sleep*. 2003;26:1029–36.
- Bonnet MH, Arand DL. Hyperarousal and insomnia: state of the science. *Sleep Med Rev*. 2010;14:9–15.
- Branco J, Atalaia A, Paiva T. Sleep cycles and alpha-delta sleep in fibromyalgia syndrome. *J Rheumatol*. 1994;21:1113–7.
- Bremer G, Smith JR, Karacan I. Automatic detection of the K-complex in sleep electroencephalograms. *IEEE Trans Biomed Eng*. 1970;17:314–23.
- Buyse DJ, Germain A, Hall ML, et al. EEG spectral analysis in primary insomnia: NREM period effects and sex differences. *Sleep*. 2008;31:1673–82.
- 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;33:585–92.
- Clawson BC, Durkin J, Aton SJ. Form and function of sleep spindles across the lifespan. *Neural Plast*. 2016;2016:6936381. doi:10.1155/2016/6936381. Epub 2016 Apr 14.
- Clemens Z, Fabo D, Halasz P. Twenty-four hours retention of visuospatial memory correlates with the number of parietal sleep spindles. *Neurosci Lett*. 2006;403:52–6.
- Collop NA. Scoring variability between polysomnography technologists in different sleep laboratories. *Sleep Med*. 2002;3:43–7.
- Cox R, Hofman WF, de Boer M, Talamini LM. Local sleep spindle modulations in relation to specific memory cues. *NeuroImage*. 2014;99:103–10.
- Crowley K, Trinder J, Kim Y, Carrington M, Colrain IM. The effects of normal aging on sleep spindle and K-complex production. *Clin Neurophysiol*. 2002;113:1615–22.
- Danker-Hopfe H, Kunz D, Gruber G, Klösch G, Lorenzo JL, Himanen SL, et al. Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. *J Sleep Res*. 2004;13:63–9.
- De Gennaro L, Ferrara M, Bertini M. The spontaneous K-complex during stage 2 sleep: is it the 'forerunner' of delta waves? *Neurosci Lett*. 2000;291:41–3.
- Devuyst S, Dutoit T, Stenuit P, Kerkhofs M. Automatic K-complexes detection in sleep EEG recordings using likelihood thresholds. *Conf Proc IEEE Eng Med Biol Soc*. 2010;2010:4658–61.
- Drake C, Richardson G, Roehrs T, Scofield H, Roth T. Vulnerability to stress-related sleep disturbance and hyperarousal. *Sleep*. 2004;27:285–91.
- Drake CL, Scofield H, Roth T. Vulnerability to insomnia: the role of familial aggregation. *Sleep Med*. 2008;9:297–302.
- Drake C, Roehrs T, Breslau N, Johnson E, Jefferson C, Scofield H, et al. The 10-year risk of verified motor vehicle crashes in relation to physiologic sleepiness. *Sleep*. 2010;33:745–52.
- Ellingson RJ, Peters JF. Development of EEG and daytime sleep patterns in trisomy-21 infants during the first year of life: longitudinal observations. *Electroencephalogr Clin Neurophysiol*. 1980;50:457–66.
- Ferrarelli F, Huber R, Peterson MJ, Massimini M, Murphy M, Riedner BA, et al. Reduced sleep spindle activity in schizophrenia patients. *Am J Psychiatry*. 2007;164:483–92.
- Ferrarelli F, Peterson MJ, Sarasso S, Riedner BA, Murphy MJ, Benca RM, et al. Thalamic dysfunction in schizophrenia suggested by whole-night deficits in slow and fast spindles. *Am J Psychiatry*. 2010;167:1339–48.
- Ferri R, Ferri P, Colognola RM, Petrella MA, Musumeci SA, Bergonzi P. Comparison between the results of an automatic and a visual scoring of sleep EEG recordings. *Sleep*. 1989;12:354–62.
- Fogel SM, Nader R, Cote KA, Smith CT. Sleep spindles and learning potential. *Behav Neurosci*. 2007;121:1–10.
- Freedman RR. EEG power spectra in sleep-onset insomnia. *Electroencephalogr Clin Neurophysiol*. 1986;63:408–13.
- Gallicchio L, Kalesan B. Sleep duration and mortality: a systematic review and meta-analysis. *J Sleep Res*. 2009;18:148–58.
- Geiger A, Huber R, Kurth S, Ringli M, Jenni OG, Achermann P. The sleep EEG as a marker of intellectual ability in school age children. *Sleep*. 2011;34:181–89.
- Godbout R, Bergeron C, Limoges E, Stip E, Mottron L. A laboratory study of sleep in Asperger's syndrome. *Neuroreport*. 2000;11:127–30.
- Guazzelli M, Feinberg I, Aminoff M, Fein G, Floyd TC, Maggini C. Sleep spindles in normal elderly: comparison with young adult patterns and relation to nocturnal awakening, cognitive function and brain atrophy. *Electroencephalogr Clin Neurophysiol*. 1986;63:526–39.
- Halasz P. K-complex, a reactive EEG graphoelement of NREM sleep: an old chap in a new garment. *Sleep Med Rev*. 2005;9:391–412.
- Halász P. The K-complex as a special reactive sleep slow wave - a theoretical update. *Sleep Med Rev*. 2015;29:34–40.
- Hauri P, Hawkins DR. Alpha-delta sleep. *Electroencephalogr Clin Neurophysiol*. 1973;34:233–7.
- Heath AC, Kendler KS, Eaves LJ, Martin NG. Evidence for genetic influences on sleep disturbance and sleep pattern in twins. *Sleep*. 1990;13:318–35.
- Horne JA, Shackell BS. Alpha-like EEG activity in non-REM sleep and the fibromyalgia (fibrositis) syndrome. *Electroencephalogr Clin Neurophysiol*. 1991;79:271–6.
- Jackson ML, Bruck D. Sleep abnormalities in chronic fatigue syndrome/myalgic encephalomyelitis: a review. *J Clin Sleep Med*. 2012;8:719–28.
- Jaimcharyatam N, Rodriguez CL, Budur K. Prevalence and correlates of alpha-delta sleep in major depressive disorders. *Innov Clin Neurosci*. 2011;8:35–49.
- Krohne LK, Hansen RB, Christensen JA, Sorensen HB, Jennum P. Detection of K-complexes based on the wavelet transform. *Conf Proc IEEE Eng Med Biol Soc*. 2014;2014:5450–3.
- Krystal AD, Edinger JD, Wohlgenuth WK, Marsh GR. NREM sleep EEG frequency spectral correlates of sleep complaints in primary insomnia subtypes. *Sleep*. 2002;25:630–40.
- Kubicki S, Herrmann WM. The future of computer-assisted investigation of the polysomnogram: sleep microstructure. *J Clin Neurophysiol*. 1996;13:285–94.
- Lajnef T, Chaibi S, Ruby P, Aguera PE, Eichenlaub JB, Samet M, et al. Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. *J Neurosci Methods*. 2015a;250:94–105.
- Lajnef T, Chaibi S, Eichenlaub JB, Ruby PM, Aguera PE, Samet M, et al. Sleep spindle and K-complex detection using tunable Q-factor wavelet transform and morphological component analysis. *Front Hum Neurosci*. 2015b;9:414.
- Latreille V, Carrier J, Lafortune M, Postuma RB, Bertrand JA, Panisset M, et al. Sleep spindles in Parkinson's disease may predict the development of dementia. *Neurobiol Aging*. 2015;36:1083–90.
- Levine B, Roehrs T, Zorick F, Roth T. Daytime sleepiness in young adults. *Sleep*. 1988;11:39–46.
- Limoges E, Mottron L, Bolduc C, Berthiaume C, Godbout R. Atypical sleep architecture and the autism phenotype. *Brain*. 2005;128:1049–61.
- Loomis AL, Harvey EN, Hobart GA. Distribution of disturbance-patterns in the human electroencephalogram, with special reference to sleep. *J Neurophysiol*. 1939;2:413–30.
- Magalang UJ, Chen NH, Cistulli PA, Fedson AC, Gislason T, Hillman D, et al. Agreement in the scoring of respiratory events and sleep among international sleep centers. *Sleep*. 2013;36:591–6.
- Mahowald MW, Mahowald ML, Bundlie SR, Ytterberg SR. Sleep fragmentation in rheumatoid arthritis. *Arthritis Rheum*. 1989;32:974–83.
- Malhotra A, Younes M, Kuna ST, Benca R, Kushida CA, Walsh J, et al. Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep*. 2013;36:573–82.
- Mallon L, Broman JE, Hetta J. High incidence of diabetes in men with sleep complaints or short sleep duration. *Diabetes Care*. 2005;28:2762–7.
- Mander BA, Rao V, Lu B, Saletin JM, Ancoli-Israel S, Jagust WJ, Walker MP. Impaired prefrontal sleep spindle regulation of hippocampal-dependent learning in older adults. *Cereb Cortex*. 2014;24:3301–9.
- Manoach DS, Thakkar KN, Stroynowski E, Ely A, McKinley SK, Wamsley E, et al. Reduced overnight consolidation of procedural learning in chronic medicated schizophrenia is related to specific sleep stages. *J Psychiatr Res*. 2010;44:112–20.
- Manoach DS, Demanuele C, Wamsley EJ, Vangel M, Montrose DM, Miewald J, et al. Sleep spindle deficits in antipsychotic-naïve early course schizophrenia and in non-psychotic first-degree relatives. *Front Hum Neurosci*. 2014;8:762. doi:10.3389/fnhum.2014.00762.eCollection2014.
- Martin SE, Wraith PK, Deary IJ, Douglas NJ. The effect of nonvisible sleep fragmentation on daytime function. *Am J Respir Crit Care Med*. 1997;155:1596–601.
- Martin N, Lafortune N, Godbout J, Barakat M, Robillard R, Poirier G, et al. Topography of age-related changes in sleep spindles. *Neurobiol Aging*. 2013;34:468–76.

- Martinez D, Breitenbach TC, Lenz MDCS. Light sleep and sleep time misperception - relationship to alpha-delta sleep. *Clin Neurophysiol*. 2010;121:704–11.
- Meza S, Giannouli E, Younes M. Enhancements to the multiple sleep latency test. *Nat Sci Sleep*. 2016;8:145–58.
- Moldofsky H, Scarisbrick P, England R, Smythe H. Musculoskeletal symptoms and non-REM sleep disturbance in patients with "fibrositis syndrome" and healthy subjects. *Psychosom Med*. 1975;37:341–51.
- Mölle M, Marshall L, Gais S, Born J. Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. *J Neurosci*. 2002;22:10941–7.
- Nicolas A, Petit D, Rompré S, Montplaisir J. Sleep spindle characteristics in healthy subjects of different age groups. *Clin Neurophysiol*. 2001;112:521–7.
- Nilsson PM, Röst M, Engström G, Hedblad B, Berglund G. Incidence of diabetes in middle-aged men is related to sleep disturbances. *Diabetes Care*. 2004;27:2464–69.
- Nissen C, Kloepper C, Nofzinger EA, Feige B, Voderholzer U, Riemann D. Impaired sleep-related memory consolidation in primary insomnia. *Sleep*. 2006;29:1068–73.
- Norman RG, Pal I, Stewart C, Walsleben JA, Rapoport DM. Interobserver agreement among sleep scorers from different centers in a large dataset. *Sleep*. 2000;23:901–8.
- Ohayon MM. Prevalence and correlates of nonrestorative sleep complaints. *Arch Intern Med*. 2005;165:35–41.
- Ohayon M, Reynolds III CF. Epidemiological and clinical relevance of insomnia diagnoses algorithms according to the DSM-IV and the international classification of sleep disorders (ICSD). *Sleep Med*. 2009;10:952–60.
- Olsen MN, Sherry DD, Boyne K, McCue R, Gallagher PR, Brooks LJ. Relationship between sleep and pain in adolescents with juvenile primary fibromyalgia syndrome. *Sleep*. 2013;36:509–16.
- Parekh A, Selesnick IW, Rapoport DM, Ayappa I. Detection of K-complexes and sleep spindles (DETOKS) using sparse optimization. *J Neurosci Methods*. 2015;251:37–46.
- Parrino L, Milioli G, De Paolis F, Grassi A, Terzano MG. Paradoxical insomnia: the role of CAP and arousals in sleep misperception. *Sleep Med*. 2009;10:1139–45.
- Patel S, Hu FB. Short sleep duration and weight gain: a systematic review. *Obesity*. 2008;16:643–53.
- Penzel T, Hirshkowitz M, Harsh J, Chervin RD, Butkov N, Kryger M, et al. Digital analysis and technical specifications. *J Clin Sleep Med*. 2007;3:109–20.
- Perlis ML, Smith MT, Andrews PJ, Orff H, Giles DE. Beta/Gamma EEG activity in patients with primary and secondary insomnia and good sleeper controls. *Sleep*. 2001;24:110–7.
- Peters KR, Rockwood K, Black SE, Hogan DB, Gauthier SG, Loy-English I, et al. Neuropsychiatric symptom clusters and functional disability in cognitively-impaired-not-demented individuals. *Am J Geriatr Psychiatry*. 2008;16:136–44.
- Philip P, Stoohs R, Guilleminault C. Sleep fragmentation in normals: a model for sleepiness associated with upper airway resistance syndrome. *Sleep*. 1994;17:242–7.
- Pittman SD, MacDonald MM, Fogel RB, Malhotra A, Todros K, Levy B, et al. Assessment of automated scoring of polysomnographic recordings in a population with suspected sleepdisordered breathing. *Sleep*. 2004;27:1394–403.
- Pivik RT, Harman K. A reconceptualization of EEG alpha activity as an index of arousal during sleep: all alpha activity is not equal. *J Sleep Res*. 1995;4:131–37.
- Principe JC, Smith JR. Sleep spindle characteristics as a function of age. *Sleep*. 1982;5:73–84.
- Punjabi NM, Shifa N, Doffner G, Patil S, Pien G, Aurora RN. Computerassisted automated scoring of polysomnograms using the somnolyzer system. *Sleep*. 2015;38:1555–66.
- Ray LB, Sockeel S, Soon M, Bore A, Myhr A, Stojanoski B, et al. Expert and crowd-sourced validation of an individualized sleep spindle detection method employing complex demodulation and individualized normalization. *Front Hum Neurosci*. 2015;9:507.
- Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Washington, DC: National Institute of Health, Publ. 204, US Government Printing Office; 1968.
- Richard C, Lengelle R. Joint time and time-frequency optimal detection of K-complexes in sleep EEG. *Comput Biomed Res*. 1998;31:209–29.
- Riedner BA, Goldstein MR, Plante DT, Rumble ME, Ferrarelli F, Tononi G, et al. Regional patterns of elevated alpha and high-frequency electroencephalographic activity during nonrapid eye movement sleep in chronic insomnia: a pilot study. *Sleep*. 2016;39:801–12.
- Riemann D, Voderholzer U. Primary insomnia: a risk factor to develop depression? *J Affect Disorders*. 2003;76:255–59.
- Riemann D, Spiegelhalter K, Feige B, Voderholzer U, Berger M, Perlis M, et al. The hyperarousal model of insomnia: a review of the concept and its evidence. *Sleep Med Rev*. 2010;14:19–31.
- Roehrs T, Merlotti L, Petrucelli N, Stepanski E, Roth T. Experimental sleep fragmentation. *Sleep*. 1994;17:438–43.
- Rosenberg RS, Van Hout S. The American academy of sleep medicine interscorer reliability program: sleep stage scoring. *J Clin Sleep Med*. 2013;9:81–7.
- Schabus M, Hodlmoser KH, Gruber G, Sauter C, Anderer P, Klösch G, et al. Sleep spindle-related activity in the human EEG and its relation to general cognitive and learning abilities. *Eur J Neurosci*. 2006;23:1738–46.
- Scheuler W, Stinshoff D, Kubicki S. The alpha-sleep pattern. Differentiation from other sleep patterns and effect of hypnotics. *Neuropsychobiology*. 1983;10:183–9.
- Selvitellia MF, Krishnamurthya KB, Herzog AG, Schomer DL, Chang BS. Sleep spindle alterations in patients with malformations of cortical development. *Brain Dev*. 2009;31:163–8.
- Spiegel K, Tasali E, Leproult R, Van Cauter E. Effects of poor and short sleep on glucose metabolism and obesity risk. *Nat Rev Endocr*. 2009;5:253–61.
- Stone KC, Taylor DJ, McCrae CS, Kalsekar A, Lichstein KL. Nonrestorative sleep. *Sleep Med Rev*. 2008;12:275–88.
- Wamsley EJ, Tucker MA, Shinn AK, Ono KE, McKinley SK, Ely AV, et al. Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation? *Biol Psychiatry*. 2012;71:154–61.
- Warby SC, Wendt SL, Welinder P, Munk EG, Carrillo O, Sorensen HB, et al. Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nat Methods*. 2014;11:385–92.
- Wei HG, Riel E, Czeisler CA, Dijk DJ. Attenuated amplitude of circadian and sleep-dependent modulation of electroencephalographic sleep spindle characteristics in elderly human subjects. *Neurosci Lett*. 1999;260:29–32.
- Wendt SL, Christensen JA, Kempfner J, Leonthin HL, Jennum P, Sorensen HB. Validation of a novel automatic sleep spindle detector with high performance during sleep in middle aged subjects. *Conf Proc IEEE Eng Med Biol Soc*. 2012;2012:4250–3.
- Wendt SL, Welinder P, Sorensen HB, Peppard PE, Jennum P, Perona P, et al. Inter-expert and intra-expert reliability in sleep spindle scoring. *Clin Neurophysiol*. 2015;126:1548–56.
- Yotsumoto Y, Sasaki Y, Chan P, Vasios CE, Bonmassar G, Ito N, et al. Location-specific cortical activation changes during sleep after training for perceptual learning. *Curr Biol*. 2009;19:1278–82.
- Younes M. Role of arousals in the pathogenesis of obstructive sleep apnea. *Amer J Respir Crit Care Med*. 2004;169:623–33.
- Younes M, Soiferman M, Thompson W, Giannouli E. Performance of a New Portable Wireless Sleep Monitor. *J Clin Sleep Med*. 2016. Acceptable for Publication with 2016 Oct 20.
- Younes M, Hanly PJ. Immediate post-arousal sleep dynamics: an important determinant of sleep stability in obstructive sleep apnea. *J Appl Physiol*. 2016;120:801–8.
- Younes M, Hanly PJ. Minimizing Interrater Variability in Staging Sleep by Use of Computer-Derived Features. *J Clin Sleep Med*. 2016. [Epub ahead of print].
- Younes M, Ostrowski M, Soiferman M, Younes H, Younes M, Raneri J, Hanly P. Odds ratio product of sleep EEG as a continuous measure of sleep state. *Sleep*. 2015a;28:641–54.
- Younes M, Thompson W, Leslie C, Egan T, Giannouli E. Utility of technologist editing of polysomnography scoring performed by a validated automatic system. *Ann Am Thorac Soc*. 2015b;12:1206–18.
- Younes M, Younes M, Giannouli E. Accuracy of automatic polysomnography scoring using frontal electrodes. *J Clin Sleep Med*. 2015c;12:735–46.
- Younes M, Raneri J, Hanly P. Staging sleep in polysomnograms: analysis of inter-scorer variability. *J Clin Sleep Med*. 2016;12:885–94.
- Zhang X, Dong X, Kantelhardt JW, Li J, Zhao L, Garcia C, et al. Process and outcome for international reliability in sleep scoring. *Sleep Breath*. 2015;19:191–5.