Sleep and Breathing (2019) 23:1309–1314
https://doi.org/10.1007/s11325-019-01937-w

NEUROLOGY • ORIGINAL ARTICLE

Interobserver reliability of ICSD-3 diagnostic criteria for disorders of arousal in adults

Giuseppe Loddo[1] · Luca Vignatelli[2] · Corrado Zenesini[2] · Francesco Lusa[1] · Luisa Sambati[1,2] · Luca Baldelli[1] · Valentina Favoni[1,2] · Laura Rosa Pisani[3] · Giulia Pierangeli[1,2] · Sabina Cevoli[2] · Federica Provini[1,2]

Received: 22 May 2019 / Revised: 28 August 2019 / Accepted: 7 September 2019 / Published online: 14 September 2019

Abstract

Purpose Disorders of arousal include confusional arousals, sleepwalking and sleep terrors. The diagnosis of disorders of arousal is based on the clinical criteria established in the International Classification of Sleep Disorders, third edition, although the interobserver reliability of these criteria has never been investigated. The aim of this study was to estimate the inter-rater reliability of the diagnostic criteria for disorders of arousal throughout the whole life in order to understand their feasibility in clinical daily activity and in multicenter observational studies.

Methods Three raters interviewed 126 subjects (patients complaining of sleep disorders, headache, and healthy subjects), aged 18–80 years, with a standardized questionnaire created by applying the International Diagnostic Criteria for Disorders of Arousal.

Results An "almost perfect" inter-rater reliability for disorders of arousal criteria and the final diagnosis was found among the raters (kappa 0.89 for confusional arousals, 0.87 for sleepwalking, and 0.87 for sleep terrors).

Conclusions The International Classification of Sleep Disorders, Third Edition criteria are adequate for a reliable diagnosis of disorders of arousal. Further validation studies, confirming DOA diagnosis with video polysomnography, are needed to investigate the predictive value of ICSD-3 criteria.

Keywords Movement disorders · Questionnaires · NREM parasomnias · Sleepwalking · Sleep terror · Confusional arousal

Introduction

Disorders of arousal (DOA) are a group of NREM-related parasomnias including confusional arousals (CA), sleepwalking (SW), and sleep terrors (ST) according to the International Classification of Sleep Disorders, third edition (ICSD-3) [1]. DOA are characterized by abnormal sleep-related complex movements and behaviors associated with various degrees of autonomic nervous system activation, limited or absent cognition or dream imagery, and partial-to-complete amnesia. As with many sleep disorders, DOA do not occur spontaneously but are thought to be the result of several interacting factors in genetically susceptible individuals [2]. Behavioral and cognitive patterns observed during a DOA episode are explained by the co-occurring deactivation of frontoparietal associative cortices (typical of sleep) and activation of the posterior cingulate and anterior cerebellum networks without deactivation of the thalamus (typical of wake) [3].

DOA are considered benign phenomena, especially in children, and do not usually have a serious impact on sleep quality and quantity. CA affect from 20 to 50% of children, mainly infants and toddlers [4, 5]. Sleep terrors have the greatest incidence in preschool children. Sleepwalking is common in older children but most episodes usually resolve after the age of 10 years [6–8]. SW has been reported in 66% of children of 3-10 years persisting after 12 years in 24%. ST has been reported in 84% of children between 3-10 years persisting after 12 years in 6% [7].

In adulthood, DOA prevalence ranges from 1.6 to 2.4%, although prevalence could be underestimated [8–14]. Although

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11325-019-01937-w) contains supplementary material, which is available to authorized users.

✉ Federica Provini
federica.provini@unibo.it

[1] Dipartimento di Scienze Biomediche e Neuromotorie, Università di Bologna, Bologna, Italy

[2] IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy

[3] UOC di Neurologia, Ospedale Cutroni Zodda, Barcellona PG, Messina, Italy

childhood DOA are often transitory and harmless, in adulthood, they have a significant harmful potential, including serious injuries to the sleeper, bed partner, or others [15, 16]. Especially in adulthood, a correct diagnosis of DOA is essential not only to prevent harmful potential DOA consequences, but also to differentiate DOA from other sleep motor manifestations. The most important differential diagnosis is with sleep-related hypermotor epilepsy (SHE), a sleep-related focal epilepsy characterized by complex, often bizarre and/or violent motor behaviors, including asymmetrical tonic or dystonic posture [17–20]. Furthermore, DOA need to be differentiated from REM sleep behavior disorder (RBD), a REM parasomnia common in men who are over 50 years of age, characterized by complex motor behaviors during REM sleep due to the loss of physiological atonia of the skeletal muscles during REM sleep [21].

Currently the diagnosis of DOA is based on clinical criteria reported in the ICSD-3 [1].

Video polysomnography is indicated only for the evaluation of atypical, complicated, and injurious DOA or to exclude other sleep disorders such as obstructive sleep apnea syndrome or periodic limb movements triggering DOA [1, 22, 23]. A recent work assessed the diagnostic value of video polysomnography for DOA and highlighted slow-wave sleep fragmentation and slow/mixed arousal indexes as relevant biomarkers to be used in clinical and research settings [24].

Although DOA diagnosis is mainly based on the ICSD-3 DOA diagnostic criteria, the interobserver reliability (IR) of these criteria has not been assessed.

Our group has just investigated IR of DOA diagnostic criteria established in 2001 in the International Classification of Sleep Disorders-Revised (ICSD-R) [25], showing "substantial" IR for CA (kappa 0.74), "slight" IR for SW (kappa 0.36), and "fair" IR for ST (Kappa 0.02) [26]. DOA diagnostic criteria have been subsequently modified in 2014 in the ICSD-3 [1].

The aim of this study is to estimate the IR of the new DOA diagnostic criteria throughout the whole life by means of a standardized interview applying the ICSD-3 criteria.

Methods

Design and ethics

For the publication of this research, we followed the recommendations of the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [27].

The study was approved by the Bologna-Imola Ethics Committee in 15/03/18 (no. 17175).

Sample and raters

The sample included patients consecutively referred to the Sleep Disorders Centre and Headache Disorders Centre of the Institute of Neurological Sciences of Bologna from April to June 2018 and subjects accompanying the patients.

The subjects included in the study were >18 years old, Italian mother tongue, and without cognitive impairment.

We included as raters one neurology resident and two general practitioners in order to understand if the questionnaire could be applied in clinical practice also by clinicians without experience in sleep medicine.

Structured interview

A structured interview was created according to the DOA ICSD-3 criteria [1].

Since ICSD-3 criteria do not provide standardized operating procedures, in order to be more pertinent to a practical history-taking procedure, we decided to ascertain firstly the presence of a single DOA and subsequently the presence of general DOA criteria (Online Resource 1, 2).

Each criterion of the ICSD-3 for the diagnosis of DOA was posed in the form of the question: "During your life, have you ever…." The English text of the ICSD-3 was translated into Italian by a sleep medicine expert and back translated into English by a mother tongue professional translator blinded with respect to the original version. Adherence to the original text was assessed by another sleep medicine doctor.

In the interview we included 5 questions regarding age at onset of DOA, episode frequency at onset and 12 months after onset, and, for patients without a stable disease course, episodes peak and last 12 months frequency (Online Resource 1, 2).

Sampling methods

Informed consent was obtained from all individual participants included in the study. After signing the informed consent, each subject underwent the structured interview in order to diagnose the occurrence of DOA at any time in the subject's life.

The interview was administered separately, in the same day, three times to each subject, one for each rater. The sequence of the interviewers was randomly defined. Parental or another observer input was collected when possible to improve the accuracy of lifelong history.

Each rater, unaware of the diagnostic classification of the other raters and of the true reason for referral of subjects, classified the results of the questions and recorded them on a form.

Statistical analysis

The overall proportion of agreement and IR were evaluated for the single and general DOA criteria and the final diagnosis of each DOA for each pair of doctors. IR was calculated by

Table 1 Mean of positive observations, overall agreement, and interobserver reliability (kappa and confidence interval—CI) among the three raters for the questions regarding general criteria of disorders of arousal according to the ICSD-3 [1]

	Mean of positive observation *n* (%)	Overall agreement (%)	kappa	95% CI	Category of agreement
Confusional arousals					
Question for Criterion B During your life, have you ever or have you ever been told that while sleeping you suddenly woke up and were confused or behaved in a confused manner?	18 (14)	94	0.85	0.70–1	Almost perfect
Question for Criterion C If yes, were night terrors and sleepwalking excluded during these episodes?	18 (14)	94	0.85	0.70–1	Almost perfect
Final diagnosis	3 (2)	99	0.89	0.74–1	Almost perfect
Sleepwalking					
Question for Criterion B During your life, have you ever or have you ever been told that you got out of bed and started walking or presenting other complex behaviors out of bed?	23 (18)	98	0.95	0.78–1	Almost perfect
Final diagnosis	2 (1)	99	0.87	0.72–1	Almost perfect
Sleep terrors					
Question for Criterion B During your life, have you ever had, or have you ever been told that during sleep you presented episodes of sudden fear typically starting with alarmed vocalizations like a shout of fear?	37 (29)	96	0.93	0.77–1	Almost perfect
Question for Criterion C If yes, was this episode accompanied by dilation of pupils, tachycardia, rapid breathing, and sweating?	28 (22)	88	0.78	0.54–1	Substantial
Final diagnosis	2 (1)	99	0.87	0.72–1	Almost perfect

kappa statistics, which is the ratio of the observed agreement beyond chance to the potential agreement, according to the formula of kappa for dichotomous data and more than two raters, proposed by Fleiss [28]. Kappa values were interpreted according to conventional groups (0.0–0.20 = slight agreement, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, 0.81–1.00 = almost perfect) [29].

Results

The sample comprised 126 subjects (76 females) aged 18–80 years at the time of the interview (mean age ± standard deviation, SD, 50 ± 15 years). Duration of education was 5–27 years (mean 13 ± 4 years).

The reasons for clinical evaluation of the 46 patients referred to the Sleep Disorders Centre included obstructive sleep apnea syndrome (22), insomnia (10), parasomnias (10), restless legs syndrome (3), and circadian rhythm sleep-wake disorders (1), respectively.

Fifty-three subjects were patients referring to the Headache Disorders Centre for primary headaches. 27 were subjects accompanying the patients. Parental or another observer input was collected for 52 subjects.

An "almost perfect" IR was found among the three raters for the final diagnosis of each DOA (kappa 0.89 for CA, 0.87 for SW, 0.87 for ST, Table 1) and for their specific criteria (except Criterion C of ST, "substantial" IR).

An "almost perfect" or "substantial" IR was found for DOA general criteria, except a "moderate" IR for C general Criterion when ST is assessed, and "slight" to "moderate" IR for D general criterion when CA and SW are assessed (Table 2).

Taking into account the small sample size, no differences were observed in IR for CA and SW, if the interviews were performed with or without witnesses. Analysis was not applicable for ST.

A "substantial"/"almost perfect" IR among the three raters (kappa range 0.64–0.99) was found also for the majority of time-dependent descriptors of DOA (age at onset, episode frequency during life) (Table 3).

The number of positive observations for DOA according to the different raters are reported in Online Resource 3, 4, 5.

Table 2 Interobserver reliability (kappa and confidence interval—CI) among the three raters for the questions regarding general criteria of disorders of arousal (A–E) according to the ICSD-3 [1]

	CA °	95% CI	Category of agreement	SW°	95% CI	Category of agreement	ST°	95% CI	Category of agreement
Question for Criterion A During these episodes were you partially awake?	0.86	0.46–1	Almost perfect	0.84	0.48–1	Almost perfect	0.76	0.52–1	Substantial
Question for Criterion B During these episodes did you ever not respond or respond inappropriately to someone's attempts to intervene or guide you?	0.76	0.39–1	Substantial	0.94	0.55–1	Almost perfect	0.81	0.56–1	Almost perfect
Question for Criterion C During these episodes were you partially conscious (for example did you see a single visual scene)? Were you unconscious? Did you see images as if you were dreaming?	0.99	0.50–1	Almost perfect	0.61	0.37–0.85	Substantial	0.58	0.40–0.76	Moderate*
Question for Criterion D Did you have little or no recall of the episode?	0.60	0.33–0.87	Moderate*	0.18	0.12–0.24	Slight*	0.84	0.58–1	Almost perfect
Question for Criterion E Can you rule out that these episodes were due to other sleep problems, medical conditions, use of drugs or other substances?	0.81	0.43–1	Almost perfect	0.72	0.42–1	Substantial	0.67	0.46–0.88	Substantial

CA confusional arousals, *SW* sleepwalking, *ST* sleep terrors

°kappa

*Less than desired level of agreement

Discussion

This work represents the first part of a questionnaire validation study created in order to screen for DOA by applying the ICSD-3 criteria.

The diagnosis of DOA is mainly based on clinical criteria established in these criteria, although their IR has not been investigated.

IR provides information about the quality of measurements. On one side, reliability represents the ability of a measurement to discriminate between subjects or objects. On the other side, agreement is the degree to which scores or ratings are identical [28].

Evaluating the IR of ICSD-3 DOA diagnostic criteria, it is important to understand if ICSD-3 criteria are reliable independently of the variability in the interpretation of patient information (interpretation variance).

Reliable criteria will therefore be relevant to provide uniform diagnostic procedures in multicenter observational studies or to compare results from different studies.

Our study on the IR of the ICSD-3 diagnostic criteria for DOA occurring at any time in the patient's life disclosed an "almost perfect" level of IR for all DOA single criteria except for criterion C of ST, whose agreement was substantial (Table 1).

These agreement rates are clearly above the usual level for clinical agreement and reveal that ICSD-3 diagnostic criteria for DOA are reliable independently of interviewers, who are not sleep experts, and independently of the possible instability of patient responses.

In comparison with the previous criteria of the ICSD-R, our study reveals that the ICSD-3 criteria are more reliable [26].

The ICSD-R criteria showed a "fair" IR for SW because of disagreement on criterion C2 (amnesia following an episode) and ST showed "slight" IR due to a disagreement based on the first criterion (the patient complains of a sudden episode of intense terror during sleep).

Two argumentations may explain the different agreement between ICSD-R and ICSD-3 criteria. First, DOA single criteria are better defined in the ICSD-3 version. Second, the shared clinical features of DOA (limited or absent cognition or

Table 3 Interobserver reliability (kappa and confidence interval—CI) among the three raters for age at onset and frequency of disorders of arousal

	Confusional arousal		Sleepwalking		Sleep terrors	
	kappa	95% CI	kappa	95% CI	kappa	95% CI
Age at onset	0.92	0.63–1	0.72	0.52–0.92	0.76	0.61–0.91
Episodes frequency at onset	0.99	0.60–1	0.73	0.46–1	0.96	0.72–1
Episodes frequency 12 months after onset	0.64	0.46–0.82	0.52*	0.41–0.63	0.60*	0.51–0.69
Peak frequency	0.65	0.36–0.94	0.57*	0.40–0.74	0.71	0.50–0.92
Episodes frequency 12 months before the interview	0.30*	0.18–0.42	0.99	0.45–1	0.67	0.39–0.95

*Less than desired level of agreement

dream imagery, and partial-to-complete amnesia) are now collected into five common general mandatory criteria. These two improvements probably reduced the variability in the interpretation of patient's information.

Considering the single items of the general criteria (Table 2), we found some disagreement for the C criterion (limited or no associated cognition or dream imagery) and the D criterion (partial or complete amnesia for the episode) and they could be possibly improved in future versions of the criteria. In particular, the source of disagreement in the C criterion is probably the interpretation variance due to the intrinsic ambiguity or to the difficulties in correctly interpreting some patients' reporting, referring to the cognition occurring soon after the investigated phenomena. The source of disagreement in the D criterion, similarly to our previous work [26], (Table 2) could be the information variance due to the instability of patient responses linked to past events difficult to recall.

However, it has to be underlined that complete amnesia of the event is not standard for adult DOA, while in children amnesia might be more common, possibly because of higher arousal thresholds [30].

Another interesting point emerging from our study is the discrepancy between the percentage of positive observations when people were interviewed about the criteria discriminating for each type of DOA (18% for CA, 23% for SW, and 29% for ST) and the percentage of positive observations on final diagnosis (2% for CA and 1% for SW and ST) (Table 1). These percentages on final diagnosis can be explained considering the low percentages of positive observations regarding A and B general criteria **(**Online Resource 3, 4, 5**).** In particular, among 18 possible CA patients only 10 were positive for A general criterion and 9 for B (Online Resource 3). Among 23 possible SW patients, only 6 were positive for A general criterion and 11 for B (Online Resource 4). Among 28 possible ST patients, only 13 were positive for A general criterion and 10 for B (Online Resource 5). In our opinion the A general criterion, requiring the presence of "incomplete awakening" during DOA episodes, could exclude patients who cannot describe clearly the real complexity of awakenings related to DOA episodes. Furthermore, the B general criterion, requiring "inappropriate or absent responsiveness to efforts of others to intervene or redirect the person during the episode," could exclude patients who sleep alone.

It is therefore possible that general DOA criteria improve specificity of ICSD-3 DOA criteria at the expense of their sensitivity. Literature data reveal indeed a higher lifetime prevalence percentage for CA (18.5%), SW (6.9–29%), and ST (10.4%) [10, 11, 14]. It is also to be highlighted that all literature epidemiologic studies on DOA have been conducted without validated questionnaires. Only video polysomnographic studies validating ICSD-3 criteria could reveal their real specificity and sensitivity value.

Our study has some limitations including the limited sample size of the possible DOA population screened and the patient's difficulties in answering some questions especially on the cognition during the episodes and on DOA frequency, in particular during childhood.

Summing up, our study offers the following conclusions:

1. For a reliable diagnosis of DOA the application of the ICSD-3 DOA criteria is probably adequate.
2. ICSD-3 DOA criteria probably define DOA with high specificity and low sensitivity.
3. Further validation studies, confirming DOA diagnosis with video polysomnography, are needed to investigate the predictive value of ICSD-3 criteria.

Acknowledgements The authors wish to thank Cecilia Baroncini and Anne Collins for English editing.

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Compliance with ethical standards

Conflict of interest Dr. Loddo, Dr. Vignatelli, Dr. Zenesini, Dr. Lusa, Dr. Sambati, Dr. Baldelli, Dr. Favoni, Dr. Pisani, Dr. Pierangeli, Dr. Cevoli report no disclosures. Dr. Provini received honoraria for speaking

engagements or consultancies from Sanofi, Bial, Fidia, Vanda Pharmaceutical, Zambon, Eisai Japan, Italfarmaco.

Ethical approval All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and national research committee (Bologna-Imola Ethics Committee, no. 17175) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. American Academy of Sleep Medicine. The International Classification of Sleep Disorders, Third Edition (American Academy of Sleep Medicine, 2014).
2. Pressman MR (2007) Factors that predispose, prime and precipitate NREM parasomnias in adults: clinical and forensic implications. Sleep Med Rev 11:5–30
3. Bassetti C, Vella S, Donati F, Wielepp P, Weder B (2000) SPECT during sleepwalking. Lancet. 356(9228):484–485
4. Broughton RJ (1968) Sleep disorders: disorders of arousal? Enuresis, somnambulism, and nightmares occur in confusional states of arousal, not in "dreaming sleep". Science (New York, NY) 159:1070–1078
5. Stores G (2009) Aspects of parasomnias in childhood and adolescence. Arch Dis Child 94:63–69
6. Klackenberg G (1982) Somnambulism in childhood–prevalence, course and behavioral correlations. A prospective longitudinal study (6-16 years). Acta Paediatr Scand 71:495–499
7. Laberge L, Tremblay RE, Vitaro F, Montplaisir J (2000) Development of parasomnias from childhood to early adolescence. Pediatrics. 106:67–74
8. Ohayon MM, Guilleminault C, Priest RG (1999) Night terrors, sleepwalking, and confusional arousals in the general population: their frequency and relationship to other sleep and mental disorders. J Clin Psychiatry 60:268–276
9. Bjorvatn B, Grønli J, Pallesen S (2010) Prevalence of different parasomnias in the general population. Sleep Med 11:1031–1034
10. Hublin C, Kaprio J, Partinen M, Heikkilä K, Koskenvuo M (1997) Prevalence and genetics of sleepwalking: a population-based twin study. Neurology. 48:177–181
11. Hublin C, Kaprio J (2003) Genetic aspects and genetic epidemiology of parasomnias. Sleep Med Rev 7:413–421
12. Moreno MA (2015) Sleep terrors and sleepwalking. JAMA Pediatr 169:704
13. Ohayon MM, Mahowald MW, Dauvilliers Y, Krystal AD, Léger D, Leger D (2012) Prevalence and comorbidity of nocturnal wandering in the US adult general population. Neurology. 78:1583–1589
14. Stallman HHM, Kohler M, Broughton R, Sadeh A, Mume C, Baddeley A et al (2016) Prevalence of sleepwalking: a systematic review and meta-analysis. PLoS One 11:e0164769
15. Moldofsky H, Gilbert R, Lue FA, MacLean AW (1995) Sleep-related violence. Sleep. 18:731–739
16. Schenck CH, Milner DM, Hurwitz TD, Bundlie SR, Mahowald MW (1989) A polysomnographic and clinical report on sleep-related injury in 100 adult patients. Am J Psychiatry 146:1166–1173
17. Bisulli F, Vignatelli L, Naldi I, Pittau F, Provini F, Plazzi G et al (2012) Diagnostic accuracy of a structured interview for nocturnal frontal lobe epilepsy (SINFLE): a proposal for developing diagnostic criteria. Sleep Med 13:81–87
18. Licchetta L, Bisulli F, Vignatelli L, Zenesini C, Di Vito L, Mostacci B et al (2017) Sleep-related hypermotor epilepsy. Neurology. 88:70–77
19. Provini F, Plazzi G, Tinuper P, Vandi S, Lugaresi E, Montagna P (1999) Nocturnal frontal lobe epilepsy: a clinical and polygraphic overview of 100 consecutive cases. Brain. 122:1017–1031
20. Tinuper P, Provini F, Bisulli F, Vignatelli L, Plazzi G, Vetrugno R et al (2007) Movement disorders in sleep: guidelines for differentiating epileptic from non-epileptic motor phenomena arising from sleep. Sleep Med Rev 11:255–267
21. Plante DT, Winkelman JW (2006) Parasomnias. Psychiatr Clin N Am 29:969–987
22. Drakatos P, Marples L, Muza R, Higgins S, Nesbitt A, Dongol EM et al (2019) Video polysomnographic findings in non-rapid eye movement parasomnia. J Sleep Res 28(2):e12772
23. Fois C, Wright M-AS, Sechi G, Walker MC, Eriksson SH (2015) The utility of polysomnography for the diagnosis of NREM parasomnias: an observational study over 4 years of clinical practice. J Neurol 262(2):385–393
24. Lopez R, Shen Y, Chenini S, Rassu AL, Evangelista E, Barateau L et al (2018) Diagnostic criteria for disorders of arousal: a video-polysomnographic assessment: disorders of arousal. Ann Neurol 83:341–351
25. American Academy of Sleep Medicine (2001) International Classification of Sleep Disorders, revised: Diagnostic and Coding Manual. Rochester, Minnesota: American Academy of Sleep Medicine
26. Vignatelli L, Bisulli F, Zaniboni A, Naldi I, Fares JE, Provini F et al (2005) Interobserver reliability of ICSD-R minimal diagnostic criteria for the parasomnias. J Neurol 252:712–717
27. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A et al (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol 64:96–106
28. Fleiss JL, Levin B, Paik MC (2003) Statistical methods for rates and proportions. Third edition. Wiley, New York
29. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics. 33:159–174
30. Pilon M, Desautels A, Montplaisir J, Zadra A (2012) Auditory arousal responses and thresholds during REM and NREM sleep of sleepwalkers and controls. Sleep Med 13:490–495