

## SHORT REPORT

# Sleep fragmentation: comparison of two definitions of short arousals during sleep in OSAS patients

M.V. Smurra<sup>\*,†</sup>, M. Dury<sup>#</sup>, G. Aubert<sup>#</sup>, D.O. Rodenstein<sup>\*</sup>, G. Liistro<sup>\*</sup>

*Sleep fragmentation: comparison of two definitions of short arousals during sleep in OSAS patients. M.V. Smurra, M. Dury, G. Aubert, D.O. Rodenstein, G. Liistro. ©ERS Journals Ltd 2001.*

**ABSTRACT:** The measurement of arousals during sleep is useful to quantify sleep fragmentation. The criteria for electroencephalography (EEG) arousals defined by the American Sleep Disorders Association (ASDA) have recently been criticized because of lack of interobserver agreement. The authors have adopted a scoring method that associates the increase in chin electromyography (EMG) with the occurrence of an  $\alpha$ -rhythm in all sleep stages (Université Catholique de Louvain (UCL) definition of arousals). The aim of the present study was to compare the two scoring definitions in terms of agreement and repeatability and the time taken for scoring in patients with obstructive sleep apnoea syndrome (OSAS) of varying severity.

Two readers using both ASDA and UCL definitions scored twenty polysomnographies (PSGs) each on two occasions. The PSGs were chosen retrospectively to represent a wide range of arousal index (from 6–82) in OSAS patients.

There was no difference in the arousal indices between readers and between scoring methods. The mean  $\pm$ SD difference between the two definitions (the bias) was  $1.1 \pm 3.76$  (95% confidence interval:  $-0.66$ – $2.86$ ). There was a strong linear relationship between the arousal index scored with the two definitions ( $r=0.981$ ,  $p<0.001$ ). Mean  $\pm$ SD scoring duration was significantly shorter for UCL than for ASDA definitions ( $18.5 \pm 5.4$  versus  $25.3 \pm 6.6$  min,  $p<0.001$ ).

In conclusion, it has been found that in obstructive sleep apnoea syndrome patients, the American Sleep Disorders Association and Université Catholique de Louvain definitions were comparable in terms of agreement and repeatability.

*Eur Respir J 2001; 17: 723–727.*

\*Pneumology and <sup>#</sup>EEG Units, Cliniques Universitaires Saint-Luc, Université Catholique de Louvain, Brussels, Belgium. <sup>†</sup>Hospital Muniz, Buenos Aires, Argentina.

Correspondence: D.O. Rodenstein  
Pneumology Division  
Cliniques Universitaires Saint-Luc  
10 av Hippocrate  
1200 Brussels  
Belgium  
Fax: 32 27643703

Keywords: Movement arousals  
polysomnography  
sleep apnoea

Received: January 21 2000  
Accepted after revision November 8 2000

Sleep fragmentation is one of the recognized determinants of excessive daytime sleepiness [1]. It is characterized by repetitive short interruptions of sleep. These interruptions can be variously defined using electroencephalography (EEG) [2] autonomic [3] or behavioural markers. An increase in the number of (short) arousals may be associated with periodic leg movements and sleep-disordered breathing [4]. The number of arousals related to the total sleep time, the arousal index, is useful to quantify sleep fragmentation.

The association between "any increase in electromyography (EMG) on any channel, accompanied by a change in pattern on any additional channel" initially described by RECHTSCHAFFEN and KALES [5] is termed movement arousal (MA). However, these authors did not use the arousals as an epoch score, but rather as a signal to a possible stage change. In 1984, STEPANSKI *et al.* [1] introduced a four-level arousal scoring system. Level I, the lower degree of arousal, included an increase in EEG frequency, and an increase in chin EMG amplitude; the intrusion of an  $\alpha$ -rhythm was considered as an intermediate level of arousal. Since 1987, the present authors' sleep laboratory (Université Catholique de Louvain; UCL) has adopted a definition

of arousals that requires an increase in chin EMG (in addition to the appearance of an  $\alpha$ -rhythm in the EEG) in all sleep stages.

The American Sleep Disorders Association (ASDA) criteria for EEG arousals were published in 1992 and are intended to serve as a basis for comparison between sleep laboratories. On an EEG tracing, arousals are identified by "an abrupt shift in EEG frequency" lasting for  $\geq 3$  s. However, the ASDA protocol has recently been criticized. Notably, DRINNAN *et al.* [7] have illustrated a lack of interobserver agreement. The fact that the ASDA definition does not take into account an increase in chin EMG in non-REM sleep was pinpointed as a possible cause of interobserver disagreement.

It has become common practice to include the arousal index in the assessment of sleep-disordered breathing disorders. The arousal index is an indirect indicator of the severity, not only of the obstructive sleep apnoea-hypopnoea syndrome (OSAS), but also of the upper airway resistance syndrome. Moreover, this index is useful to guide treatment decisions.

The aim of the present study was to compare UCL and ASDA scoring methods in terms of agreement and repeatability in a group of patients referred to the

authors' institution for suspicion of sleep-disordered breathing. The hypothesis was that the UCL definition of arousals, including EMG increase, would give better repeatability, and would require less time for scoring.

## Methods

### Polysomnography

Full-night diagnostic polysomnography (PSG) was performed in each subject according to standard criteria as previously described [5]. A microphone was glued on the anterior face of the patients' neck, over the larynx. Airflow was monitored by three thermocouples placed in front of the mouth and each nostril and linked to independent channels. Body position was recorded (Pro-Tech body position sensor, Woodinville, USA) via one channel.

All signals were recorded with a digital acquisition system (OSG Brainlab, Antwerp, Belgium). Sleep and respiratory parameters were recorded with the following sampling rate: Electrooculography (EOG) (2 channels, right and left) 128 Hz, chin EMG (1 channel) 512 Hz, EEG (3 channels: C4-A1, C3-A2, and C4-O2) 128 Hz, electrocardiography (ECG) 128 Hz, thoraco-abdominal movements 64 Hz,  $S_aO_2$  and pulse rate 16 Hz, as previously described [9].

The desaturation index (DI) was the number of >4% arterial oxygen desaturations per hour of sleep.

**UCL definition.** A UCL arousal was defined as the reappearance of an  $\alpha$ -rhythm in the EEG during a sleep epoch, accompanied by an increase in EMG, both lasting for at least 2 s. The increase in EMG activity was detected visually, with respect to the baseline of the sleep epoch (fig. 1).

**ASDA definition.** An ASDA arousal was defined as "an abrupt shift in EEG frequency, which may include theta, alpha and/or frequencies greater than 16 Hz but not spindles". Ten seconds of continuous sleep must precede the arousal. The arousal must last  $\geq 3$  s and it must be accompanied by an increase in chin EMG if it occurs during rapid-eye-movement sleep.

Twenty polysomnographic recordings were chosen retrospectively to represent a wide range of sleep fragmentation according to the UCL definition. The subjects had varying severity of OSAS but the choice of the PSG included in the study was guided only by the arousal index. Each PSG had been scored routinely using the UCL definition before the study. Routine scoring had been performed by one of 4 readers, and systematically reviewed. The basis of

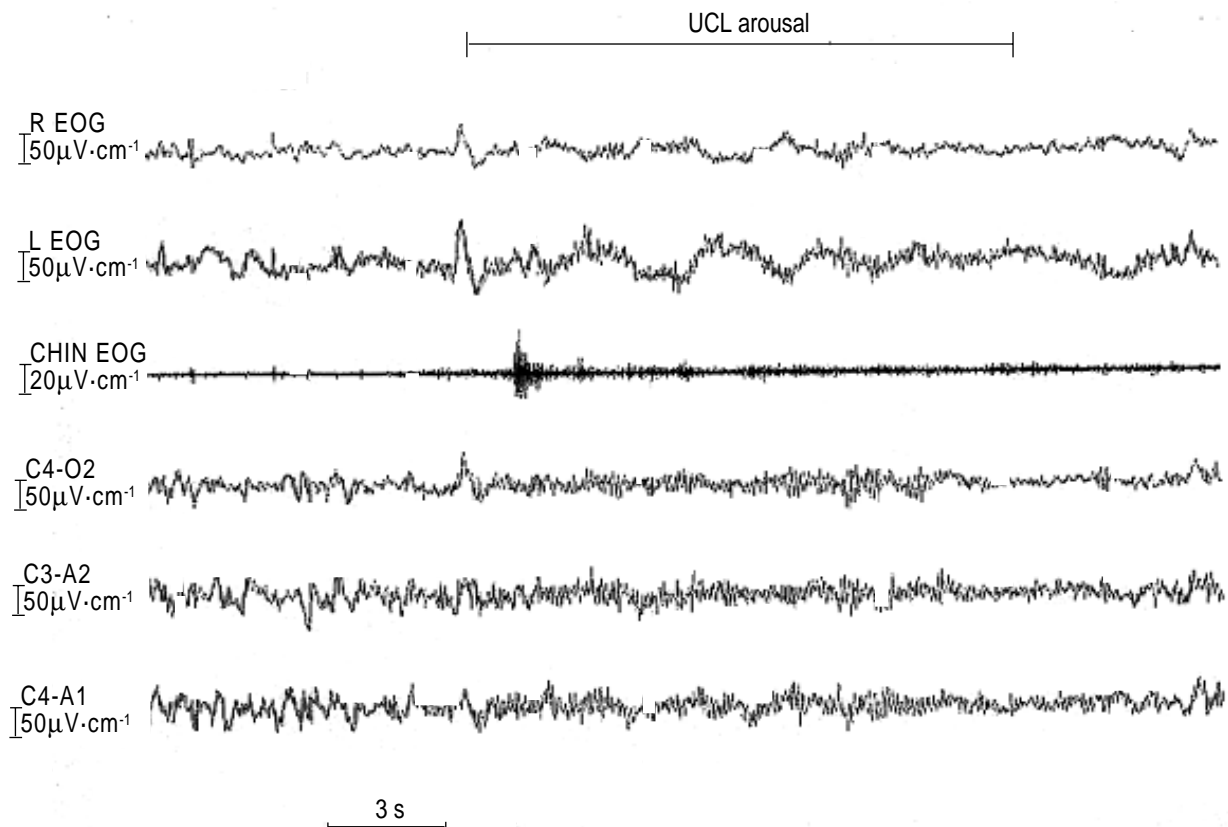


Fig. 1. – Example of an arousal according to the Université Catholique de Louvain (UCL) definition during stage I sleep. R EOG: right electrooculogram; L EOG: left electrooculogram, CHIN EMG: chin electromyogram, C4-O2, C3-A2, C4-A1: electroencephalographic derivations.

selection of the 20 PSGs was the UCL arousal index during routine analysis.

Two readers scored the twenty chosen PSGs using both the UCL and ASDA definitions of arousal. The PSGs were read in random order and each was read using each definition. After a first reading, all marks were erased and at least one week later, a second reading was done performed under the same conditions. The readers visualized only the EEG, EMG and EOG channels on the PC screen. The UCL arousal index (AIUCL) and ASDA arousal index (AIASDA) were defined as the number of arousals per hour of sleep. The first reader was an expert reader with more than 10 yrs practice in both routine and research studies. The second reader was much less experienced, and participated in the study at the end of two months' training in PSG. The arousal index measured on the 20 PSGs during the repeated readings by the two readers with the 2 scoring definitions were recorded (total:  $20 \times ((2+2)+(2+2))=160$  readings). The time spent to score each PSG by the expert reader with both definitions was also recorded.

#### Statistical methods

The intrareader repeatability was compared by one-way analysis of variance (ANOVA). The repeatability coefficient of the British Standards Institution was calculated. This coefficient equals 2 sds of the mean difference between two measurements. The smaller the coefficient the better the repeatability.

No difference was found between the repeated measurements of arousal index and the average of each pair of values of arousal index was used for further analysis ( $(2+2) \times 20=80$  values of arousal index).

Arousal indices were compared between readers and between definitions by two-way ANOVA. The agree-

ment between both methods of scoring was assessed by linear regression analysis and also according to the method of BLAND and ALTMAN [11]. The relationship between arousal index scored using the two definitions was compared between the readers using covariance analysis.

The time spent to score the PSG by the expert reader was compared by paired t-test. The relationship between the scoring time of the expert reader and the arousal index was assessed by linear regression. All data are presented as mean $\pm$ SD.

## Results

Table 1 shows the individual anthropometric data of the subjects, their desaturation index, AIUCL and AIASDA measured by the expert reader.

The repeatability coefficients of the arousal index measured for the two readers were 4.35 (arousals per hour of sleep) and 5.00 for the UCL, and 4.39 and 6.23 for the ASDA protocol. There was no difference in the arousal indices compared between readers and between definitions by two-way ANOVA.

There was a highly significant linear relationship between AIUCL and AIASDA measured by the two readers (fig. 2). There was no difference in the relationship between arousal index scored according to the two protocols within the two readers.

Figure 2 shows the relationship between the arousal index measured according to the two scoring methods by the expert reader and figure 3 shows the corresponding BLAND and ALTMAN plot. The latter presents the relationship between the difference in arousals ascending to the two definitions *versus* the average value of arousal index. The mean difference between the 2 definitions (the bias) was  $1.1 \pm 3.76$  (95% confidence interval (CI): -0.66 to 2.86). The difference was normally

Table 1. – Anthropometric data, desaturation index and arousal index of the subjects.

Subject	Sex	Age	BMI	DI	AIUCL	AIASDA
1	F	54	45.3	31	7	4
2	M	74	40.8	67	14	14
3	M	50	34.1	71	48	46
4	M	46	37.7	58	45	44
5	M	50	46.2	16	5	7
6	M	23	25.2	66	50	39
7	M	60	28.3	13	25	25
8	M	53	30.9	41	39	33
9	M	48	30.1	25	34	32
10	M	27	30.0	13	8	7
11	M	38	34.5	12	7	10
12	F	27	35.6	18	26	24
13	M	47	31.6	51	32	37
14	M	47	31.1	44	26	23
15	M	56	30.5	39	17	16
16	M	73	35.4	26	12	13
17	F	76	29.3	94	54	49
18	M	44	28.7	6	14	20
19	M	41	35.4	79	61	59
20	M	45	47.0	81	58	58
Mean $\pm$ SD		49 $\pm$ 14	34.4 $\pm$ 6.2	42.6 $\pm$ 26.9	29.1 $\pm$ 18.6	28.0 $\pm$ 17.1

BMI: body mass index ( $\text{kg}\cdot\text{m}^{-2}$ ), DI: desaturation index; AIUC: arousal index according to the protocol; AIASDA: arousal index according to the American Sleep Disorders Association protocol.

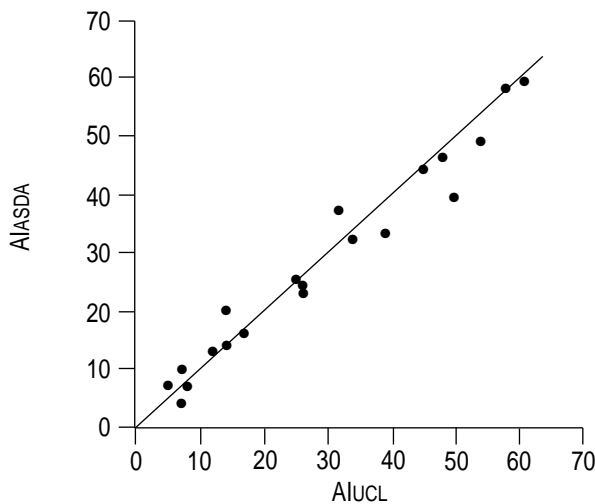


Fig. 2. – Relationship between the arousal index (AI) measured by the expert reader according to the Université Catholique de Louvain (AIUCL) and the American Sleep Disorders Association (AIASDA) definitions. The identity line is shown.  $r=0.981$ ,  $p<0.001$ .

distributed (skewness and kurtosis coefficients  $<2$ ). The limits of agreement between both methods were  $-6.28$  and  $8.49$  ( $\text{mean} \pm 1.96 \text{ SD}$ , fig. 2). The 95% CI for the lower limits of agreement were  $-3.22$  and  $-9.33$ , respectively. The 95% CI for the upper limits were  $5.43$  and  $11.54$ , respectively.

The average duration of scoring by the expert reader was greater with the ASDA ( $25.3 \pm 6.6$  min) than with the UCL definition ( $18.5 \pm 5.4$  min),  $p<0.001$ . There was a positive linear relationship between the arousal index and the duration of scoring:  $r=0.54$ ,  $p=0.01$  and  $r=0.60$ ,  $p=0.005$  for the UCL and ASDA definitions, respectively, so that the greater the sleep fragmentation, the longer it took to read the PSGs.

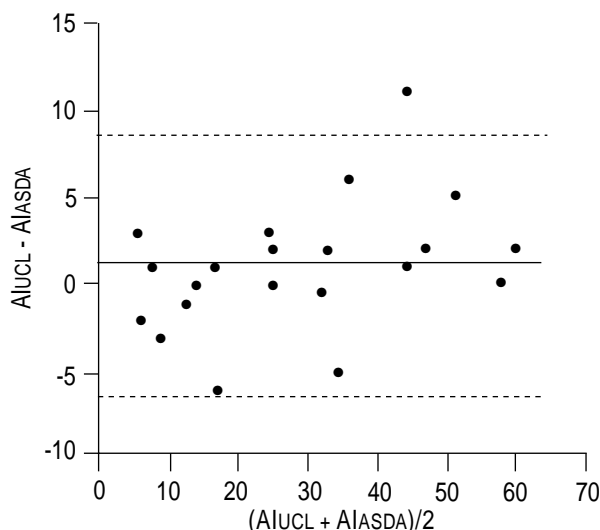


Fig. 3. – Bland and Altman plot showing no relationship between the difference in arousal index (AI) measured with the American Sleep Disorders Association (AIASDA) and Université Catholique de Louvain (AIUCL) definitions against average arousal index by the two methods: mean; ----: 1.96 SD of the difference.

## Discussion

There was good agreement in scoring using two different definitions of arousals in a group of patients with OSAS. The ASDA and UCL definitions were comparable in terms of agreement and also in their repeatability. The time spent by the expert reader was significantly shorter with the UCL than with the ASDA scoring methods and was related to sleep fragmentation.

Contrary to the hypothesis, the UCL definition of arousals, including EMG increase in all sleep stages, did not allow for better repeatability. It was expected that this method would permit better pattern recognition for the visual assessment of arousal on screen. However this was not the case. The scoring method is not easier and the experience of the reader did not influence the reproducibility.

The only difference between the scoring methods was the time spent to read the PSGs. The shorter time to score a PSG using the UCL definition of arousals may represent a true advantage due to the inclusion of the EMG increase, allowing, if not for better, at least for faster visual pattern recognition. It could, however, also be due to the fact that the readers were more familiar with the UCL protocol.

A recent report showed poor interobserver agreement in the total number of arousals with the ASDA scoring method [7]. The ASDA definition was criticized as a possible cause of the interobserver disagreement. These authors suggested that extending EMG criteria to all sleep stages might improve the interobserver agreement. The present authors have not found significant discrepancies between the scoring of this study's two readers with the two definitions. However, a different method was used to quantify the interobserver agreement. In the report of DRINNAN *et al.* [7], the different readers scored individual epochs, while this study did not assess whether a given event was scored as an arousal with one definition and not with the other. However, the overall indices were similar and statistically the same as the routine initial reading. It is hard to believe that there could have been a systematic difference between the events identified with the ASDA definition and those identified with the UCL definition.

More recently, LOREDO *et al.* [12] reported high interscorer reliability with the ASDA definition. These authors used the intraclass correlation (ICC) test to compare the interscorer reliability of five definitions of arousals. The ASDA definition had a good interscorer reliability (ICC 0.84) but the UCL definition was not tested. The ICC calculated with the present data were 0.96 with the ASDA and 0.98 with the UCL definitions.

Quantifying sleep fragmentation is believed to be tedious and cumbersome. Indeed, the present authors have found that scoring time is related linearly to the arousal index: the more sleep is fragmented, the more time it takes to measure it. However, it is interesting to note that the time needed to count the arousals with indices as high as 60, rarely exceeds 30 min with the UCL definition.

Several recent studies have suggested that EEG arousals are only one form of sleep interruption (so-called

cortical arousals), but that there could be other, more subtle subcortical, or autonomic arousals which might be easier to identify and especially to detect using automated methods. Until much more research material is available, the arousal index will remain of prime importance in the description of sleep fragmentation.

In conclusion, it has been found that the American Sleep Disorders Association and the Université Catholique de Louvain definitions were comparable in terms of agreement and repeatability, in patients with varying degrees of obstructive sleep apnoea syndrome. The inclusion of an electromyography increase in all sleep stages did not result in better reproducibility, but seemed to lead to a faster reading.

### References

1. Stepanski E, Lamphere J, Badia P, Zorick F, Roth T. Sleep fragmentation and daytime sleepiness. *Sleep* 1984; 7: 18–26.
2. Atlas task force of the American Sleep Disorders Association. EEG Arousals: scoring rules and examples. *Sleep* 1992; 15: 174–184.
3. Davies RJ, Belt PJ, Roberts SJ, Ali NJ, Stradling JR. Arterial blood pressure responses to graded transient arousal from sleep in normal humans. *J Appl Physiol* 1993; 74: 1123–1130.
4. Arousal responses to respiratory stimuli during sleep. In: Saunders N, Sullivan CE, eds. *Sleep and breathing*. New York, Marcel Dekker, 1984; pp. 137–161.
5. A manual of standardized terminology, techniques and scoring system for sleep stages in human subjects. Los Angeles, UCLA Brain information service, 1968.
6. Aubert Tulkens G, Hamoir M, Van den Eeckhaut J, Rodenstein DO. Failure of tonsil and nose surgery in adults with long-standing severe sleep apnea syndrome. *Arch Intern Med* 1989; 149: 2118–2121.
7. Drinnan MJ, Murray A, Griffiths CJ, Gibson GJ. Interobserver variability in recognizing arousal in respiratory sleep disorders. *Am J Respir Crit Care Med* 1998; 158: 358–362.
8. Collard P, Dury M, Delguste P, Aubert G, Rodenstein DO. Movement arousals and sleep-related disordered breathing in adults. *Am J Respir Crit Care Med* 1996; 154: 454–459.
9. Parreira VF, Delguste P, Jounieaux V, Aubert G, Dury M, Rodenstein DO. Glottic aperture and effective minute ventilation during nasal two-level positive pressure ventilation in spontaneous mode. *Am J Respir Crit Care Med* 1996; 154: 1857–1863.
10. British Standards Institution. Precision of test methods. I: Guide for the determination and reproducibility for a standard test method (BS 5497, part 1). London, BSI, 197.
11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
12. Lored JS, Clausen JL, Ancoli Israel S, Dimsdale JE. Night-to-night arousal variability and interscorer reliability of arousal measurements. *Sleep* 1999; 22: 916–920.
13. Pitson DJ, Stradling JR. Autonomic markers of arousal during sleep in patients undergoing investigation for obstructive sleep apnoea, their relationship to EEG arousals, respiratory events and subjective sleepiness. *J Sleep Res* 1998; 7: 53–59.
14. Argod J, Pepin JL, Levy P. Differentiating obstructive and central sleep respiratory events through pulse transit time. *Am J Respir Crit Care Med* 1998; 158: 1778–1783.
15. Martin SE, Engleman HM, Kingshott RN, Douglas NJ. Microarousals in patients with sleep apnoea/hypopnoea syndrome. *J Sleep Res* 1997; 6: 276–280.