The guidelines of Rechtschaffen and Kales (R&K) were meant as a reference method. However, it became, unintentionally, a gold standard. The rules have never been appropriately validated. It is used for the scoring of recordings of, for instance, pathological sleep, for which it was not designed. Today R&K is an insufficient description of sleep processes. In practice, the rules are often difficult or impossible to follow and deviations are common, although not reported.

The major drawbacks are: low temporal resolution, ignorance of spatial information, insufficient number of stages, low correspondence between electrophysiological activity and stages, and ignorance of other physiological parameters such as autonomous nervous system activity and body motility.

The original idea of the committee should be followed: the guidelines should be revised with the accumulation of new knowledge and development of technology. This is especially true for the analysis of delta waves, which should replace visual scoring of the NREM sleep stages.

Many of the problems related to the use of the manual are not due to the system itself but the way it has been applied. This has evidently had a serious impact on the way of thinking and the development of the field in general.

Key words: sleep stage scoring, Rechtschaffen and Kales, sleep analysis, visual scoring.

The scoring system in the manual edited by Rechtschaffen and Kales [1] has been, since its publication, the major set of guidelines for sleep analysis despite its drawbacks and criticisms presented. The guidelines can be viewed from different points of view. They can be regarded as a reference method, a gold standard, or a model of the sleep process(es). Originally it was meant to be a reference method. However, in practice it became both the major method of sleep analysis and the gold standard to which all other methods have to be compared.

Many, or even most, of the ideas in this review are not new. Similar thoughts have been presented previously [2–5]. In this paper reference to computerized electroencephalography (EEG) analysis is often made. More detailed description of the methods is beyond the scope of this paper. The reader is asked to turn to recent reviews in this field for detailed descriptions [4–6].
Background

Before the invention of EEG by Hans Berger in 1924 sleep was studied by other means [7]. Variations in waking threshold, motility, respiration and other parameters throughout the night were observed. They could not be explained at that time because the differences between the psychophysiological stages were not known, as sleep was thought to be a uniform state.

Sleep stages based on the EEG were first introduced by Loomis et al. [8]. The letters A–E were assigned to the stages. Stage A corresponded to early drowsiness and E to sleep with high-amplitude delta waves. Stage REM (SREM) was not recognized until the discovery of rapid eye movements (REMs) by Aserinsky and Kleitman [9]. Finally, Dement and Kleitman [10] introduced the cyclic patterns of sleep stages based on their large normative study. The four non-REM (NREM) stages and SREM formed the basis of subsequent polygraphic sleep studies.

The need of a common platform for exchanging data and results was emphasized by the study of Monroe [11] in which he showed that the inter-rater agreement between different laboratories was low. This led to the establishment of the committee led by Rechtschaffen and Kales. The rules of the manual of Rechtschaffen and Kales (R&K) are more or less a formalization of the Dement and Kleitman criteria.

The main goals of the Committee were to standardize recording and scoring techniques in order to increase the comparability of results between laboratories. The manual provided the minimum requirements for comparison of polygraphic sleep studies of adult humans. The Committee also encouraged the use of other concepts and revisions of the manual were suggested, with the acknowledgement of new information. However, instead of providing the necessary reference to novel developments the rules of the manual became in practice the only method of sleep analysis. As a consequence it became a gold standard and, unintentionally, a restriction to the development of subsequent sleep research.

Recording procedures

The R&K system was designed for paper recordings including specifications for filters, gains, paper speed, pen deflection, number of channels, etc. Some of the requirements and guidelines have become obsolete with the use of modern digital equipment. With sufficient dynamic range and sampling rates all these can be set in a non-restrictive way. In analysis more limited settings can be used.

One profound weakness of the standard rules is that only the EEG, and to some extent eye movements (EOG) and muscle activity, are recorded. In the two-channel electrooculography (EOG) eye movements are seen as out-of-phase deflections, whereas the EEG signals produce in-phase deflections. One weakness of the R&K is that blinks are seen only by the upper channel. An improved derivation giving more symmetrical deflections for blinks has been developed [12].

From the EEG information only one derivation is taken into account because the Committee did not find regional differences between the scalp areas critical for the scoring of sleep stages. A single central EEG derivation for recordings was specified. Either the C4/A1 or C3/A2 may be used, since the EEG patterns from homologous areas were thought to be generally synchronous. No studies where this would have been confirmed were conducted. Using only one EEG derivation means that only one
part of the brain surface is evaluated. Most of the information obtained even from this part is ignored because it is not needed for sleep staging. In addition, it is not always known whether the recorded activity comes from the scalp or the reference electrode. If EEG activity from both ears or mastoids is recorded on one channel, for instance, sleep spindles are seen (personal observation). It is also probable that much of the alpha activity during drowsiness and sleep is picked up by the reference rather than the scalp electrodes [13,14].

Scoring by epochs

R&K sleep stage scoring is performed in epochs of equal lengths so that the whole epoch is assigned to one stage. If signs of two or more stages are present the epoch is assigned to the stage of the longest duration. With paper recordings the one-page epoch minimizes the time used for scoring and later calculations. The chosen epoch length of 20 or 30 sec is clearly a compromise between accuracy and laboriousness. Dement and Kleitman [10] used 10 sec epochs in borderline cases to differentiate NREM stages. Different results are obtained with different epoch lengths [3]. With 20 sec epochs 11 sec of a certain stage is sufficient, while with 30 sec epochs 16 sec is required. Short stage changes are ignored and the microstructure of sleep cannot be taken into account. In many cases it can also be very difficult to tell which stage is dominating the epoch. The problem is less with good sleepers who have fewer stage changes and electrophysiological characteristics that fit the scoring criteria well. In disturbed sleep with a great number of stage changes the scoring results do not reflect reality.

In sleep studies other physiological parameters, arousal threshold, reactivity and performance are correlated to epoch scored sleep stages. With long epoch lengths some epochs include two, sometimes even three, electrophysiologically different states. This increases variability of correlations because it is not always known what was the actual physiological state of the subject when, for instance, the stimulus was given. Also, automatic analysis is often made epoch by epoch which is misleading because of the heterogenic content of the epochs. The presumption with most methods is that the signals are stationary within the segment analysed [5].

Characteristics of R&K stages

The division into R&K stages is generally based on the physiological knowledge of the sleep processes at that time. The principle of the Committee was to retain, as much as possible, the terminology and criteria which had had the greatest use. Sleep had already been defined polygraphically by stages 1, 2, 3, 4 and REM. The stages were included in the manual, although the physiological correlates of sleep stages determined by the EEG were not completely known.

No objective studies examining whether these stages are the true representatives of the morphology and the physiological processes behind the events visible on the traces were conducted. According to Lairy: "...it seems clear that ... sleep stages represent nothing but one means of data reduction"[2]. On the other hand the purpose was not to limit the use of other descriptions when needed. Revision of the manual was also suggested with the accumulation of new knowledge. Criticism pointing out that the
number of stages is too limited appeared [2]. However, the rules have never been revised.

Wakefulness

Morphologically there are at least two types of wakefulness stages. Subjects with high alpha background show occipital alpha rhythm with closed eyes and low-amplitude mixed-frequency EEG with fast eye movements and blinks with open eyes. Poor alpha producers do not necessarily have any differences in their EEG whether eyes are closed or open.

It is not clear if all alpha periods during sleep are signs of wakefulness, or even arousal [3]. There are at least two different types of alpha activities which have a different frequency and topography [3,13]. One is the wakefulness-related occipital alpha rhythm of 8–12 Hz. The other is an approximately 2 Hz slower centro–frontal rhythm that is present during drowsiness and sleep. Slow eye movements, which are, according to the manual, a characteristic of stage 1 (S1) can already be present with alpha activity. Loomis et al. [8] classified this as stage A. The pattern is a sign of drowsiness [15] and it would be justified to classify it as a separate stage.

Even if wakefulness constitutes a negligible part of the night in good sleepers, and thus one stage category is justified, this might not be the case in subjects with disturbed sleep [16]. Even if the manual gives room for, or even demands, additional definitions of stages, the restrictive application of the rules has had a negative impact on the development of the field.

Stage 1

S1 is defined by a low voltage, mixed frequency EEG with a prominence of activity in the 2–7 Hz range [1]. In healthy adults S1 constitutes a small proportion of whole-night sleep [7]. If sleep is uninterrupted the only occurrence of S1 can be at sleep onset. It is therefore mostly regarded as a transition phase.

S1 is also scored if EEG is of a low amplitude and sleep spindles and REMs are lacking for more than 3 min during otherwise sound sleep. Even according to the manual this 3 min rule is arbitrary. It was based upon the judgement of the committee that “inter-spindle intervals of that length might occur without stage change although such occasions would be rare” [1]. In some subjects this stage takes up a considerably large proportion of the night. According to some authors the period is too long and therefore considerable parts of S1 would be overlooked [4]. On the other hand, there is no evidence that even a longer period without significant landmarks of any stage would be physiologically equal to S1 of falling asleep [15].

In disturbed sleep S1 can take up a very large proportion of the night. Thus, its more accurate definition and division into subcategories might become meaningful. This has so far not been performed.
Stage 2

Onset of S2 is defined by the first appearance of a 13–14 Hz sleep spindle or K-complex (KC) on a low voltage background EEG activity. "Low-voltage activity at 12–14 cps may begin to appear as the transition to S2 approaches, but this activity is not to be defined as sleep spindles until the rhythmic bursts are clearly visible for at least 0.5 sec..." [1]. Well-formed spindles have to last 0.5 sec before S2 can be scored. By R&K rules KCs are defined as EEG wave forms having a well delineated negative sharp wave, which is immediately followed by a positive component. The total duration of the complex should exceed 0.5 sec. Waves of 12–14 Hz may or may not constitute a part of the complex.

By definition, spindles and KCs should appear on a background of low-voltage mixed frequency EEG. However, in practice stage 2 is heterogeneous and could be divided into substages [2]. Episodes with delta activity not exceeding 75 μV in amplitude are often scored as S2 even if spindles and definable KCs are lacking. If the rules were strictly obeyed these epochs should be scored as S1. If such a period occurred between two S4 epochs an artificial stage change would occur [3].

In EEG practice the duration criterion for KCs is not routinely applied. Amzica and Steriade found that the duration of the KCs varies between 250 msec and 1050 msec [18]. Some definitions require an associated sleep spindle while others state that a spindle may, or may not, be part of the complex [1,19]. In summary, the definition of the KC by morphology is variable and unclear. The start endpoints of KCs are not clearly stated in any of the definitions, which causes additional confusion.

Spindles are sometimes visible only parietally or frontally [3]. In addition to localization spindles can also vary somewhat in frequency [3,13,20]. Occasionally it seems that the amplitudes of the spindles are variable. However, with topographical studies it can be shown that in many cases it is the localization of the amplitude maximum that varies [13]. Considerable interindividual and low internight variability in spindle density and localization is seen [20,21]. Already Roth et al. [22] have noted that the maximum of every component of the KC is not necessarily at the vertex. According to various studies, approximately one third of the KCs are located centrally, one third can be seen only frontally and one third in both locations [23,24].

Although the events are often also visible more laterally, some of them can be reliably recognized only at the midline [13,20]. Especially at sleep onset, the amplitude and duration of both the KCs and spindle activity increase gradually. It is possible that in some cases sleep onset could be shortened by using midline electrodes which should be systematically studied.

Studies on inter-rater agreements of the visual detection of spindles and KCs are few. The inter-rater agreement in sleep spindle detection is between 80 and 90% [25]. No clear statistics exists on inter-rater agreements for poorly defined spindles. Bremer et al. obtained 50% agreement in KC detection between two human judges [26].

In summary, the definition of onset of S2 is highly dependent on the decision of what is considered a sufficiently recognizable phasic event. Also its continuity is defined by the presence of the short-lasting, variable phasic events and the absence of other (SREM related) phasic events. All events are not even visible by the EEG derivation used. No difference is made between cases where there is only one occasional, poorly defined spindle or KC within each 3-min interval and the other extreme when well-formed phasic events appear at regular intervals of a few seconds. Thus scoring of an epoch can be highly dependent on the preceding and following epochs [2].
**Stages 3 and 4**

According to R&K the basic characteristic of NREM sleep S3 and S4 is the increasing amount of delta activity in the frequency range of 0.5–2 Hz. The upper limit of 2 Hz was chosen even though 3 Hz had commonly been used [7,27] and common EEG practice usually defines delta as being waves of less than 4 Hz [19]. The committee did not present any arguments for this narrow band.

The NREM stages S2, S3 and S4 are separated by the percentages of time (<20%, 20–50% and >50%, respectively) occupied by delta waves exceeding the amplitude of 75 µV within the epochs. No clear scientific basis for taking these values and not others can be found. The need of an amplitude criterion was discussed by the Committee but it was maintained because most studies until then had used some criterion. The amplitude criterion used by Dement and Kleitman with different derivations was 100 µV [10]. Alternative measures of slow activity were, however, encouraged.

The 75 µV threshold might work satisfactorily with young people who have a delta amplitude well over 100 µV, leaving a sufficient margin for inter-individual variations. The slow wave amplitude declines with increasing age [7,28,29]. This does not necessarily reflect impairment of sleep quality. It is therefore doubtful whether similar amplitude criteria can be utilized for all age groups. Webb [30] has suggested that no amplitude criterion should be used for elderly people.

A general finding is that the amplitude of slow activity can be different in C3/A2 and C4/A1. In some instances slow activity is of too low an amplitude in the central area but is well represented frontally [3]. Thus the differentiation between NREM stages S2, S3 and S4 is clearly dependent on which EEG derivation is chosen for scoring.

We would like to argue whether the whole concept of estimating the slow wave process is wrong. Counting waves above a certain threshold does not give any information about the amplitude distribution above or below the threshold. Martin et al. have already noted that quantification of slow activity gives more realistic results than sleep stage scoring [31].

**Inter-rater agreements**

Studies on inter-scorer agreements between laboratories are few and the results are poor [11,32]. Even within laboratories inter-scorer agreements obtained range from 88% to 94% [5,31,33]. Usually high, i.e. >90%, agreements, are obtained for stable S2 and SREM, whereas the agreements for S1 and S3 are lower. In general, the more transitions, the lower the agreements. Agreements between 52% and 94% have been obtained for visual scoring of S3, and between 44% and 94% in scoring of S4, depending on subject groups and whether the comparisons were made within or between laboratories [31–33]. It is clear that the results are highly dependent on the age and quality of sleep. The more fluctuations there are and the closer the amplitude of the delta waves is to the 75 µV limit, the poorer the results.

It is clear that in situations where rules of R&K are difficult or impossible to obey, different smoothing and other methods are utilized [3]. However, in contrast to the recommendations of the committee these are not reported.
NREM-process—new concepts

According to recent experimental studies the principal process behind NREM sleep seems to be a cortically generated slow oscillation of <1 Hz [34,35, for latest review see 36]. This oscillation is synchronized widely over the cortical surface by intracortical linkages [37]. The slow oscillation gives rise to the KC activity which is its most visible outcome [18].

At sleep onset, the synchronization of the cortical network is rather low. At the onset of slow oscillation, vertex waves can be seen in the EEG. With further synchronization and spreading of the slow oscillation over the cortical surface, the vertex waves become more visible and their amplitude increases until they get the form of the KCs [38]. With deepening of sleep, KCs become more rhythmic [18]. Thus spontaneous KCs can be regarded as an oscillatory phenomenon in contrast to the evoked KCs [38].

The KC activity is transferred from the cortex to the thalamus, where it drives thalamic reticular and thalamocortical cells to produce spindles [38]. The spindles are generated in the thalamus and transferred to the cortex through the thalamocortical network [39]. Spindles are considered sleep maintaining events [40] which block the transfer of sensory information into the cortex allowing the evolution of sleep into deeper stages [see discussion after 41,42]. During light sleep every cycle of the slow oscillation generally leads to a sequence of spindle waves [36].

With increasing hyperpolarization of the thalamocortical cells, spindles are gradually replaced by intrinsically generated delta activity [41] which is also synchronized by the KCs [38]. The slow oscillation groups cortically generated slow waves as well. The result is the complex waveforms seen in the EEG during NREM sleep [36]. That the slow oscillation and delta activity are also two distinct entities in humans has recently been demonstrated by Achermann and Borbély [35].

Delta waves and delta oscillation are different phenomena. The former are slow waves with a duration between 0.25 and 2 sec, the latter are phenomena that appear at this interval. Both have a contribution to the spectrum of the EEG delta band [18, 36,38].

Impact of new concepts on NREM sleep modelling

Spindles

If sleep spindles have a sleep protective function then it is possible that S2, with a small number of spindles, has insufficient sleep protection. However, it has to be taken into account that there is inter-individual genetically determined variability of the spindle and KC production that might have nothing to do with sleep quality [4,20,21].

The R&K rules do not include measures of spindle intensity. Naitoh et al. have suggested that NREM sleep should be defined as spindle-dominant and delta-dominant sleep rather than arbitrary sleep stages [40]. It has been suggested that the number of sleep spindles per minute during sleep S2 should be counted [4]. However, recently attention has been paid to the appearance of spindles at regular intervals of 3–5 sec [44]. This is close to the lowest part of the frequency band of the slow oscillation and this periodicity should be taken into account when defining spindle activity [45].

The spindles appear in blocks of up to 10 [4]. The question is whether these spindle
blocks should be scored separately. It has been found that good spindle producers have more spindle blocks than poor spindle producers. However, the inter-spindle interval remains unchanged. Benzodiazepines also increase the number of the spindle blocks without affecting the interval [see 4]. It would be of interest to see whether the proportions of “spindling blocks” versus “non-spindling blocks” are altered in disturbed sleep, and it is not known whether spindle rhythmicity is altered in disturbed sleep. It would also be of interest to study whether sleep disturbances are reflected in an impairment of spindle activity in a specific localization. It is possible that spindle intensity may have a different importance during slow wave sleep than during spindle sleep on a low-voltage background activity [4].

K complexes

In the early days of sleep research the separation of the vertex waves and KCs was justified. Today the definitions of these waves has to be re-evaluated. It seems that both are expressions of the same cellular mechanisms. Thus the old morphological definitions could be abandoned. This would evidently also have an effect on the scoring of S1 versus S2.

There has been controversy about the meaning of the KCs. On one hand KCs have been regarded as non-specific evoked potentials either to external or internal stimuli. They could either indicate transient changes toward arousal or they could be sleep protective signs of dampening of the response-related arousal [for overviews see 22, 46,47].

On the other hand, recent findings indicate that spontaneous KCs may be part of the NREM process, being the most visible outcome of the slow oscillation. Their increase would lead to the deepening of NREM sleep [for overview see 48]. The spontaneous, periodic KCs may thus be regarded as favourable for sleep. All spontaneous KCs do not have to be rhythmic since, at least in the initial stages of sleep, the cortical network undergoes competing influences that would result in asynchronous and arrhythmic or isolated KCs[18]. In sound sleep single, randomly occurring KCs in response to stimuli would thus be a minority. The differentiation of single, evoked and aperiodically occurring KCs from oscillatory ones might be important for the assessment of sleep quality. The former would indicate sleep disturbances, the latter undisturbed sleep.

Slow waves

It has been claimed that slow wave sleep is related to cerebral recovery and bodily restitution. Slow wave activity can be regarded as a measure of NREM sleep intensity [49]. According to present understanding, sleep is regulated by at least a homeostatic and circadian component. It has been postulated that the homeostatic component is directly reflected by the quantity of slow wave activity [50]. Amplitude (and power) of the delta activity is higher at the beginning of the night, reflecting the pressure for sleep, and declines towards the end of sleep [51].

It is not clear what the essential characteristics of slow activity of good and poor sleep are. Is it the amplitude or the continuity? We could argue that low amplitude but regular slow activity does not necessarily mean poor sleep, whereas substantial
irregularities, even with abundant high-amplitude waves, could be signs of underlying disturbances. These matters are highly speculative but possible in the light of recent literature. They should be studied both in subjects with subjectively normal and disturbed sleep. However, such characteristics of sleep microstructure are poorly reflected by the rules of R&K and this is why new approaches and methods are badly needed.

It is evident that a large number of waves that are considered KCs by the group of Steriade are generally included as delta activity by sleep scorers and thus contribute to the scoring of the S3 and S4. It is not clear whether KCs and delta waves can be visually separated or whether this is even needed.

One has to remember that the concept of slow oscillation and its expression is based on the work of a few research groups. The ideas should therefore be considered with caution. Possible methodological pitfalls have to be examined and the results have to be repeatable in humans by methods other than spectral analysis. However, the description here is, in our opinion, justified in order to demonstrate what kind of matters focus should be turned to in sleep studies. The most important thing is to renew the concepts from stereotyped classifications into accurate description of the true biological processes. We do not know whether studies focusing on slow oscillation and its outcomes would be beneficial. Still, we suggest that future studies should, instead of only counting waves and phasic events, focus on regularities and irregularities of spindles and KCs and other slow wave patterns. Examining the slope of the decline of delta activity in various conditions could provide useful information as well. The validity of all new concepts should be examined both by the differences between normal and disturbed sleep, and different age groups.

REM sleep

During SREM the background EEG is characterized by low-voltage mixed-frequency activity similar to that of early S1, except that vertex waves are not as prominent as in S1. Characteristic sawtooth waves are frequently, though not inevitably, recorded [1]. These are 2–6 Hz, sharply contoured triangular waves that usually occur serially for several seconds and are highest in amplitude over the Cz and Fz electrodes [13].

Scoring of SREM also requires the presence of REMs and the absence of sleep spindles and KCs. The muscle tonus has to be low in comparison to other stages although sudden, occasional twitches occur.

Alpha frequencies are often present in SREM and may be more persistent than in S1. The alpha frequencies are usually 1–2 Hz less than the subjects waking rhythm [1]. Kubicki et al. have pointed out that all alpha activity does not necessarily mean arousal but is probably a physiological part of sleep [3]. This is also reflected in the definitions of arousals by the guidelines of the ASDA Task Force [52] where elevation of muscle tone in SREM is required in conjunction with alpha activity. As there are no guidelines for scoring epochs with alpha activity it is expected that the outcomes are variable. Some may use the 50% rule and score epochs with more than half of this activity as wakefulness, while others might regard them as REM sleep if the muscle tonus remains low. Also the physiological significance of sleep related alpha activity has remained obscure since a lack of rules has not encouraged the research.

As with S2, the problem of scoring SREM is that it is so heavily based on the presence of phasic events rather than continuous parameters. The standard manual
includes a vast number of complicated rules for scoring. Despite the many examples, every single case cannot be scored according to the manual. The onset and the end of the stage are sometimes especially difficult to score when the landmarks of the stages do not appear, and disappear simultaneously.

Often the scoring of several sequential epochs is dependent on the change of a single parameter. Long periods with relatively low-voltage EEG, especially at the end of the night, could either be scored SREM, S2 or even S1, depending on whether there is an intervening REM, spindle or KC. Striking examples where a period between S2 and SREM without phasic events would be scored as either S1, S2, SREM or remain unclassified purely depending on the muscle tonus have been demonstrated by Kubicki et al. [3]. Unlike scoring of S2 the 3-min rule is not used. It is possible that long periods of sleep that are scored SREM are in fact NREM sleep or some intermediate stage that has not yet been defined.

Dement and Kleitman [10] have noted that there are periods with both REM and NREM related phasic events. They called the pattern “slipping” down to S2. It was particularly common during the first REM period. This pattern may be seen in 1-8% of normal subjects [53]. There are two possibilities that could produce this kind of pattern. The stage could be alternating so fast that the epoch that has to be scored into a single stage is too long, in other words temporal resolution is insufficient; the other possibility is that parts of REM and NREM related brain processes could coexist.

It is evident that even when being meticulously followed, at some points the rules provide unspecific tools for sleep characterization. The problem of lacking phasic events for scoring SREM is especially frequent within the first REM period. The EEG of REM sleep is different from that of NREM sleep. Most often the EEG patterns can be separated by experienced electroencephalographists. Kubicki et al. have claimed that sometimes when the guidelines are strictly obeyed the first REM episode can be unnoticed which would artificially lengthen the SREM latency [3]. Even if REMs are not present the episodes could be scored SREM if typical EEG with sawtooth waves are accompanied by low muscle tonus [2,3].

On the other hand, Kubicki and Herrmann have postulated that long periods of S1 are often scored as S2 or SREM, even if the interpreter felt that they were actually S1 [4]. However, we think that the situation might not be that straightforward and therefore more physiological studies on the mechanisms are needed. Epochs with no clear landmarks should be scored as separate entities at least until their neurophysiological substrates have been revealed. Autonomic nervous system parameters, especially heart rate variation, could provide an aid in the differentiation between the states because they are known to be different in REM and NREM sleep [54]. The small Twitches in the distal parts of the extremities characteristic of SREM could also be used as an aid in scoring [55].

If coexistence of REMs and spindles in the same epoch is dependent on rapid stage shifts, visual adaptive scoring (VAS, S6, see below) might provide a solution. If there is physiological coexistence of both elements new stage categories could be developed. So far it is not clear whether subdivision of REM sleep into subcategories is possible and biologically meaningful. There are no clues as to whether these subdivisions could aid in the clinical diagnosis of sleep disorders. REM parameters have been studied, especially in psychiatric conditions [57]. It could also be of interest, although laborious, to calculate the sawtooth waves, their periodicity, lengths of periods and other possible parameters.
Clinical consequences of the limitations

Many organic sleep disorders have patterns that do not fit into the scoring criteria of R&K. The rules of R&K are clearly designed for normal, usual sleep patterns, not abnormal or deviant normal electrophysiological patterns. During normal aging the amplitude of the EEG decreases. Thus the slow activity might not exceed the 75 μV limit. The consequences are the flattening of the hypnogram with a lack of stages S3 and S4, even if they are physiologically present [27–30]. The rules could either be modified [29,30] or scoring of NREM stages could be replaced by quantitative analysis [56].

Alpha–delta sleep is characterized by the persistence of alpha activity during NREM sleep [58]. It can seen in a variety of conditions, but also in healthy sleepers [59]. Moldofsky has suggested that the alpha–delta pattern is a reflection of non-restorative sleep of fibromyalgia patients [60]. As there are no rules for scoring this pattern it is unclear how it should be classified. If alpha activity during sleep is abundant it can considerably hamper visual scoring. The determination of sleep onset and wakefulness episodes within sleep can be difficult. With low-alpha subjects the problem is reversed: the attenuation of alpha activity cannot be used as a marker of S1 and one has to rely on the theta activity.

Sleep apnoeas are respiratory pauses of 10–20 sec. They are as a rule terminated by an awakening or arousal. If the sleep episode remains shorter than half an epoch because of an apnoea related arousal the whole epoch has to be scored as wakefulness (Fig. 1). As apnoeas during wakefulness are not taken into account, the apnoea index will be too low. In multiple sleep latency test (MSLT) and MWT sleep latencies can become too long if apnoeas interfere with too short sleep episodes. This leads to the underestimation of the sleepiness of the subject. The problem can, for instance, be overcome by VASS (see below, Fig. 3). Some laboratories calculate sleep latencies to be the first sleep apnoea, even if the sleep episode would be less than half the epoch (personal information). Our opinion is that it would be preferable to have rules that fit the reality rather than making exceptions from originally too strict guidelines.

Sleep apnoeas are often aggravated in SREM. This relation to SREM can be overlooked if sleep is repeatedly terminated by respiratory-related arousals before the appearance of REMs. One can hardly get meaningful results about relationships between respiratory events and sleep stages if a considerable proportion of the events taking place in R&K scored S1 are actually in SREM.

The cyclic alternating pattern (CAP) described by Terzano et al. is characterized by sequences of two recurrent EEG-patterns each lasting more than 2, and less than 60, sec (for a recent brief overview and references, see 61). CAP consists of pseudoperiodic clusters of arousal-related phasic events (phase A) alternating with EEG background activities (phase B). CAP rate is reduced after sleep deprivation. It is increased, for instance, in environmental disturbances, organic sleep disorders and insomnia. The CAP cycles do not coincide with the fixed boundaries of the 20 or 30 sec epochs. Artificial stage changes can be obtained, or short sleep episodes within CAP sequences can be ignored, by R&K scoring. This is dependent on the duration of the phases and the epoch length used. Sleep stage scoring is often complemented by separate scoring of arousals. However, according to the ASDA criteria only a few types of A phases are considered as arousals [52]. In addition, attention is not usually paid to their periodicity.
Figure 1 Twenty-seven seconds of polygraphic recording with part of the channels shown. In the beginning the subject is drowsy with diffuse alpha activity and slow eye movements (drowsy-alpha). This is followed by a short episode of S1 with disappearance of the alpha and increasing theta activity. There is an apnoea during which the subject wakes up again. The alpha episodes seen on the occipital channels during wake-low are less than 1 sec and therefore ignored. (From reference no. 55, with permission).

Parasomnias, especially sleepwalking, express delta waves with concomitant body motility, high EMG-tonus and sometimes eye movements [62]. There are no rules for scoring such a state and therefore scoring results can sometimes be misleading. It might be scored as wakefulness, movement time, or one of the NREM stages, depending on the scorer and the electrophysiological picture.

Scoring of SREM is especially sensitive to the existence of REMs [3]. Usually the REMs do not appear immediately after onset of REM sleep. If the episode remains very short, for instance due to an arousal, it might be totally omitted even if all other SREM-related electrophysiological parameters, the typical EEG pattern, the low muscle tonus and the high heart rate variability are present (Fig. 2, see also Fig. 7 in [1]). If this happens in the first sleep cycle where the REM sleep episode is naturally short, the SREM latency will be markedly and artificially prolonged [3]. This can have clinical implications, especially in the diagnosis of narcolepsy or forms of depression associated with a shortened SREM latency. Also in experimental sleep studies the uncertainty of scoring SREM causes excessive variability in the SREM latency parameter.

In narcolepsy elevated submental muscle tonus can be seen in REM sleep [63]. This would affect the classification in two ways: at least in some subjects the state could be scored as wakefulness. In other instances most of the REM sleep period would be scored SREM but the onset would be delayed, as there would be no diminution of muscle tonus in REM sleep epochs preceding the appearance of the REMs [3].

REM sleep behaviour disorder is characterized by elevated submental muscle tonus and excessive chin or limb movements (for recent review see 64). Corresponding scoring problems to those of recordings from subjects with narcolepsy and parasomnia can appear. The REM sleep episodes could be scored either as wakefulness, SREM or sometimes partly S1, depending on the case.
Figure 2 Twenty-seven seconds of polygraphic recording with part of the channels shown. The EEG and other physiological activity of SREM without REMs (a) is similar to the activity of SREM with REMs (b). Usually the REMs appear some time after the onset of REM sleep. If the REM sleep episode remains short, for instance due to an arousal, the whole episode cannot be scored SREM and the SREM latency will be considerably prolonged.

The parameters obtained by sleep stage scoring give a superficial picture of the differences between good and poor sleep. Persistent psychophysiological insomnia is only reflected by an excess of S1 and sometimes an inadequacy of slow wave sleep (see [64] for review). In idiopathic insomnia sleep latency is increased, sleep efficiency is reduced and the number and duration of awakenings are increased. In sleep state misperception sleep parameters are normal even if the patient subjectively has poor...
sleep. No real clues about the possible mechanisms behind the various forms of insomnia are obtained. This has already raised the question of whether the sleep stages obtained by standard sleep stage scoring are at least insufficient, if not inadequate, measures of sleep quality. Recent experimental studies give evidence that analysis of sleep micro structure might provide the required insight into basic sleep mechanisms (see above). There is experimental evidence that sleep spindles are sleep protective, by blocking the transfer of external information to the cortex at the thalamic level (see above and [4] for review). This is further emphasized by the findings that benzodiazepines, which are used as hypnotics, also increase spindle activity [4] which can be one of the reasons for subjective perception of improved sleep quality. Although it is not known what would be the exact value of analysing sleep micro structure in clinical patients, it is clear that this is the direction towards which novel clinical studies should be directed.

**VASS—an alternative to R&K**

Visual adaptive scoring system (VASS [56]) could provide a solution to the problems with epoch scoring, the insufficient number of sleep stages and the ignorance of EEG topography of the standard rules. By VASS the stage change is assigned to the point when it actually takes place, and the stages are defined according to their electrophysiological content. All recorded EEG derivations, as well as other parameters, are taken into account.

Our first application of VASS was in the multiple sleep latency test (MSLT). VASS was more sensitive to short sleep spells than standard scoring and gave a shorter, and in our opinion, a more accurate sleep onset latency. The short lapses of alertness and short sleep episodes and arousals were more precisely revealed.

Digital polysomnography is in practice a necessity because with paper recordings the method would be too time consuming. With long, stationary periods VASS is less time consuming and it might be useful for evaluating slow wave sleep over longer periods. VASS fits the sleep/wake process more closely than standard scoring and can therefore be regarded as a better model. It might also increase inter-scorer reliability because 1–2 sec differences do not result in different scorings of whole 30 sec epochs. The improved temporal resolution gives more stationary epochs electrophysiologically for computer analysis.

VASS leads to greater accuracy, especially in the analysis of disturbed sleep with repetitive, short-term stage changes (Fig. 3). Common applications would be vigilance studies and polysomnograms of patients with respiratory sleep disorders.

**Conclusions**

The guidelines of Rechtschaffen and Kales were meant as a reference method for healthy adults with undisturbed sleep. The validation was evidently, even for this restricted group, both inadequate and insufficient. Studies dealing with validation of the rules for subjects with different disorders do not exist and it is surprising that stage scoring of disturbed sleep by the standard method is constantly used without critical evaluation of its validity. The low inter-rater agreements in visual scoring,
Figure 3 A rapid sleep onset and two major sleep episodes with some wakefulness in between are seen in the lower hypnogram scored by R&K. Using VASS (upper hypnogram) sleep onset is fragmented until S2. The second sleep episode is constantly interrupted by repetitive arousals (A). DA and DL are drowsiness stages. (From reference no. 55, with permission).

together with the low human/computer agreements, should have raised questions about the validity of the sleep stages as a description of sleep electrophysiology.

The use of the manual of Rechtschaffen and Kales as a gold standard has probably had a negative impact on the field. This has been, in our opinion, a result of lack of knowledge or understanding about the intentions of the guidelines. It is clearly stated in the manual itself that “the existence of established categories does not free the investigator from the task of devising a descriptive system which better communicates the unique features of the phenomena”[1]. With accumulating knowledge R&K can today be seen as an insufficient and inappropriate description of the sleep processes both in terms of biology and morphology. The necessary reevaluation and revision of the concepts which should constantly be performed does not exist.

Paper recordings should be replaced by digital systems in order to get a better dynamic range and bandwidth. More channels can be used in order to take into consideration topographical aspects. Temporal resolution can be increased with the visual adaptive scoring systems (VASS). Quantitative methods can be used either alone or in combination with visual analysis. There is evidence that computer analysis of slow activity is a more reliable tool than visual estimation of sleep stages [66] and the separation of NREM sleep stages should be replaced by computer analysis. This would
also enhance speed of analysis. This far the computer analysis methods have mostly been rejected because of the discrepancies with results obtained by R&K [5].

Stage determination should be based on continuous parameters rather than the appearance of phasic events. In this case use of other parameters, for instance autonomic nervous system activity, is among the most important ones. Stage division into subclasses should also be examined where appropriate. Other new approaches, for instance analysis of the cyclic alternating pattern, should be examined more widely. One possibility for overcoming the problem of lacking or coexisting landmarks is to abandon the stage concept and treat the NREM and REM as processes with a certain probability. This would require continuous measures that have not yet been validated.

Acknowledgements

This study was supported by grants from The Research Fund of The Tampere University Hospital and Biomed-2 BMH4-CT97-2040 (SIESTA) funded by the EU Commission.

Practice points
1. Alternative descriptions should be used when the guidelines do not fit the electrophysiological content of the episode.
2. Quantification of slow waves is a more useful and accurate method than sleep stage scoring of NREM stages.
3. Autonomic nervous system variables could be used as an aid in state determination.
4. Assessment of the validity of a new method should be based on how well it describes physiology or pathology, not how close it is to the guidelines.
5. More channels than those recommended by the guidelines should be used.

Research agenda
1. New analysis methods and definitions that describe the actual sleep processes more closely than the sleep stages should be developed and validated.
2. More attention should be paid to the infrastructure, the phasic events and their periodicities.
3. Effects of increased temporal resolution by, for instance, visual adaptive scoring systems (VASS) should be examined.
4. Alternative descriptions, for instance, the scoring of Cyclic Alternating Pattern (CAP) could provide useful information and be more accurate than stage scoring in many circumstances.
5. Methods should be validated on subject groups for which they are applied.
References


* The most important references are denoted by an asterisk.


