Advanced Analysis of Pharmaco-Sleep Data in Humans

Peter Anderer

Department of Psychiatry and Psychotherapy, Medical University of Vienna, Siesta Group Schlafanalyse GmbH, and Sleep and Respiratory Care, Home Healthcare Solutions, Philips Austria GmbH, Vienna, Austria

Key Words
Sleep macrostructure · Sleep microstructure · Sleep classification · Interrater reliability · Validity

Abstract
Pharmaco-sleep studies in humans aim at the description of the effects of drugs, most frequently substances that act on the central nervous system, by means of quantitative analysis of biosignals recorded in subjects during sleep. Up to 2007, the only standard for the classification of sleep macrostructure that found worldwide acceptance were the rules published in 1968 by Rechtschaffen and Kales. In May 2007, the AASM Manual for the Scoring of Sleep and Associated Events was published by the American Academy of Sleep Medicine, and concerning the classification of sleep stages, these new rules are supposed to replace those developed by Rechtschaffen and Kales. As compared to the rather low interrater reliability of manual sleep scoring, semiautomated approaches may achieve a reliability close to 1 (Cohen's kappa 0.99 for 2 semiautomated scorings as compared to 0.76 for 2 manual scorings) without any decline in validity. Depending on the aim of the pharmaco-sleep study, additional analyses concerning sleep fragmentation, sleep microstructure, sleep depth, sleep processes and local aspects of sleep should be considered. For some of these additional features, rules for visual scoring have been established, while for others automatic analysis is obligatory. Generally, for reasons of cost-effectiveness but also reliability, automatic analysis is preferable to visual analysis. However, the validity of the automatic method applied has to be proven.

Introduction
At the behavioral level, sleep can be defined as a ‘reversible behavioral state of perceptual disengagement from and unresponsiveness to environmental stimuli’ [1]. The sleep-wake cycle and the structure of sleep reflect the spontaneous activity of autoregulatory central nervous system processes. Within sleep, 2 separate states have been defined on the basis of a constellation of physiological parameters. These 2 states, nonrapid eye movement (NREM) and rapid eye movement (REM) sleep, are distinct from one another as each is from wakefulness.

Conventionally, NREM sleep is subdivided into 4 stages – S1, S2, S3 and S4 according to Rechtschaffen and Kales [2] – or 3 stages – N1, N2 and N3 according to the American Academy of Sleep Medicine (AASM) [3]. They reflect light sleep (S1 and N1), middle-deep sleep (S2 and N2) and deep sleep (S3, S4 and N3), respectively. These stages are distinguished from each other principally on the basis of their different patterns of brain electrical ac-
activity, as measured by electroencephalography (EEG). Indeed, EEG is considered the ‘core measurement of poly-
somnography (PSG)’ by Carskadon and Rechtschaffen [4]. The EEG pattern in NREM sleep is synchronous, with characteristic waveforms such as sleep spindles, K complexes and slow-frequency, high-amplitude waves (delta waves). It is characterized by a periodic alternation of so-called up- and downstates (the ‘slow oscillation’, which seems to orchestrate the appearance of the aforementioned waveforms) [5–7].

By contrast, REM sleep is defined by EEG activation, muscle atonia and episodic bursts of rapid eye movements. Generally, REM sleep is not subdivided into substates, although tonic and phasic types of REM sleep are often distinguished for research purposes. This distinction between tonic and phasic REM sleep is based on short, transient events occurring in clusters separated by episodes of relative quiescence [1].

In addition to sleep macrostructure described by the sleep stages, sleep microstructure may be described by EEG arousals or by cyclic alternating patterns (CAPs). According to the conceptual framework of the criteria developed by the American Sleep Disorders Association [8], arousals are a marker of sleep disruption, and thus the number of arousals per hour of sleep represents a measure of sleep fragmentation. Also during normal sleep a certain number of arousals – often associated with changes in body position – is found, but excessive numbers of arousals disturb sleep considerably. Arousals are defined as transient shifts in EEG frequencies for a minimum of 3 s, which may be accompanied by an increase in EMG muscle tone. They may occur during any sleep stage. Arousals in REM sleep require a concurrent increase in submental electromyography (EMG). These arousal definitions by the American Sleep Disorders Association have been included in the AASM manual without any changes [3]. According to an alternative conceptual framework suggested by Terzano et al. [9] and Halász et al. [10], arousals embedded in CAP sequences are viewed as elements weaved into the texture of sleep taking part in the regulation of the sleep process. CAPs are defined as sequences of EEG activity in NREM sleep which repeat themselves cyclically. Each cycle of the sequence consists of phase A (transient EEG events constituting a central nervous system arousal) and phase B (background EEG activity). The scoring rules and recording techniques for CAPs have been published as a consensus report [9].

Diagnosis and differential diagnosis of the vast majority of sleep disorders require recording and analyzing additional biosignals. For this so-called cardiorespiratory PSG, a minimum of 12 physiological signals is needed. Sleep evaluations may comprise additional variables such as electrocardiography, pulse, snoring, airflow, respiratory effort (chest and abdomen), esophageal pressure, oxymetry, body position, (periodic) leg movements, nocturnal penile tumescence, temperature, hormones, electrolytes, etc. These additional signals may be selected on the basis of the disorders to be diagnosed in a particular patient. Further EEG leads are useful for evaluating the spatial distribution of sleep-related EEG patterns. In sleep-related epilepsy or other neurological disorders with associated sleep disturbances, additional EEG leads are indispensable.

Pharmaco-sleep research based on manual and/or computer-assisted evaluation of the sleep EEG has been performed in both healthy volunteers and patients with sleep disorders since the late 1960s [11, 12]. The objectives of these studies are multifarious such as supporting registration of a new drug, providing evidence of efficacy, evaluating central nervous system side effects or providing basic pharmacodynamic data at an early stage in drug development [13–15]. Generally, pharmaco-sleep studies have an important role in characterizing the effect of central nervous system-active drugs since an influence on sleep initiation, continuity and architecture may have a significant impact on the clinical use of the drug [16, 17].

Recording

As published by Rechtschaffen and Kales [2] in 1968, at least one central EEG channel (C3-A2 or C4-A1) is obligatory for sleep scoring in humans. In addition, 2 electro-oculography (EOG) leads are required in order to distinguish between eye movements and prefrontal EEG interference. The EOG electrodes are arranged in a way that eyeball movements result in signals going in opposite directions, whereas head movement and EEG artifacts produce signals going in the same direction. One electromyogram lead placed on the chin is required for detection of REM sleep atonia.

Due to the high temporal resolution of the EEG, microstructures of sleep, such as slow waves, K complexes, sleep spindles, vertex sharp waves or sawtooth waves can be identified and separated from alpha bursts, microarousals, CAP phases or artifacts. To allow these discriminations, the sampling rate of the EEG recording has to be high enough to enable the identification of waveform, amplitude, frequency and duration of the events with ad-
equate accuracy. A time resolution in the range of milliseconds, for instance, is necessary to discriminate a K-alpha complex (indicating an arousal) from a K complex with a superimposed sleep spindle (indicating consolidated sleep). According to the AASM manual, the minimal sampling rate for EEG, EOG and EMG channels is 200 Hz [3].

While the value of a high temporal resolution of the electroencephalogram is unequivocally accepted, that of its spatial resolution is still underestimated. According to the AASM manual [3], whose rules are intended to replace those developed by Rechtschaffen and Kales, only 2 additional EEG channels, one frontal lead (F4-A1) and one occipital lead (O2-A1), are required for scoring sleep. The occipital derivation should be used for identifying alpha activity and the frontal derivation for identifying K complexes and slow-wave activity.

The addition of a parietal EEG electrode makes it possible to evaluate brain topography along the anterior-posterior axis by re-referring the EEG leads to bipolar derivations (e.g. F3-C3, C3-P3, P3-O1). Such fronto-occipital EEG power gradients in the sleep of young healthy subjects demonstrated topographic power shifts at NREM-REM sleep transitions as well as across and within NREM periods [18]. Since the regional EEG power spectra showed state-related and frequency-specific differences, the study highlighted the additional information that may be gained by a – though rather limited – spatial analysis of sleep EEG data. As early as in 1993, a topographic study performed by our group with 18 EEG leads showed an increase in delta power (which was most pronounced frontally) and a decrease in alpha power (which was most pronounced parieto-occipitally) from sleep stages S1 to S4 [19]. A follow-up study demonstrated that not only slow waves, but also sleep spindles were not uniformly distributed across the scalp. Slow sleep spindles were generally distributed over anterior and fast sleep spindles over parietal regions [20]. By means of low-resolution brain electromagnetic tomography developed by Pascual-Marqui et al. [21], we identified cortical spindle sources predominantly medially in the frontal and parietal lobes. Weaker bilateral frontal and parietal sources showed a left-hemispheric predominance [22]. Interestingly, the prefrontal sources (Brodmann areas 9 and 10) oscillated with a frequency below 13 Hz, and the precuneus sources (Brodmann area 7), with a frequency above 13 Hz. These topographic and tomographic studies indicate that the neuronal processes underlying the sleep EEG differ between brain regions, which supports the hypothesis of local aspects of sleep.

Consequently, a comprehensive sleep analysis should include spatial information. Today, a number of sleep recording systems allow the acquisition of data from multiple EEG channels, and thus, limiting factors for increasing the number of electrodes are subjects’ comfort, the time needed for applying the electrodes and the increasing time and effort required for artifact handling. In 2004, Huber et al. [23] published a study using a 256-channel EEG recording system with an electrode cap for revealing local changes in slow-wave activity in a night after a learning task of rotation adaptation. The high-density EEG analysis revealed a remarkably stable slow-wave activity that was most pronounced over frontal regions, confirming our early topographic delta distribution obtained with 18 electrodes. Indeed, caps and electrodes with improved comfort and systems combining PSG and high-density EEG devices are now available and will hopefully stimulate new studies on the spatial analysis of sleep EEG.

Recommendations, specifically intended for pharmaco-sleep studies in man, for subject selection, environmental conditions and recording settings including adaptation nights as well as for digital recordings, calibration and biosignals, can be found in the ‘Guidelines for the recording and evaluation of pharmaco-sleep studies in man’ recently published by the International PharmacoeEG Society [24].

Analysis

In any pharmaco-sleep study, the first step of the analysis has to be the classification of sleep stages. Up to 2007, the only standard for the classification of sleep macrostructure that found worldwide acceptance were the rules published by Rechtschaffen and Kales [2] entitled A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects in 1968. In May 2007, the AASM Manual for the Scoring of Sleep and Associated Events was published by the AASM [3]. Concerning the visual classification of sleep stages, these new rules are intended to replace those developed by Rechtschaffen and Kales. However, there are still ongoing discussions on the pros and cons of the visual AASM rules, and thus, no final recommendation for using the Rechtschaffen and Kales or the AASM rules can be given at present. For further commentaries on the AASM manual for scoring sleep, see those by Schulz [25] and Parrino et al. [26]. The effects of the scoring rules on interrater reliability were evaluated by Danker-Hopfe et al. [27]. The authors summarize that on the one hand the integration of frontal,
central and occipital leads improves interrater reliability. On the other hand, however, this advantage is counteracted by the impairment of interrater reliability due to the new rule according to which any cortical arousal, whether or not it is associated with an EMG increase, determines the end of stage N2. Cohen’s kappa was 0.76 for visual scoring according to the AASM standard as compared with 0.68 for visual scoring according to Rechtschaffen and Kales. To allow a direct comparison, both kappa values were based on 5 classes, i.e. for Rechtschaffen and Kales scorings stage S3 and S4 were combined with a slow-wave sleep stage, and epochs scored as movement time were ignored. In a second paper, Moser et al. [28] investigated the effects of the scoring standard on sleep parameters in the same data set. In summary, the new standard showed only a minor influence on total sleep time, sleep efficiency and stage of REM sleep, but affected wake after sleep onset as well as the distribution of NREM sleep stages (stage 1 and slow-wave sleep increased by approx. 10 min each, stage 2 sleep decreased by approx. 20 min). In the process of adapting the automated sleep staging system Somnolyzer 24 × 7 – which in 2005 had been developed and validated for scoring sleep according to Rechtschaffen and Kales – to the new AASM rules, we performed a stage transition analysis to evaluate step by step the effects the change of the rules had on sleep staging [29, 30]. On average, the inclusion of the occipital lead for detecting alpha activity affected only 2–3 epochs per recording. In these rare cases, epochs changed from stage S1 to stage W, since the occipital alpha activity was not detectable at central leads. In the vast majority of cases, however, alpha activity was detectable not only at occipital, but also at central leads and thus the inclusion of the occipital lead had only a minor effect on sleep stage scoring. More epochs were affected by the addition of the frontal lead, as epochs with slow waves (just) below 75 μV at central leads may change from S2 to N3 if their amplitudes are above 75 μV at frontal leads. As a consequence, an average of 19 epochs per recording changed from S2 to N3. The exact influence of the change in the scoring rules depends on individual EEG characteristics, such as the anterior-posterior gradient of slow-wave amplitudes or densities of sleep spindles and K complexes on the one hand, and arousal density and alpha topography on the other, and may range from nearly no changes at all to increases in slow-wave sleep of up to 53 min and in stage 1 sleep of up to 44 min, with a maximal decrease in stage 2 sleep of up to 55 min [28–30]. In any case, scoring results obtained with one method cannot directly be compared with those obtained by the other.

Since the first publication of the Rechtschaffen and Kales scoring standards in 1968, numerous attempts at a computer-assisted identification of sleep stages have been published. According to Schulz [25], sleep experts should concentrate on 3 major tasks related to automated sleep scoring: (1) definition and development of an algorithm, (2) surveillance of the analyzing process, including artifact decontamination, and (3) quality control of sleep analysis. In the paper on digital analysis and technical specifications, which accompanied the AASM manual of 2007, Penzel et al. [31] presented a comprehensive review of papers on computer application and validation of sleep staging. The authors identified 119 papers suitable for evidence review and defined a grading for performance evaluation (from level 1 with a sample size >64 without nesting within subject; sequential or representative sample; event-by-event or epoch-by-epoch comparison to level 5 for case series) with 8 additional evaluation factors for grading within level: (1) normal controls and clinically relevant group used in sample; (2) no recording selection for quality and/or discarded <5% of records; (3) clinical standard used for group classification; (4) standard used for sleep scoring and recording; (5) blind, independent scoring; (6) multiple human scorers used to set comparison standard; (7) entire recordings used, and (8) description adequate for replication. The paper on the automated sleep classification system Somnolyzer 24 × 7 was assigned grading level 1 and listed on top of 44 papers on general computer applications for sleep staging [31]. Indeed, the paper published by Anderer et al. [29] was the only study that fulfilled all criteria for level 1 and all 8 additional evaluation factors for grading within level 1. By then, also the study published in 2010 reporting the validation of the AASM version of the Somnolyzer system met all these criteria [30]. Since the automated classification by Somnolyzer typically requires only few corrections during visual editing, interrater reliability of 2 Somnolyzer-assisted sleep scorings is close to 1. To perform the standardized quality control procedure most efficiently, sleep experts receive information from various steps of the analysis including raw data, artifacts and sleep/wake-related features (see fig. 1 for an example). While on average between 1 and 5% of the epochs are changed during expert review, in single cases up to 20% of the epochs might be affected, and thus, at present, unsupervised automatic analysis is not recommended. Cohen’s kappa for interrater reliability between 2 Somnolyzer-assisted classifications was 0.99 for Rechtschaffen and Kales scorings based on 286 PSGs [29] and again 0.99 for AASM scorings based on 72
In contrast, interrater reliability for 2 manual scorings resulted in Cohen’s kappa of 0.68 for the 286 Rechtschaffen and Kales studies and 0.76 for the 72 AASM studies [29, 30]. Cohen’s kappa for the agreement between the 2 Somnolyzer-assisted and the 2 visual scorings for the 72 AASM studies was between 0.75 and 0.76 and thus was comparable to the interrater reliability between 2 manual scorings. Thus, the validity of Somnolyzer-assisted scoring is comparable to that of human experts, while the reliability is close to 1. Since the time required for expert review of the automated Somnolyzer scoring is at least 5 times shorter than for manual scoring, Somnolyzer-assisted scoring reduces interrater variability as well as scoring time to a minimum. In pharmaco-sleep studies, typically comparing drug-induced versus placebo-induced changes, the reliability of PSG scoring is essential.
First applications of Somnolyzer in a double-blind, placebo-controlled pharmacosleep study on the potential interaction of 20 mg paroxetine and 1 mg alprazolam in healthy young volunteers and in a single-blind, placebo-controlled clinical sleep study on the acute effects of 100 mg trazodone in patients with nonorganic insomnia related to somatoform pain disorder confirmed the validity of the classifier for scoring PSGs after treatment with psychotropic drugs [32, 33]. This validity was further confirmed by Svetnik et al. [34] in 164 PSGs of 82 subjects in a clinical trial using zolpidem in a phase advance model of transient insomnia. Recently, Somnolyzer-assisted scoring has been successfully used in a multicenter, double-blind, randomized, placebo-controlled, 2-way crossover study in 161 primary insomnia patients with the dual orexin receptor antagonist almorexant [35].

Irrespective of the method used for analyzing sleep EEG data, the significance of the results critically depends on the treatment of artifacts. It is obvious that analysis of data contaminated by artifacts can lead to spurious results. In sleep EEG data, simple artifact elimination methods based on the definition of a threshold for maximal EEG amplitudes, as frequently used in evoked potential studies for instance, are certainly not appropriate. As part of the EU-funded project SIESTA, we reviewed in detail types and treatment of artifacts in the sleep EEG [36]. A reliable and valid artifact processing strategy should include: (1) high-quality recording techniques in order to minimize the occurrence of avoidable artifacts (e.g. technical artifacts); (2) artifact minimization procedures in order to minimize the loss of data by estimating the interference of different artifacts in the EEG recordings, thus
allowing the calculation of the ‘corrected’ EEG (e.g. ocular and electrocardiographic artifacts), and finally (3) artifact identification procedures in order to define and eliminate epochs contaminated by remaining artifacts (e.g. movement and muscle artifacts). Indeed, artifacts can mimic almost any kind of EEG pattern [37, 38] and artifacts included in automatic analysis can seriously affect the results. For instance, Brunner et al. [39] demonstrated that the rejection of short-lasting muscle bursts significantly reduced power spectral density in all frequencies from 0.25 to 32 Hz, most prominently of course in the faster frequency bands. Thus, the careful handling of artifacts is of the utmost importance for EEG data processing.

For a review on automated detection and analysis of EEG waveforms such as arousals, CAPs, sleep spindles, K complexes, slow waves, as well as spike and wave activity or specific pattern related to hypoglycemia, see Penzel et al. [31]. In our paper in the textbook for the training course of the International Pharmaco-EEG Society in 2006, we presented methods and validation data for the automatic detection of a variety of patterns such as the artifact detector, the spindle detector, the K complex detector, the detector for rapid and slow eye movements (REMs and SEMs) as well as the arousal detector [40]. In addition to the hypnogram according to the AASM standard, figure 2 displays arousals and CAP A phases for the entire night (8 h 1 min 30 s). In the lower window, raw data (left and right EOG, chin EMG, F4-A1, C4-A1, O2-A1 EEG) as well as the automatically detected arousals and SEM and REM events are displayed for a 1-min window. Note that in REM sleep, arousals have to be associated with an increase in chin EMG activity. The Somnolyzer defines REMs and SEMs with a temporal resolution of 2 s. Note that REMs might be superimposed on SEMs. Thus, as can be seen in the upper part, SEMs do not only occur at sleep onset or after awakenings, but are also characteristic features of REM sleep (compare also Pizza et al. [46]).

Fig. 3. REMs and SEMs in a period of phasic REM sleep (22-year-old male subject). In the upper window, in addition to the arousals and the hypnogram according to the AASM manual [45], automatically detected SEM and REM events are displayed for the entire night (8 h 1 min 30 s). In the lower window, raw data (left and right EOG, chin EMG, F4-A1, C4-A1, O2-A1 EEG) as well as the automatically detected arousals and SEM and REM events are displayed for a 1-min window. Note that in REM sleep, arousals have to be associated with an increase in chin EMG activity. The Somnolyzer defines REMs and SEMs with a temporal resolution of 2 s. Note that REMs might be superimposed on SEMs. Thus, as can be seen in the upper part, SEMs do not only occur at sleep onset or after awakenings, but are also characteristic features of REM sleep (compare also Pizza et al. [46]).
example the description of sleep with a better time resolution than 30 s using continuous sleep analyzers based on complexity measures, autoregressive models, hidden Markov models or gaussian mixture models [40]. However, the sleep electroencephalogram is most frequently quantified by means of spectral analysis, which decomposes it into its constituent frequency components (see Achermann [41] for theory and application of spectral analysis on sleep EEG). Typically, fast Fourier transform is applied to short EEG segments (usually between 2 and 10 s), and the power spectral density spectrum (scaled in μV²/Hz) is averaged over artifact-free segments for a given time (e.g. for a 30-second epoch) or condition (e.g. for NREM or REM periods). Based on the averaged spectra, target variables such as absolute power (in μV²) and relative power (in % of total power) may be derived for various frequency bands and displayed as a function of time for an individual EEG channel (see fig. 4 for the distribution of delta power in the course of the night). In recent years, studies on REM sleep behavior disorder showed increased evidence that the loss of REM sleep atonia may be an early manifestation of degenerative neurological disease [42, 43]. A simple measure to automatically quantify the time in REM sleep without atonia – the atonia index – was developed and validated by Ferri et al. [44] in 2010.

**Conclusion**

Technical and digital specifications, including sensor types, maximal electrode impedance, minimum digital resolution, sampling rates and filter settings for routine PSG recordings were summarized in the 2007 AASM...
In pharmaco-sleep studies, sleep classification should be performed at specialized centers by experienced, trained and certified scorers. While unsupervised automatic scoring cannot be recommended for sleep staging at present, automatic sleep analysis has proven its usefulness in a semiautomated scoring approach or for obtaining a second independent scoring. In the first case, visual editing of the automatic scoring has to be performed by experienced and trained scorers, and in the second case, comparisons between visual and automatic scoring can be used to identify problematic recordings that should be considered for consensus scoring [30]. Obviously, an automatic method that requires only few corrections during visual editing should be used so that the influence of interindividual differences in sleep scoring can be reduced to a minimum. Specifically in pharmaco-sleep studies, both the validity and the reliability of sleep scoring are essential.

The main hypotheses in pharmaco-sleep studies are usually based on measures of sleep derived from Rechtschaffen and Kales or the AASM (e.g. sleep efficiency, wake after sleep onset, percentage of slow-wave sleep, etc.). Nevertheless, additional analyses concerning sleep fragmentation (e.g. arousals, CAPs, etc.), sleep microstructure (e.g. spindles, phasic REM episodes, etc.), sleep depth (e.g. delta plots via fast Fourier transform analysis, ‘microcontinuity’ of slow waves, probability of slow-wave sleep, etc.), sleep processes (e.g. continuous physiological and probabilistic models of sleep, etc.), REM sleep atonia and local aspects of sleep (e.g. topography, source localization by means of low-resolution brain electromagnetic tomography, etc.) should be considered. For some of these additional features, rules for visual scoring have been established, while for others automatic analysis is obligatory. Generally, for reasons of cost-effectiveness but also reliability, automatic analysis is preferable to visual analysis. However, the validity of the automatic method applied has to be proven. Last but not least, it is important to note that careful handling of artifacts is a prerequisite for all automatic methods [36].

References


