Scoring of increased upper airway resistance events

a robust standard scoring rule would increase reliability

Bjartur Thorlacius

Thesis for a Master of Science degree in Computational Engineering

Supervisors: Erna Sif Arnardóttir, Rögnvaldur Sæmundsson, and Tómas Philip

University of Iceland

School of Engineering and Natural Sciences

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science

2022

SCORING OF INCREASED UPPER AIRWAY RESISTANCE EVENTS:

A ROBUST STANDARD SCORING RULE WOULD INCREASE RELIABILITY

This thesis satisfies 60 credits towards an M.Sc. degree in Computational Engineering in the Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, School of Engineering and Natural Sciences, University of Iceland

Email address: bjt9@hi.is

Abstract

Study Objectives: Upper airway resistance during sleep has been shown to cause daytime fatigue. Esophageal pressure measurement is the gold standard for measuring upper airway resistance. The literature documents breath-by-breath changes in resistance. Inconsistencies abound, however, between scoring rules for identifying which changes are considered pathologic and labeled increased upper airway resistance events. Individual scoring rules are underspecified. The prevalence of the events varies between surveys from being rare to being more common than apneas and hypopneas combined. The choice of hypopnea definition cannot explain the variability. The aim of this study was to quantify how reliably increased upper airway resistance events can be scored.

Methods: Fifteen different scoring rules were employed for scoring increased upper airway resistance events in 26 polysomnograms, irrespective of cortical arousal. Event frequencies and second-by-second agreement were tabulated. Two scoring rules were scored visually by professional somnologists. The others were scored programmatically.

Results: Depending on the scoring rule employed, the hourly event rate varied from 0.18/h to 103/h. The ranking of polysomnograms by event frequency differed depending on the scoring rule employed, not least depending on whether events were required to start with lower than baseline pressures. The mean Kendall's τ was 25%; a low, but positive, correlation between scoring rules.

Conclusions: The inconsistency between scoring rules, and interpretations thereof, was very high. Increased upper airway resistance events cannot be reliably scored without a robust standard scoring rule. It remains to be studied which scoring rules, if any, correlate with arousals, daytime sleepiness, hypotension, or hypertension.

Keywords: sleep-disordered breathing, esophageal pressure, upper airway resistance syndrome, obstructive sleep apnea, respiratory effort

Statement of significance

Novel: Inconsistencies between previous studies have raised questions regarding their comparability. Differences in the scoring rules employed for identifying increased upper airway resistance events are common. This study is the first to quantify the implications empirically.

Innovative: Previous studies comparing scoring rules for identifying sleep-disordered breathing only compared the visual identification of events nominally in accordance with defined scoring rules. This study compares both the visual and programmatic identification of events. This eliminates intra-rater and inter-rater reliability issues, and codifies the algorithms implementing the scoring rules.

Important: This study found a plausible explanation for inconsistent results of epidemiological studies. Inconsistent scoring rules can explain the inconsistencies in event rates between previous studies.

INTRODUCTION

Apneas, hypopneas, and respiratory effort-related arousals (RERAs) are the most common respiratory disturbances during sleep. In patients who are alert and not sleepy during the day, the International Classification of Sleep Disorders, 3rd edition, recommends a diagnosis of obstructive sleep apnea syndrome if the average frequency of respiratory disturbance reaches 15 events/h.[1]

To be a clinically useful diagnostic test, polysomnography (PSG) must be able to predict some sequela or whether an intervention will be effective. There is some evidence that increased upper airway resistance (IUAR) predicts both sequela and the effectiveness of treatment for insomnia: IUAR during sleep induces progressively increasing respiratory effort and sleep disruption.[2-4] For women with IUAR and insomnia, various oronasal treatments reduce respiratory effort, actigraphy activity, and daytime fatigue.[3]

To be able to reliably confirm or exclude a diagnosis, PSG needs to yield reliable results. A RERA starts with IUAR and ends with arousal. It is not clear from the current literature if the scoring of IUAR events is reliable. Airway resistance cannot be precisely measured during sleep, so respiratory effort is measured instead. In previous research, different sensors have been used to measure respiratory effort. Esophageal manometry, a measure of intrathoracic pressure, has been referred to as the "gold standard" since respiratory effort is mediated by pressure fluctuations.[5] Inspiratory effort takes the form of respiratory muscles contracting, expanding the thorax and abdomen, thereby lowering intrathoracic pressure.[25] If the airway is open, air is immediately drawn into the lungs, normalizing the intrathoracic pressure. Although baseline respiratory effort differs across individuals and sleep stages,[38] scoring rules of incremental, breath-by-breath changes in inspiratory effort have been used to define and identify IUAR. Breath-by-breath increases in respiratory effort are referred to as a P_{es} *crescendo* when measured by esophageal manometry. According to Reese, et al., a scoring rule for identifying an IUAR event had yet been fully standardized in 1999.[2] In 1999, a task force of the American Academy of Sleep Medicine (AASM) came to the *Chicago consensus* of requiring an event to be terminated by "an arousal with resumption of more normal pressures."[6] In 2012, the AASM published a wider definition of RERA, in *Rules for Scoring Respiratory Events in Sleep,* where resumption of more normal pressures was not required.[7] These are taken to implicitly define IUAR the same way as RERA but without the requirement of a cortical arousal. Scoring rules for identifying IUAR events have varied across publications, even differing between publications with shared authorship. Scoring rules differ as to whether they stipulate a threshold of respiratory effort which is exceeded in every IUAR, either relative to some baseline effort or effortless breathing; minimum and maximum event duration, or a breath-by-breath pattern in respiratory effort (see Appendix 1). In a 1999 study, Exar and Collop discussed papers by Berg et al.,[8] Lofaso et al.,[9] and others. While maintaining that all papers examined the subject of a single syndrome, termed upper airway resistance syndrome (UARS), Exar and

Collop cautioned that it was unclear whether the same patients would fit the diagnostic criteria of all papers.[10]

Epidemiologic surveys have yielded a high spread of estimates of the ratio of RERAs to apneas and hypopneas. Anywhere between 5.7% and 55.5% of respiratory disturbances have been classified as RERAs in diagnostic surveys.[11] A previously postulated explanation was that this spread stemmed from inconsistency as to whether a restriction in airflow without an oxygen desaturation counted as a hypopnea, IUAR, or neither.[12,13] Inconsistencies in the definition of a hypopnea do, however, not explain the high RERA ratio spread in epidemiologic studies.[12] The admittedly small study that classified only 5.7% of respiratory disturbances as RERAs used the most conservative definition of a hypopnea recommended by the AASM, requiring a 4% drop in oxygen saturation in each hypopnea.[14]

Unreliable identification of RERAs is a plausible explanation for this high spread. The identification of a RERA builds on the identification of an IUAR event and associating it with an arousal. Future investigations into the reliability of associating an IUAR event with an arousal are warranted, but outside of the scope of this study. The reliability of scoring arousals in electroencephalograms has been studied,[15] and found to be variable but moderate on average.[16]

In case of unreliable RERA identification, it is conceivable that diagnostic thresholds can be adjusted to correct for any differences in the number of RERAs found, enabling reliable diagnosis. This is possible if, despite unreliability in scoring the exact number of RERAs in a polysomnogram (PSG), the PSG can be reliably classified by whether it has relatively many or relatively few RERAs. Promisingly, Ayappa et al. found no daytime sleepiness in subjects with fewer than 15 events/h. [17] It is also conceivable that there are two kinds of RERAs: those who are reliably identified, and those who are not. Arousal has been used to validate an ambiguous, poorly defined IUAR,[17] implying that reliability of IUAR alone is perceived to be low. Reliability in the timing of IUAR is necessary, but not sufficient, for the reliable diagnosis of REM-dominant and supine-dominant obstructive sleep apnea syndrome. Factoring the timing of events into interrater reliability has been hampered by a lack of computerization.[17] The state of the art was advanced by computing reliability on a second-by-second basis.

The aim of this study was to quantify the reliability of identifying IUAR using the "gold standard" technology, esophageal manometry, and investigate how reliability could be improved.

METHOD

Polysomnograms

A total of 31 participants were studied. Some participants had a pre-existing diagnosis of obstructive sleep apnea (OSA), some were on a waiting list for an OSA diagnostic test, and others had no known sleep issues. The mean body mass index of the study population, after exclusions, was 30 kg/m^2 (range: 22–49 kg/m^2, standard deviation [SD]: 6 kg/m^2), mean age was 47 years (range: 20–69 years, SD: 13 years), mean apnea-hypopnea index was 10/h (range: 0–35/h, SD: 10/h), the mean oxygen desaturation index was 9/h (range: 0–32/h; SD: 10/h), the mean arousal index was 12/h (range: 4–31/h, SD: 7/h), and the mean Epworth sleepiness score was 10 (range 0–21, SD 5).

Respiratory effort was measured using esophageal manometry. Manometry was performed using a catheter (Gaeltec Devices Ltd., Dunvegan, UK) that was threaded through the nose and to within 10 cm of the lower esophageal sphincter (LES). The catheter included four pressure sensors placed 5 cm apart, starting with one sensor at the tip of the catheter, proximal to the LES. Ambient air pressure was not measured, nor were the pressure sensors calibrated.

Electrooculo- and -encephalography were performed. Electromyography was based on the voltage sensors on the chest, chin, and tibialis anteriors. Airflow was measured using a nasal cannula-transducer system and a thermistor. Audio, body position, and movements were recorded. Thorax and abdominal movements were measured using respiratory inductance plethysmography belts. Oxygen saturation and pulse were measured using finger photoplethysmography. All signals were recorded using an A1 device and monitored with

Noxturnal 4.3 software (Nox Medical, Reykjavik, Iceland). Sensors were connected to the A1 recording device using a Bluegiga Bluetooth system (Silicon Labs, Austin, TX, USA).[5] PSG with incomplete electroencephalograms or esophageal manometry with strong cardiac interference were excluded, leaving 26 polysomnograms. The cardiac signal was not subtracted from the esophageal signal. All wakeful epochs were excluded from analysis.

The anonymized PSGs underlying this article will be shared on request to the corresponding author with the permission of the custodian of the data, Þórarinn Gíslason. Participants consented to the polysomnography. The PSG was performed as approved by the National Bioethics Committee (approval VSN-14-080) and Landspítali, the National University Hospital of Iceland.[5] The custodian of the data approved this secondary analysis. The Online Supplement contains software source code and breath-by-breath information on inspiratory effort and identified IUARs along with additional figures and statistics.

IUAR events

In this study, IUAR events were scored independently of the presence of apneas, hypopneas, and arousals. An IUAR event could thus occur concomitant or without an arousal, hypopnea, or apnea. In this study, the presence of an apnea or hypopnea never precluded the presence of an IUAR event.

Breath-by-breath increases in respiratory effort as measured by esophageal manometry are referred to as a P_{es} crescendo.[7] If a P_{es} crescendo was followed by respiratory effort at baseline level, the event was said to have been terminated by P_{es} reversal.

In addition to the presence of a P_{es} reversal, studied criteria included minimum duration, the requirement for respiratory effort to surpass and stay above baseline, and for the effort to abruptly return to below baseline.

Fifteen respiratory effort scoring rules were chosen, scored in 31 PSGs, and then compared to each other. They were chosen in cooperation with a certified expert somnologist. The scoring rules were variously similar to, or valid interpretations of, AASM recommended scoring rules or scoring rules seen in the literature, as detailed in Table 1 and Appendix 1. None of scoring rules used a cut-off at an absolute pressure nor a fixed offset from ambient pressure.

Table 1. *The scoring rules varied in the minimum duration of an event and other criteria.*

Minimum duration	**Family**		
	Pes crescendo	**+ Pes reversal**	**& above baseline**
2 breaths	Cres2[Stoohs]	Cresrev2[Masa]	Cresbaserev2
10 s	Cres10s[AASM]	Technologist 1[Chicago]	*[not studied]*
3 breaths	Cres3	Cresrev3	Cresbaserev3[Poyares] Technologist 2[Poyares]
4 breaths	Cres4	Cresrev4[Guilleminault]	Cresbaserev4
5 breaths	Cres5	Cresrev5	Cresbaserev5

Note that the scoring rules used by sleep technologists differed slightly from the scoring rules used in programmatic scoring. Cres stands for P_{es} crescendo. Cresrev stands for P_{es} crescendo followed by an abrupt reversal. Cresbase stands for P_{es} crescendo surpassing baseline effort followed by an abrupt reversal.

Stoohs) An interpretation of a scoring rule described by Stoohs et al. in 1993[33]
Masa) An interpretation of a scoring rule described by Masa et al. in 2003,[49] comparable to Ayappa et al., 2000[17]
AASM) Complies with the Rules for Scoring Respiratory Events from 2012[7]
Chicago) Abides by the Chicago consensus[6]
Poyares) A similar scoring rule was described by Poyares et al. in 2002 but with a different reversal[42]
Guilleminault) An interpretation of a scoring rule described by Guilleminault et al. in September 2001[18]

Table 2. *Four sleep technologists contributed to this study.*

Technologist 1	Technologist 2	Technologist 3	Technologist 4
Professional sleep technologist who performed PSG, staged sleep and wake, and identified IUAR events in sleep based on a scoring rule she invented.	Certified expert sleep technologist who identified IUAR events, based on a new scoring rule based on a programmed scoring rule.	Certified expert sleep technologist whose sleep-wake staging was used by sleep technologist 2 and all programmed scoring rules.	Certified expert somnologist consulted during the selection of the scoring rules for identifying IUAR events.

PSG stands for polysomnography. IUAR stands for Increased Upper Airway Resistance.

Figure 1.

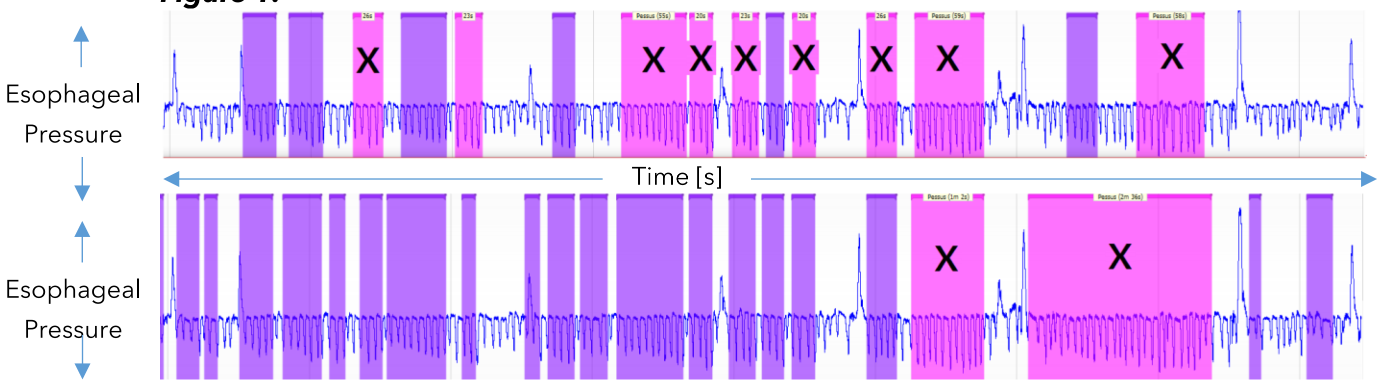


Figure 1, upper: *P_{es} crescendo IUAR (purple) and P_{es} sustained events (pink with an 'x') as identified by sleep technologist 2. The P_{es} sustained events were disregarded in this study.*
Figure 1, lower: *P_{es} crescendo IUAR (purple) and P_{es} sustained events (pink with an 'x') as identified by sleep technologist 2. The P_{es} sustained events were disregarded in this study.*

Visually identified respiratory effort events

Sleep technologist 1 had identified IUAR events described as "progressive, breath-by-breath, more negative inspiratory waveform, lasting at least 10 s and not associated with oxygen desaturation," terminated by "an abrupt drop in respiratory effort, indicated by a less negative end inspiratory waveform, after a sequence of variations in respiratory efforts."[5] This respiratory effort scoring rule was based on a study by Guilleminault et al.[18] Sleep technologist 2, who was not blinded to the aforementioned scoring, identified IUAR events that lasted for at least three breaths with a crescendo in respiratory effort, with pressure under baseline, and terminated by an abrupt return to baseline. This scoring rule was based on the programmed scoring rule Cresbaserev3, defined below. Both sleep technologists had access to all PSG channels, including four esophageal pressure tracings, during scoring. See Table 2 for further information on the sleep technologists. Each sleep technologist additionally identified events called P_{es} sustained.[5] These events were not considered IUAR events. See Figure 1 for examples of both kinds of events.

Programmatically identified respiratory effort events

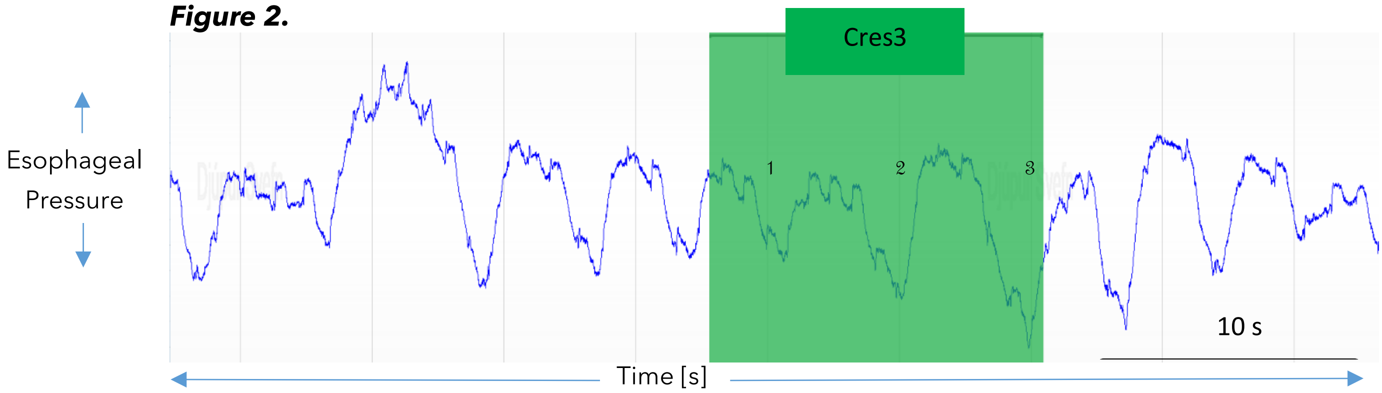
An additional 13 scoring rules were programmed in Haskell (Haskell.org, Inc, New York, NY, USA). Programmatic scoring used only abdominal movements and esophageal manometry. Each PSG was segmented into *attempted breaths* by passing the abdomen respiratory inductance plethysmogram to the proprietary software package Nox Reader library (Nox Medical, Reykjavík, Iceland). For each PSG, the pressure signal from one pressure sensor was chosen as the input for all 13 programmed scoring rules. The criteria used to choose among the four pressure tracings in each case were negative pressure, plausible pressure swings, and the absence of a heartbeat component. Enumerating pressure sensors from the mouth towards the abdomen, the first pressure sensor was never chosen as the input for the programmatic scoring of IUAR. For four PSGs, the second sensor was chosen. For 21 polysomnograms, the third sensor was chosen. For the remaining PSGs, the fourth sensor was chosen. Python 3.5.6 (Python Software Foundation, Beaverton, OR, USA) was used to extract the peak inspiratory effort of the first few attempted breaths and those attempted breaths who did not reach a pressure over four times the average expiratory pressure.

The scoring rules were categorized in three families. The programmatic scoring rules within each class differed only in minimum duration. The first five programmatic scoring rules belonged to the class *cres*, short for P_{es} crescendo. See Figure 2 for an example of a cres event. One of the cres scoring rules is an interpretation of AASM recommendations from 2012. The

2012 AASM recommendation was to identify any "sequence of breaths lasting at least 10 seconds characterized by increasing respiratory effort."[7] The next four respiratory effort scoring rules belonged to the second class: *cresrev*, short for P_{es} crescendo followed by P_{es} reversal, where P_{es} reversal is a breath that is less laborious than the average breath since the last event. See Figure 3 for an example event. They were similar to a Chicago consensus scoring rule from 1999,[6] but with different minimum durations. As advised by a fourth certified expert somnologist, the last four respiratory effort scoring rules, *cresbase,* additionally stipulated that during each breath in an event of IUAR, the inspiratory effort had to surpass the average of the peak-inspiratory efforts of breaths since the last event. See Figure 4 for an example. This was done to preclude an unusually effortless breath followed by a steady return to baseline effort being scored as an IUAR event.

The scoring rule *Cres10* stipulated a minimum duration of 10 seconds. A minimum of two consecutive breaths is stipulated for a *Cres2* event to be identified, each with a higher peak inspiratory effort than that of the previous breath. *Cres3* was defined as three or more consecutive breaths, each with a higher peak inspiratory effort than that of the previous breath. *Cres4* was defined as four or more consecutive such breaths and *cres5* as five or more.

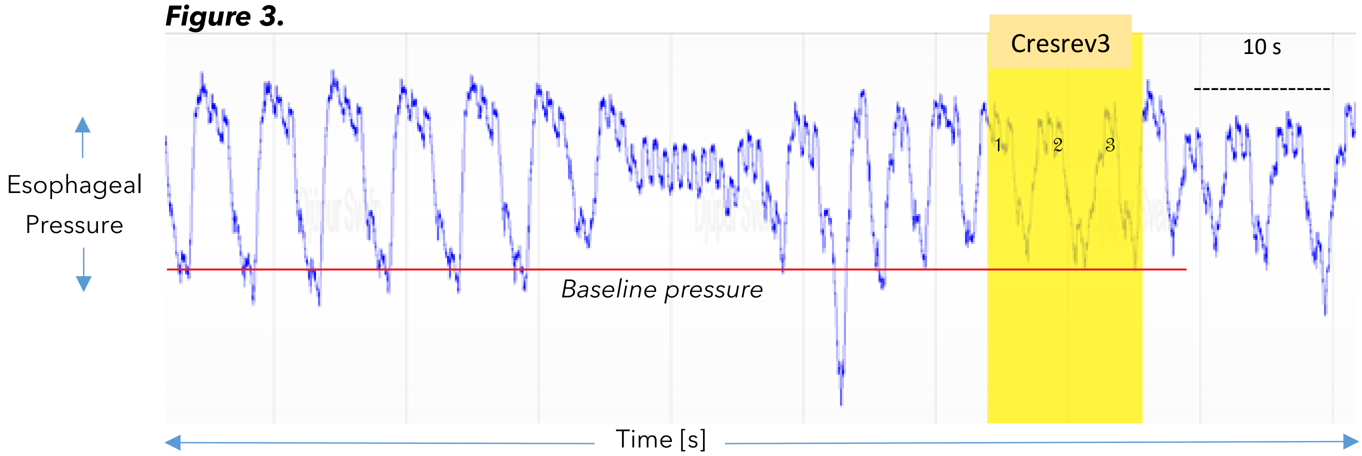
Figure 2.



Cres3 matches a crescendo in respiratory effort–even if a decrescendo, and not an abrupt return to baseline, follows–as long as the inspiratory effort increases for at least three consecutive breaths. Cres3 stands for P_{es} crescendo for at least three breaths. In this case the IUAR happens to last for precisely three breaths.

Cresrev2 was defined as two or more consecutive breaths, each with a higher peak inspiratory effort than that of the previous breath, with the event followed immediately by a breath with below baseline peak inspiratory effort. The baseline is the average peak inspiratory effort between since last event. *Cresrev3*, *Cresrev4*, and *Cresrev5* were defined analogously, with a minimum duration of 3, 4, and 5 breaths, respectively. Respiratory effort was estimated as the unitless inverse of unitless esophageal pressure.

Figure 3.

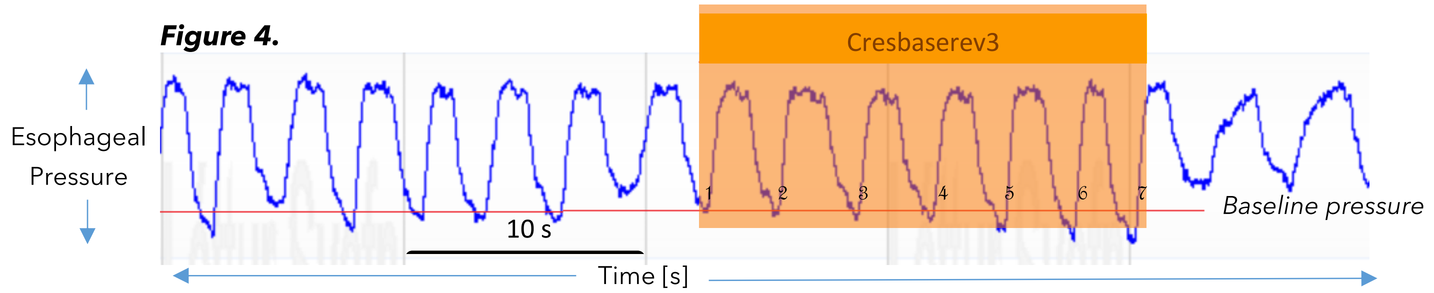


Cresrev3 matches a crescendo in respiratory effort followed by below baseline respiratory effort breath. Note how in this example the respiratory effort of all three breaths in the event happened to be close to or below baseline. Cresrev3 stands for P_{es} crescendo for at least three breaths followed a P_{es} reversal.

Cresbaserev2 was defined as two or more consecutive breaths, each with above baseline peak inspiratory effort that is also higher than the peak inspiratory effort of the previous breath, followed immediately by a breath with a below baseline peak inspiratory effort. *Cresbaserev3*, *Cresbaserev4*, and *Cresbaserev5* were defined analogously, with minimum durations of 3, 4, and 5 breaths, respectively.

The Online Supplement contains the computer program source code of the implementation of the algorithms for programmatic identification of IUAR events.

Figure 4.



Cresbaserev3 matches a crescendo in inspiratory effort in breaths with above baseline peak inspiratory effort, but only if immediately followed by a breath with below baseline inspiratory effort. Cresbaserev3 stands for P_{es} crescendo for at least three breaths, which all surpass baseline inspiratory effort, followed a P_{es} reversal. Note how this example happens to be longer than three breaths.

Analysis

Both pre-existing and bespoke software was used in this study. The bespoke software has not yet seen clinical or previous research use. To minimize the risk of software defects, the most complicated calculations and data manipulations were tested using both manually defined unit tests and randomized property tests.[19] Some, but not all, of the bespoke software was reviewed by the development team. To maximize numerical stability, calculations were performed using integers and rational numbers.

For each scoring rule, all IUAR events in the study population were identified. No scoring rule for identifying events was taken as authoritative.

Event rate

To discern whether the same number of events was identified, irrespective of the scoring rule used, the average event rate, across the study population, was tabulated by scoring rule. The proportion of the PSGs with more than 15 IUAR events/h was also listed.

Rank by event rate

To gauge if adjusting the diagnostic threshold depending on the scoring rule could yield a reliable (consistent) diagnosis, the PSGs were ranked by the hourly rate of events. Each scoring rule yielded a particular ranking of the PSGs. The rankings were compared pairwise using Kendall's τ. The probability of two randomly chosen scoring rules yielding a ranking more similar than not was given by the number of positive Kendall's τs divided by the total number of Kendall's τs computed. Similarly, the probability of two randomly chosen scoring rules yielding a ranking less similar than not was given by the number of negative Kendall's τs divided by the total number of Kendall's τs computed.

Timing and duration of events

To assess whether there existed IUAR events that were reliably identified irrespective of the scoring rule used, the polysomnograms were analyzed second-by-second. The seconds which were not IUAR events by any scoring rule were tallied. The seconds which belonged to IUAR events according to any single scoring rule, but not according to any other scoring scoring rule, were tallied. Seconds which exactly two scoring rules agreed belonged to an IUAR event were tallied—summing across all pairs of scoring rules that ever agreed. Seconds which exactly three scoring rules agreed belonged to an IUAR event were tallied. Seconds which exactly four scoring rules agreed belonged to an IUAR event were tallied. Seconds which five or more scoring rules agreed belonged to an IUAR event were tallied.

Agreement between pairs of scoring rules was quantified on a second-by-second basis. Jaccard indices and overlap coefficients[20] were computed. An overlap coefficient is equivalent to either the recall or the precision of one scoring rule as an approximation of the other, whichever is greater, as proven in Appendix 3.

RESULTS

Out of 31 participants (13 females), five (three females) were excluded, and 26 (10 females) were included in this study.

Event rate

The frequency of IUAR events varied by over three orders of magnitude by the scoring rule used to identify events. This implies that the intraclass correlation was low. Sleep technologist 1 identified 16 events/h on average, but sleep technologist 2 identified 9 events/h. Sleep technologist 1 reported between two and three times as many patients as having more than 15 events/h than did sleep technologist 2 (see Table 3).

Table 3. *The hourly rate of IUAR events, and proportion of study population with ≥15 IUAR events/h, according to two sleep technologists.*

Scoring rule	Event rate [1/h]	≥15 events/h [%]
Cres2	103	100
Cres3	39	100
Cres10sec	48	100
Cresrev2	33	100
Technologist 1	**16**	**38**
Cresbaserev2	14	42
Cres4	14	23
Cresrev3	11	8
Technologist 2	**9.0**	**15**
Cres5	4.7	0
Cresrev4	3.5	0
Cresbaserev3	2.5	0
Cresrev5	1.3	0
Cresbaserev4	0.49	0
Cresbaserev5	0.18	0

Scoring an average of 16 events/h, sleep technologist 1 found 38% of the participants to have 15 or more IUAR events/h. Sleep technologist 2 found 15% of the participants to have 15 or more events/h.

Cres stands for P_{es} crescendo. Cresrev stands for P_{es} crescendo followed by an abrupt reversal. Cresbaserev stands for P_{es} crescendo surpassing baseline effort followed by an abrupt reversal.

Rank by event rate

When the PSGs were ranked by event rate by each scoring rule, and the rankings compared pairwise using Kendall's τ, over 90% of all 182 pairs of rankings were more similar than a random shuffle (had a positive Kendall's τ). Fifteen pairs of scoring rules ranked the patients less similarly than expected from a random shuffle (Kendall's rank coefficient negative). On average, the Kendall's rank correlation coefficient, τ, was 25%. Technologist 1 and 2 ranked the PSGs more similarly to Cres4, with τ 48% and 45%, respectively, than to each other. The τ between the manual scoring rules used for visually identifying events was 39%. Technologist 2 ranked the PSGs most similarly to Cres5, with a τ of 52% (see Table 4). No matter which two scoring rules were compared, the resulting Kendall's rank correlation coefficients was consistently lower than 75%.

Table 4. *Kendall's τ between two sleep technologists and the cres family of scoring rules.*

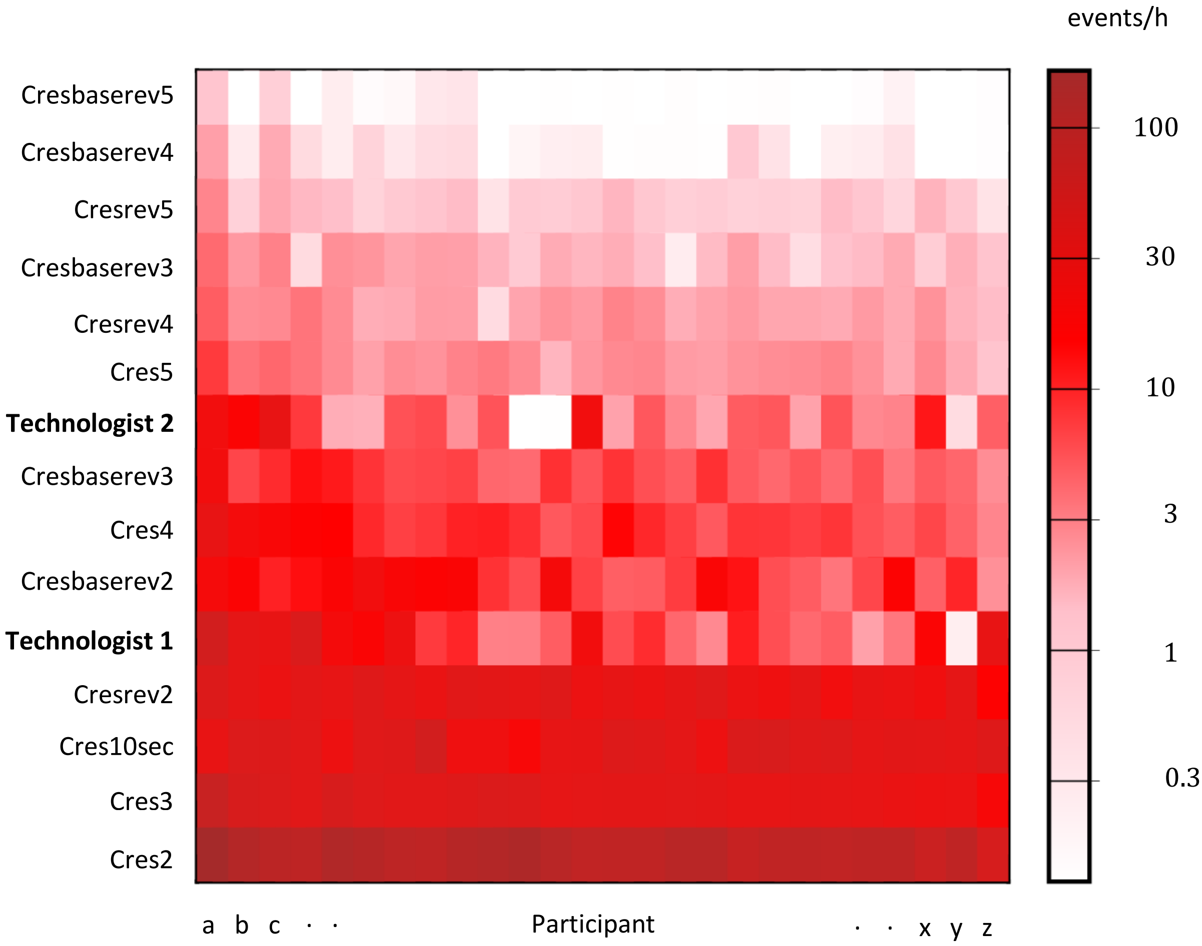
Technologist 1	Technologist 2	Cres2	Cres3	Cres4	Cres5	Cres10sec	Scoring rule
-	**39%**	4%	35%	48%	32%	22%	**Technologist 1**
	-	-2%	28%	45%	52%	27%	**Technologist 2**
		-	65%	40%	25%	-12%	**Cres2**
			-	72%	52%	8%	**Cres3**
				-	68%	24%	**Cres4**
					-	14%	**Cres5**
						-	**Cres10sec**

Each cell contains the Kendall's rank correlation coefficient, τ, of how the polysomnograms were ranked by the score of events matching the scoring rule corresponding to the row as compared to how they were ranked using the scoring rule corresponding to the column. The expected value of τ is zero for uncorrelated classifiers. Kendall's τ can range from -100% (perfect anticorrelation) to 100% (perfect correlation).

Cres stands for P_{es} crescendo.

Each PSGs had multiple different IUAR event rates, depending on the scoring rule, as can be seen in Figure 5. Sleep technologist 1 found, on average, more IUAR events than sleep technologist 2. This does not at all imply that this is the case across all participants.

Figure 5.



Cres stands for P_{es} crescendo. Cresrev stands for P_{es} crescendo followed by an abrupt reversal. Cresbaserev stands for P_{es} crescendo surpassing baseline effort followed by an abrupt reversal. The hourly rate of events identified by scoring rule and polysomnogram. Each row represents a scoring rule. Each column represents one polysomnogram of a unique participant. The shade of each tile reflects the hourly rate of IUAR events. Specifically, the darkness of each tile indicates the number of events matching the corresponding scoring rule per hour of the corresponding polysomnogram. Scoring rules (rows) are ordered as in Table 3. Similarly, PSGs (columns) are ordered by their event rate averaged over all scoring rules. This layout is inspired by Drinnan et al., 1998.[15]

Timing and duration of events

The timing of the studied IUAR events was more similar across scoring rules than would have been expected by chance alone.

Over 90% of the events identified by Cresbaserev5, of which there were 115, were identified by both Cresbaserev2 and Cresbaserev4, on a second-by-second basis. The overlap coefficient between Cresbaserev2 and Cresbaserev4 themselves, however, was only 68%.

On average, there was a total of one hour of each PSG considered uneventful by 14 scoring rules but eventful by one scoring rule, albeit not always the same scoring rule, neither here nor in the rest of the paragraph. A further half an hour per PSG, on average, was considered uneventful by 13 scoring rules but eventful by two. Additional 24 minutes per PSG, on average, were considered uneventful by 12 scoring rules but eventful by three scoring rules. A further 17 minutes per PSG, on average, were considered uneventful by 11 scoring rules but eventful by four scoring rules. A further 15 minutes per PSG were considered uneventful by 10 scoring rules but eventful by five scoring rules. On average, just over 21 minutes per PSG were considered a part of IUAR events by more than five scoring rules.

Sleep technologist 2 found fewer, but longer, events than sleep technologist 1. Their mean event durations were 46 s and 27 s. The sleep technologists also disagreed on the timing of IUAR events the majority of the time, as can be seen in Figure 6. The overlap coefficient between the scorings of the sleep technologists was 34%.

Out of 26 PSGs totaling 166.7 hours of sleep time, there were in total just under four minutes of sleep that were identified as part of IUAR events by all scoring rules studied. Close to three additional minutes of sleep matched all programmatic scoring rules but were not visually identified by either technologist as IUAR.

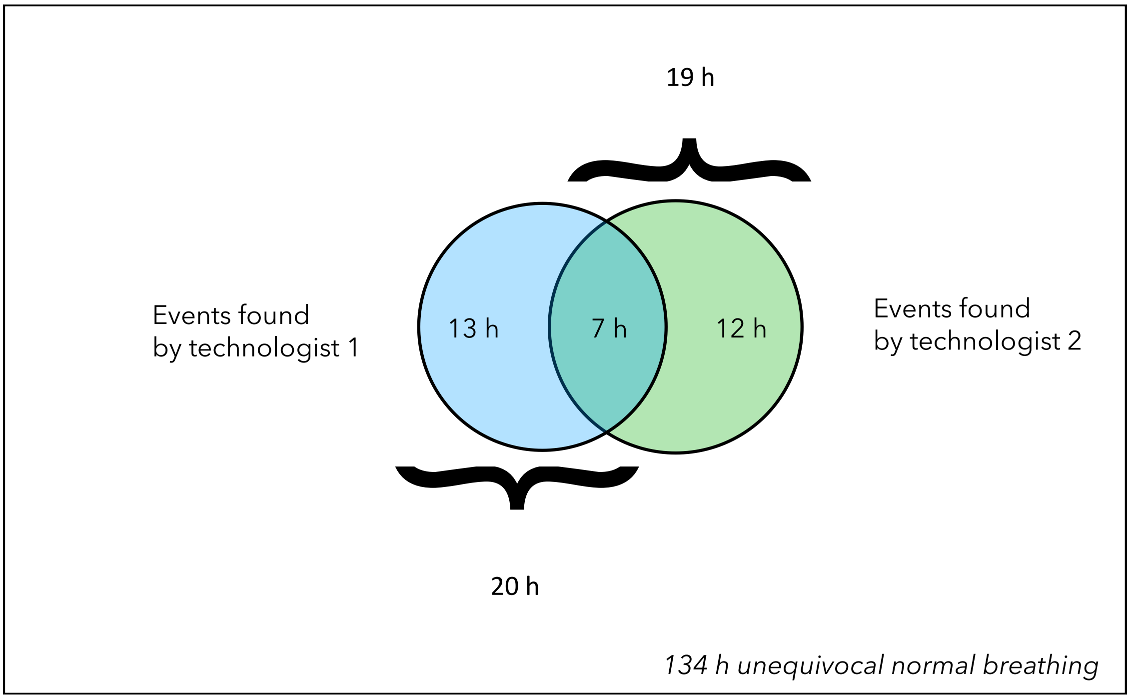
Just under five minutes were identified as part of IUAR events by all programmatic scoring rules and by technologist 1. Five minutes and 41 seconds were identified as part of IUAR events by sleep technologist 1 and 2 and all programmatic scoring rules except Cresbaserev5.

Just under 90% of Cres3 events, by time, lasted 10 seconds or longer, but less than half of the cres10sec events, by time, were Cres3 events. The difference between Cresrev3 and Cresbase3 was even greater, as can be seen in Figure 7.

As an example of a conflict between scoring rules, consider several consecutive epochs in one of the PSGs. The first two each contained an event according to sleep technologist 2 but containing no Cresbaserev3 event. The third epoch contained a Cresbaserev3 event, but no event according to sleep technologist 2. The fourth epoch contained no Cresbaserev3 event but did contain an event according to sleep technologist 2. For every scoring rule, plenty of such examples can be found between that scoring rule and most, if not all, other scoring rules.

The overlap coefficient of 51% between Cresrev2 and Cresrev5 means that almost half of the IUAR events, weighted by duration, identified by Cresrev5 were not identified as such by Cresrev2. The overlap coefficient of only 81% between Cresrev3 and Cresbaserev3 shows that adding a criterion to a scoring rule does not simply make it stricter or less sensitive but can lead to breaths being newly identified as IUAR events. (Figure 7(b))

Figure 6.



A third of the breaths belonging to an IUAR event according to technologist 1 also belonged to an IUAR event according to technologist 2. The technologists agreed on 7 h of IUAR (Figure 6). In contrast, two random processes classifying on average 20 h as sleep-disordered breathing would have been expected to agree, on average, on 3 h of IUAR.

Figure 7(a).

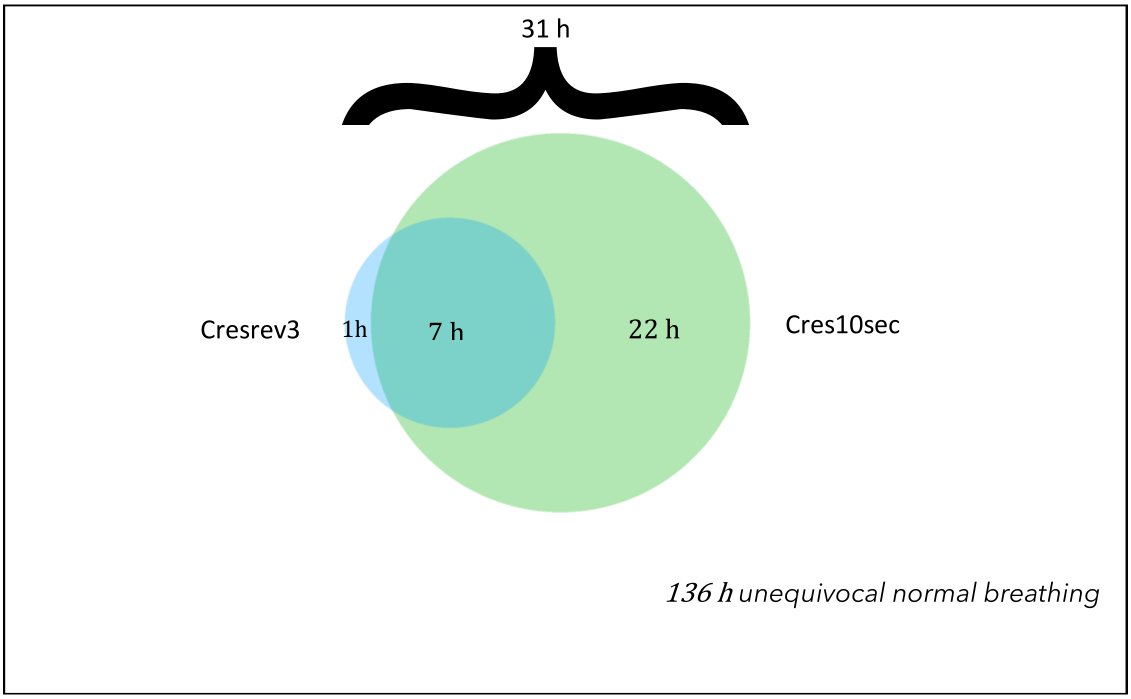


Figure 7(b).

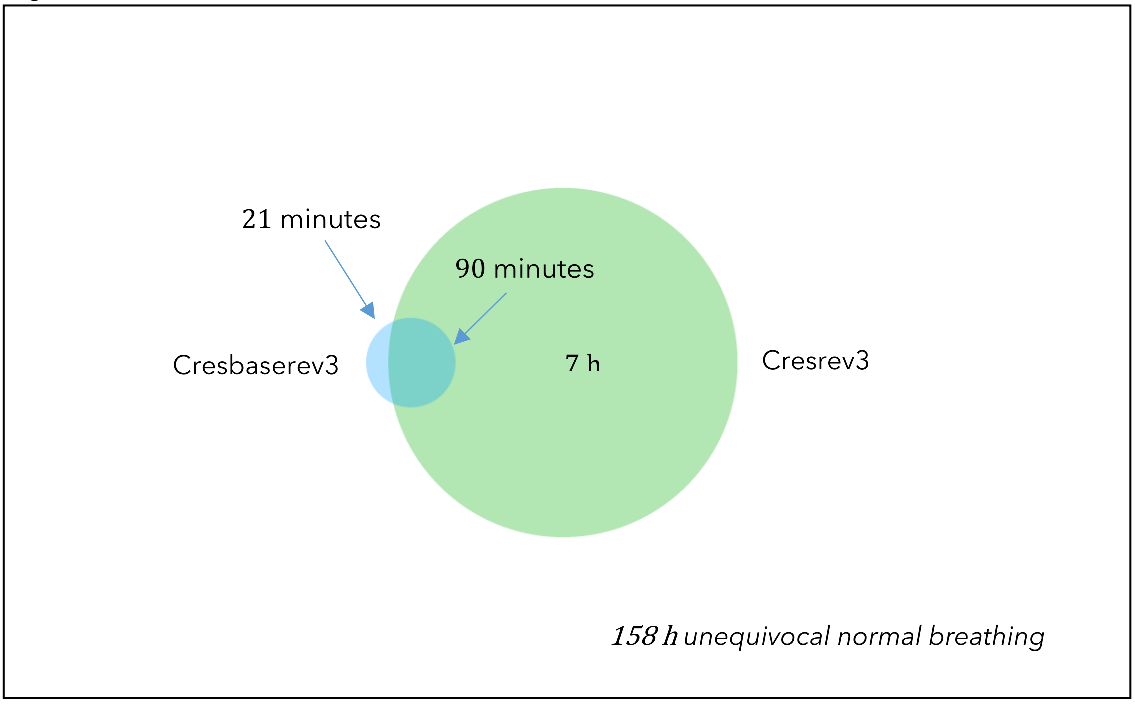


Figure 7(a): Out of the 30 h of cres10sec events identified in the study population, and the 8 h of the Cresrev3 events identified, 7 h were simultaneous cres10sec and Cresrev3 events.

Cresrev3 stands for Pes crescendo for at least three breaths followed a P_{es} reversal.

Figure 7(b): Adding the requirement of effort during an IUAR surpassing baseline decimates the proportion of sleep identified as having IUAR, but also adds new events. Cres10sec stands for P_{es} crescendo for at least 10 seconds.

Cresbaserev3 stands for Pes crescendo for at least three breaths, all of which surpassed baseline, followed by a P_{es} reversal.

DISCUSSION

The scoring rules studied yielded very different numbers of events/h in each PSG (as shown in Table 3 and Figure 5). The scoring rules correlate only weakly in the sense of identifying the same events.

The literature does not quantify expectations of similarity between scoring rules. Refer to a test-retest reliability study of St. Mary's Sleep Questionnaire that used Kendall's rank correlation coefficient. [21] No two scoring rules ranked PSGs as similarly as any individual question in the St. Mary's Sleep Questionnaire ranked patients, let alone the questionnaire as a whole. It is impossible to rank the PSGs in a manner remotely consistent with all scoring rules.

The difference between scoring rules can also be seen by comparing the colors between rows (scoring rules) in Figure 5. If all scoring rules ranked PSGs identically by event rate, every row (scoring rule) in Figure 5 would contain the same progression of colors, from darkest (highest event rate) to the left to lightest (lowest event rate) to the right, because the columns are ordered by the average event rate of the corresponding PSG. To the contrary, Figure 5 does not display such a pattern.

Thus, the scoring of IUAR is not reliable when the rules for scoring IUAR events are allowed to vary. Since the scoring rules for identifying IUAR events vary across studies, the identification of IUAR events cannot be assumed to be reliable across studies. Studies on IUAR can, thus, not be assumed to be comparable. By extension, this casts doubt on the comparability of studies on RERAs and may explain the large inconsistencies in the results on the prevalence of RERAs in the literature. Similarly, AASM recommended rules for scoring IUAR events both change between publications[6,7] and leave ample room for interpretation.

This underlines the importance of clearly stating what scoring rule was used for scoring IUAR events, both when diagnosing a patient with OSA or UARS, and when explaining the methods of research into IUAR or RERA events. For example, while Guilleminault et al. showed in a landmark 2002 study that upper airway resistance can cause daytime fatigue, they did not describe how they defined and identified IUAR events,[3] hampering the reproducibility and translatability of their important findings.

It may be tempting to borrow a suggested consistency remedy from the hypopnea literature. Namely, to transition from a standard diagnostic threshold of 15 events/h to different thresholds for different scoring rules, [22] or even a personalized threshold for every sleep technologist. However, this study demonstrated that scoring rules for identifying IUAR do not agree, not even on which participants have the most IUAR events. Therefore, this remedy, changing the diagnostic threshold, can not be expected to bring about consistent diagnoses.

Scoring rules did not mention apneas, hypopneas, arousals, nor, save for sleep technologist 1, oxygen desaturations. Thus, inconsistencies in the scoring of IUAR were attributed to different interpretations of P_{es} or, in the case of sleep technologist 1, oxygen saturation and sleep stages, rather than interscorer disagreement on hypopneas or arousals.[16,22-23]

Of note were the findings that not only did the different scoring rules find a different number of IUAR events, but both the duration of these events, and their timing, was considerably different between scoring rules. A potential explanation for this might be that most of the scoring rules studied refer to a baseline of previous normal breathing. They are chaotic in the sense that they are sensitive to which previous breathing has been scored as IUAR events, and thus excluded from the baseline of previous normal breathing. Therefore, seemingly minor differences between scoring rules incorporating such a baseline can be magnified by consequential shifts in baselines, leading to completely different segments of the PSGs being scored as IUAR events.

Limitations

The technologists had access to more signals than strictly needed to follow the scoring rules. No measures were taken to verify that they in fact followed the scoring rules they claimed to use. Nor did the technologists review all of the events identified programmatically.

The technologists marked the start and end of events by pointing to a location in a polysomnogram displayed visually using the Noxturnal software program (Nox Medical, Reykjavík, Iceland). This might exaggerate differences in the duration of visually scored events. However, any such exaggeration must have been relatively small, by the following logic. The margin of error can be assumed to be at most one breath per event. A breath usually lasted for just over 3 s. With the mean event durations 27 s and 46 s, the margin of error was less than 8% per average event. The measured disagreement between the technologists was 80% or ten times greater than the margin of error.

The programmatic scoring rules did not consistently delineate breathing attempts during obstructive apneas. They sometimes delineated each breathing attempt but in other instances they merged the obstructive apnea with an adjacent breath. Central apneas were always merged with an adjacent breath.

Conclusion

In conclusion, rules for scoring IUAR events are not reconcilable, let alone interchangeable. Changing the diagnostic threshold of 15 events/h would not bring about consistency. Hopefully, the array of scoring rules precisely defined here and implemented in the supplementary software can serve as a solid base for future studies into whether any of the scoring rules predict arousals,[8] daytime sleepiness,[3] hypotension,[24] or hypertension.[9] Which rules for scoring IUAR events are useful for the diagnosis of sleep-disordered breathing, remains an open question.

Practice points

| ***I**ncreased upper airway resistance event* can refer to an event scored in accordance with any of dozens of scoring rules.

| **A**n event scored in accordance with one of the scoring rules studied is unlikely to be an event according to all the other scoring rules studied.

| **A** polysomnogram can be interpreted as showing the presence of increased upper airway resistance, and its absence, depending on how it is scored.

Technical leap forward

| **A**lgorithms for scoring increased upper airway resistance events are now available in the Haskell programming language. Polysomnography software can be developed to incorporate any of the algorithms for assisted or automated scoring of increased upper airway resistance events.

Research agenda

| **S**tudy how different respiratory effort scoring rules yield different decisions with respect to whether an arousal is respiratory effort-related or spontaneous.

| **S**tudy the epidemiology, sequelae, and pathophysiology of increased upper airway resistance as defined by precisely formulated respiratory effort scoring rules.

| **V**alidate an alternative sensor of respiratory effort against esophageal manometry the same way esophageal manometry was validated against intrapleural manometry during inhalation; plot both measures of respiratory effort using the same units, one on top of the other.

| **B**efore a diagnostic cut-off number of respiratory events per hour can be standardized on as a diagnostic threshold for sleep apnea, a robust scoring rule needs to be standardized on.

REFERENCES

1] American Academy of Sleep Medicine. International classification of sleep disorders, 3rd ed. Darient, IL. American Academy of Sleep Medicine, 2014.

2] Rees K; Kingshott RN; Wraith PK; Douglas NJ. Frequency and significance of increased upper airway resistance during sleep. *Am J. Respir. Crit. Care Med*. 2000;162(4):1210–1214. doi:10.1164/ajrccm.162.4.9908052.

3] Guilleminault C; Palombini L; Poyares D; Chowdhuri S. Chronic insomnia, premenopausal [sic] women and sleep disordered breathing Part 2. Comparison of nondrug treatment trials in normal breathing and UARS post menopausal women complaining of chronic insomnia. *J. Psychosom. Res*. 2002;53(1):617–623.

4] Berry RB; Gleeson K. Respiratory Arousal From Sleep: Mechanisms and Significance. *Sleep*. 1997;20(8):654–675. https://academic.oup.com/sleep/article/20/8/654/2725999. Accessed January 14, 2021.

5] Marta Serwatko. Validation of a new method to assess respiratory effort non-invasively. *Skemman*. https://skemman.is/bitstream/1946/23804/1/Serwatko_Msc_MS_yfirfarid.pdf. Accepted February 18, 2016. Accessed October 25, 2017.

6] An AASM task force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep*. 1999;22(5):667–689. doi:10.1093/sleep/22.5.667. https://academic.oup.com/sleep/article/22/5/667/2726040. Accessed February 15, 2020.

7] Berry RB; Budhiraja R; Gottlieb DJ; et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events, Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med*. 2012;8(5):597–619. doi:10.5664/jcsm.2172.

8] Berg S; Nash S; Cole P; Hoffstein V. Arousals and nocturnal respiration in symptomatic snorers and nonsnorers. *Sleep*. 1997;20(12):1157–1161. doi:10.1093/sleep/20.12.1157.

9] Lofaso F; Coste A; Gilain L; Harf A; Guilleminault C. Sleep fragmentation as a risk factor for hypertension in middle-aged nonapneic snorers. *Chest*. 1996;109(4):896–900. doi:10.1378/chest.109.4.896. https://journal.chestnet.org/article/S0012-3692(15)46019-2/fulltext. Accessed August 28, 2020.

10] Exar EN; Collop NA. The upper airway resistance syndrome. *Chest*. 1999;115(4): 1127–1139.

11] Krakow B; Krakow J; Ulibarri VA; McIver ND. Frequency and accuracy of "RERA" and "RDI" Terms in the Journal of Clinical Sleep Medicine from 2006 through 2012. *J Clin Sleep Med*. 2014;10(2):121–124. doi:105664/jcsm.3432.

12] Douglas NJ. Upper Airway Resistance Syndrome Is Not a Distinct Syndrome. *Am J Respir Crit Care Med*. 2000; 161 (5): 1413–1415.

13] Duce B, Milosavljevic J, Hukins C. The 2012 AASM Respiratory Event Criteria Increase the Incidence of Hypopneas in an Adult Sleep Center Population. *J Clin Sleep Med*. 2015;11(12):1425-1431. doi:10.5664/jcsm.5280

14] Masdeu MJ; Seelall V; Patel AV; Ayappa I; Rapoport DM. Awake measures of nasal resistance and upper airway resistance on CPAP during sleep. *J Clin Sleep Med*. 2011;7(1):31–40.

15] 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(2):358–362.

16] Kuna ST; Benca R; Kushida CA; et al. Agreement in Computer-Assisted Manual Scoring of Polysomnograms across Sleep Centers. *Sleep*. 2013; 36(4):583–589.

17] Ayappa I; Norman RG; Krieger AC; Rosen A; O'Malley RL; Rapoport DM. Non-Invasive detection of Respiratory Effort-Related Arousals (RERAs) by a nasal cannula/pressure transducer system. *Sleep*. 2000;23(6):763–771. doi:10.1093/sleep/23.6.763.34] Stoohs R; Guilleminault C. Snoring during NREM sleep: Respiratory timing, esophageal pressure and EEG arousal. *Respir Physiol*. 1991;85:151–167.

18] Guilleminault C; Poyares D; Palombini L; Koester U; Pelin Z; Black J. Variability of respiratory effort in relation to sleep stages in normal controls and upper airway resistance syndrome patients. *Sleep Med*. 2001;2(5):397–405.

19] Claessen K; Hughes JM. QuickCheck: a lightweight tool for random testing of Haskell programs. *ACM SIGPLAN Notices*. 2011;46(4):53–64. doi:10.1145/1988042.1988046.

20] Raayab; Kubina; Hardy M; et al. Overlap coefficient. *Wikipedia, The Free Encyclopedia*. 2021 (May). https://en.m.wikipedia.org/w/index.php?title=Overlap_coefficient&oldid=1022016130. Accessed August 20, 2021.

21] Ellis BW; Johns MW; Lancaster R; Raptopoulos P; Angelopoulos N; Pries RG. The St. Mary's hospital sleep questionnaire: a study of reliability. *Sleep*. 1981;4(1):93–97.

22] Ruehland WR; Rochford P; O'Donoghue FJ; Pierce RJ; Singh P; Thornton AT. The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. *Sleep*. 2009;32(2):150–157. doi:10.1093/sleep/32.2.150.

23] Norman RG; Ahmed MM; Walsleben JA; Rapoport DM. Detection of respiratory events during npsg: nasal cannula/pressure sensor versus thermistor. *Sleep*. 1997;20(12):1175–1184. https://academic.oup.com/sleep/article/20/12/1175/2750010. Accessed August 16, 2020.

24] Guilleminault C; Faul JL; Stoohs R. Sleep-disordered breathing and hypotension. *Am J Respir Crit Care Med*. 2000;164(7):1442–1246. doi:10.1164/ajrccm.164.7.2011036.

25] Mead J; Gaensler EA. Esophageal and pleural pressures in man, upright and supine. J Appl Physiol. 1959;14(1):81–83. doi:10.1152/jappl.1959.14.1.81. https://journals.physiology.org/doi/pdf/10.1152/jappl.1959.14.1.81. Accessed September 3, 2018.

26] Tvinnereim M; Miljeteig H. pressure recordings—a method for detecting site of upper airway obstruction in obstructive sleep apnea syndrome. *Acta Oto-Laryngologica*. 1992;112(Supp 492):132–140. doi:10.3109/00016489209136832.

27] Tvinnereim M; Mateika S; Cole P; Haight J; Hoffstein V. Diagnosis of obstructive sleep apnea using a portable transducer catheter. *Am J Respir Crit Care Med.* 1995;152(2):775–779. doi:10.1164/ajrccm.152.2.7633741.
28] Kushida CA; Morgenthaler TI; Littner MR, et al. Practice parameters for the indications for polysomnography and related procedures: An update for 2005. *Sleep*. 2005;28(4):499-521.
29] Gould GA; Whyte KF; Rhind MA; et al. Sleep hypopnea syndrome. Am Rev Respir Dis. 1988;137:895–8.
30] Guilleminalt C; Stoohs R; Duncan S. Snoring: daytime sleepiness in regular heavy snorers. *Chest*. 1991;99(1):40–48.
31] Guilleminault C; Stoohs R; Clerk A; Simmons J; Labanowski M. From obstructive sleep apnea syndrome to upper airway resistance syndrome: consistency of daytime sleepiness. *Sleep*. 1992;15(6):13–16. https://academic.oup.com/sleep/article/15/suppl_6/S13/274929.
32] Guilleminault C; Stoohs R; Clerk A; Cetel M; Maistros I. A cause of daytime sleepiness: the upper airway resistance syndrome. *Chest*. 1993;104(September):781–787. doi:10.1378/chest.104.3.781. https://journal.chestnet.org/action/showPdf?pii=S0012-3692(16)38870-5.
33] Stoohs R; Skrobal A; Guilleminault C. Does snoring intensity predict flow limitation or respiratory effort during sleep? *Respir Physiol*. 1993;92(1):27–38.
35] Shiomi T; Stoohs R; Guilleminault C. Aging, respiratory efforts during sleep, and pulsus paradoxus. *Lung*. 1993;171:203–211.
36] Watanabe T; Mikami A; Kumano-Og T; et al. The relationship between esophageal pressure and apnea hypopnea index in obstructive sleep apnea-hypopnea syndrome. *Sleep Res Online*. 2000;3(4):169–172.
37] Loube DI; Gay PC; Strohl KP; Pack AI; White DP; Collop NA. Indications for positive airway pressure treatment of adult obstructive sleep apnea patients: a consensus statement. *Chest*. 1999;115:863–866.

38] Loube DI; Andrada TF; Howard RS. Accuracy of respiratory inductive plethysmography for the diagnosis of upper airway resistance syndrome. *Chest*. 1999;115(5):1333–1337.

39] Loube DI; Andrada TF. Comparison of respiratory polysomnographic parameters in matched cohorts of upper airway resistance and obstructive sleep apnea patients. *Chest*. 1999;115(5):1519–1524. doi:10.1378/chest.115.6.1519.

40] Montserrat JM; Badia JR. Upper airway resistance syndrome. *Sleep Med Rev*. 1999;3(1):5–21. doi:10.1016/S1087-0792(99)90011-4.

41] Black JE; Guilleminault C; Colrain IM; Carrillo O. Upper airway resistance syndrome: central electroencephalographic power and changes in breathing effort. *Am J Respir Crit Care Med.* 2000;162(2): 406–411. doi:10.1164/ajrccm.162.2.9901026.

42] Poyares D; Guilleminault C; Rosa A; Ohayon M; Koester U. Arousal, EEG spectral power and pulse transit time in UARS and mild OSAS subjects. *Clin Neurophysiol*. 2002;113:1598–1606.

43] Guilleminault C; Kim YD; Stoohs R. Upper airway resistance syndrome. *OMF Surg Clin North Am*. 1995;7:201–214.

44] Vandenbussche NL; Overeem S; van Dijk JP; Simons PJ; Pevernagie DA. Assessment of respiratory effort during sleep: Esophageal versus noninvasive monitoring techniques. *Sleep Med Rev*. 2015;24:28–36.

45] Bachour A; Herrala J; Maasilta P. Is there a cost-effective way to diagnose mild sleep-disordered breathing? *Respir Med.* 2002;96(8):586–593.

46] Argod J; Pépin J; Smith RP; Lévy P. Comparison of esophageal pressure with pulse transit time as a measure of respiratory effort for scoring obstructive nonapneic respiratory events. *Am J Respir Crit Care Med*. 2000;162(1):87–93.

47] Chervin RD; Ruzicka DL; Hoban TF. Esophageal pressures, polysomnography, and neurobehavioral outcomes of adenotonsillectomy in children. *Chest*. 2012; 142 (1):101–110.

48] Katz ES, Lutz J, Black C, Marcus CL. Pulse transit time as a measure of arousal and respiratory effort in children with sleep-disordered breathing. *Pediatr Res*. 2003;53(4):580–588.

49] Masa JF, Corral J, Martin MJ, Riesco JA, Sojo A, Hernandez M, et al. Assessment of thoracoabdominal bands to detect respiratory effort-related arousal. *Eur Respir J*. 2003;22(4):661–667.

50] Hutter DA; Holland BK; Ashtyani H. Occult sleep apnea: the dilemma of negative polysomnography in symptomatic patients. *Sleep Med*. 2004;5(5):501–506.

51] Johnson PL; Edwards N; Burgess KR; Sullivan CE. Detection of increased upper airway resistance during overnight polysomnography. *Sleep*. 2005;28(1):85–90. doi:10.1093/sleep/28.1.85. https://academic.oup.com/sleep/article-abstract/28/1/85/2696720. Accessed March 1, 2020

52] Kristo DA; Lettieri CJ; Andrada T; Taylor Y; Eliasson AH. Silent upper airway resistance syndrome: prevalence in a mixed military population. *Chest*. 2005;127(5):1654–1657.

53] Masa JF; Corral J; Teran J; et al. Apnoeic and obstructive nonapneic sleep respiratory events. *Eur Respir J* 2009;34(1):156–161.

Acknowledgements and disclosure

Thanks to Rögnvaldur Sæmundsson, Erna Sif Arnardóttir, Marta Serwatko, Kristín Anna Ólafsdóttir, Tumi Steingrímsson, Pétur Ólafur Aðalgeirsson, Eggert Karl Hafsteinsson, Kristín María Tómasdóttir, Benedikt Þórðarson, Sólon Örn Sævarsson, Jacqueline Clare Mallett, Gunnhildur Halla Ármannsdóttir, and Hannes Kristján Hannesson for support and advice. Thanks to the University Hospital of Iceland Science Fund for grant A-2018-054. There is nothing else to disclose.

SUPPLEMENTS

Online supplement: The source code for the software written for this study is available at https://github.com/bjartur/master/tree/5b8474

The appendices contain data and a proof that was produced as part of this study and is cited in the manuscript.

Appendix 1. *Scoring rules for identifying IUAR by esophageal manometry documented in the literature.*

Article	***Organ***	***Baseline***	***Event characterized by…***	***Minimum duration***	***Event termination***	***Source cited***
Tvinnereim M & Miljeteig H, 1992[26]	Pharynx & upper esophagus	Preceding epoch	Flattening of a cranial pressure transducer while pressure oscillations, at least as negative as before, continue caudally.	10 s (or 8s)	?	None
Tvinnereim M & al., 1995[27]	Pharynx & esophagus	Preceding epoch	"Progressively increasing pressure fluctuations" with three (four) times more negative pressure than baseline (excluding oscillations with period 13s or shorter) indicates a hypopnea (apnea).	10 s	Observed to be abrupt.	Tvinnereim, 1992[26]

American Academy of Sleep Medicine Task Force, 1999[6]	Esophagus	?	Progressively more negative esophageal pressure.	10 s	Sudden change in pressure to a less negative level and an arousal.	? (10s cutoff from Gould, 1988[29])
Kushida & al., 2005[28]	Esophagus	?	Sequence of breaths with increasing respiratory effort leading to an arousal from sleep, as shown by progressively more negative pressure.	10 s	Arousal with resumption of more normal pressures.	AASM Task Force in Chicago, 1999[6]
Guillemin ault & al., 1991[30]	Esopagues	3 epochs preceding an arousal	The ultimate or penultimate breath before an arousal has the most negative P_{es} in 3	None	Less negative P_{es} during and after an arousal with increased airflow.	None

			epochs coinciding with an abrupt decrease in airflow.			
Exar & Collop, 1999[10]	Esophagus	From one arousal to the next	The ultimate or penultimate breath before an arousal has the greater P_{es} nadir than any breath in the baseline and is accompanied by a simultaneous decrease in airflow.	One breath	Arousal and less negative P_{es} and increased airflow during the two breaths after the arousal.	Guilleminault & al., 1991[30]
Guillemin ault & al., 1992[31]	Esopagus	Wakeful breathing and sleep from the last arousal up until the penultimate before the current arousal	The ultimate or penultimate breath before an arousal has the greater P_{es} nadir than any breath in either baseline.	One breath	Arousal and less negative P_{es} and increased airflow during the two breaths after the arousal.	None

Exar & Collop, 1999[10]	Esophagus	30 min of quiet supine wakefulness	The most negative P_{es} nadir of a snoring period (the period from the last arousal) if it is 1SD more negative than the mean baseline P_{es} nadir.	One breath	Arousal followed by a breath with a less negative P_{es} nadir.	Guilleminault & al., 1991[30]
Guilleminault & al., 1993[32]	Esophagus	30 min during quiet supine wakefulness	(Abnormal) increase in negative esophageal pressure.	?	Transient arousal (a breath or two after P_{es} nadir).	None
Stoohs & al., 1993[33]	Esophagus	Quiet, relaxed, supine wakefulness with lights on	Increase in peak-negative inspiratory efforts.	Two breaths	Arousal	Stoohs and Guilleminault, 1991[34]
Shiomi & al., 1993[35]	Esophagus	Presumably atmospheric pressure	Heavy snoring with P_{es} nadir below -10cmH_20.	None	?	None

Berg et al., 1997[8]	8, 12, and 16 cm from the nares and in the middle third of the esophagus	Awake, unobstructed breathing	20% increase in peak-to-peak pressure amplitude	15 s	?	None
Watanabe & al., 2000[36]	Lower third of the esophagus	?	Drop in the P_{es} nadir of more than 135mmH_2O.	None	?	Shiomi &al., 1993[35]
Loube DI & al., March 1999[37]	Esophagus	?	Pattern of progressive negative esophageal pressures	10 s	Change in pressure to a less negative pressure level associated with an arousal.	? (ambiguous due to missing inline citation)
Loube DI & al., May 1999[38]	Esophagus	Wake minimum negative P_{es}	Crescendo changes in P_{es} with pressure exceeding baseline by 50% and dropping below -12cmH_2O.	?	EEG arousal.	AASM Task Force in Chicago, 1999[6]

Loube DI & al., June 1999[39]	Esophagus	Unclear, but a P_{es} nadir was calculated for each epoch and the mean taken by sleep-wake stage	Pattern of progressive negative P_{es}.	10 s	Change in pressure to a less negative pressure level associated with a (micro)arousal	Loube DI et al, May 1999[38]
Montserrat JM; Badia JR, 1999[40]	Esophagus	?	Progressive fast decrease of P_{es}.	?	Abruptly returns to normal, or almost normal, P_{es} after a neurological arousal.	None
Ayappa I & al., 2000[17]	Esophagus	?	Crescendo pattern of negative inspiratory pressure swings (although figure 3[17] shows an increase without crescendo that was identified as a RERA).	10 s	Rapid decrease of the swings to baseline level.	? (ambiguous due to missing inline citation)

Black & al., 2000[41]	Esophagus	The 4s interval 26s–24s before the P_{es} reversal	Consecutively increasing negative inspiratory P_{es} values.	10 s (typically three breaths)	Followed abruptly by a series of breaths of reduced negative P_{es} (in specific, the breath following the final breath of the crescendo had to be less than 75% of the Pes of the previous breath).	Guilleminault & al., 1993[32]
Poyares etal., 2002[42]	Esophagus	Prior breathing	A more negative peak end inspiratory esophageal pressure.	3 successive breaths	Abrupt shift of at least 25% of the peak negative end inspiratory esophageal pressure toward less effort associated with improvement in nasal flow.	Guilleminault & al., 1995[43]
Vandenbu ssche etal.[44]	Esophagus	?	One or two breath increases in P_{es}.	One breath	P_{es} reversal	Guilleminault & al., September 2001[18]

Guilleminault et al., September 2001[18]	Esophagus	?	A more negative peak end inspiratory pressure with each successive breath.	Unclear; probably 4 breaths	P_{es} reversal.	Guilleminault et al., 1995[43]
Guilleminault et al., September 2001[18]	Esophagus	Mean peak end inspiratory P_{es} of preceding normal breaths	A clear and sudden increase in inspiratory effort without 'crescendo' with negative peak end inspiratory P_{es} more negative than at least 2 SD below baseline.	4 breaths	P_{es} reversal.	Guilleminault &al., 1995[43]
Bachour et al., 2002[45]	Esophagus	?	Irregular respiration with crescendo in P_{es}.	?	Rapid return from negative P_{es} of at least 5 cmH_2O (or 50%) below baseline back to baseline.	None

Bachour et al., 2002[45]	Esophagus	?	Progressively more negative esophageal pressure.	10 s*	Sudden change in pressure to a less negative level and an arousal.	AASM Task Force in Chicago, 1999[6]
Bachour et al., 2002[45]	Esophagus	?	Progressively more negative esophageal pressure.	10 s	A sudden return to the baseline.	None
Argod etal., 2000[46]	Esophagus	?	Increase in respiratory effort in a crescendo scoring rule.	?	Unmentioned.	None
Argod etal., 2000[46]	Esophagus	?	Flow limitation without flow reduction occurred concurrently with a crescendo in P_{es}.	?	Arousal followed by P_{es} going back to resting levels.	None

* The events in the figures in the paper were actually shorter than 10 s.

Chervin etal., 2012[47]	Esophagus	The peak expiratory pressure within the same breath	Sleep epoch spent with most esophageal pressure swings more negative than -10cmH_2O.	One epoch	End of epoch.	?
Katz et al., 2003[48]	Esophagus	?	Graded increases in P_{es}.	10 s	Abrupt P_{es} reversal.	Guilleminault & al., 1993[32]
Masa etal., 2003[49]	Esophagus	?	Increasing negative esophageal pressures which did not coincide with increased oral-nasal flow.	Two breaths	Arousal with esophageal pressures less negative.	Ambiguous due to missing inline citation
Hutter et al., 2004[50]	Esophagus	? (All patients had baseline values <−5 cmH_2O.)	Decrescendo scoring rule with a nadir >−10 cmH_2O.	?	Crescendo.	Unclear due to missing inline citation

Johnson etal., 2005[51]	Esophagus	Previous 2 minutes of baseline stable breathing	Increased negative P_{es}.	One breath but less than 1 min	EEG arousal and a return to a less-negative P_{es}.	None
Kristo et al., 2005/2008[52]	Esophagus	?	Crescendo scoring rule of negative inspiratory pressures ≤−12 cmH_2O	?	Arousal followed by normalization of P_{es}.	Loube et al., May 1999[38]
Masa et al., 2009[53]	Esophagus	?	Progressive increase in esophageal pressure.	10 s	Arousal.	None
AASM, 2012 (2007 v2.0)[7]	Esophagus	?	Sequence of breaths characterized by increasing respiratory effort.	10 s	?	Unknown
Serwatko, 2016[5]	Distal esophagus	?	A progressive, breath-by-breath, more negative	10 s	An abrupt drop in respiratory effort, indicated by a less	Guilleminault et al.,

			inspiratory waveform not associated with oxygen desaturation followed by a sequence of variations in respiratory efforts.		negative end inspiratory waveform.	September 2001.[18]
Serwatko, 2016[5]	Distal esophagus	?	A relatively stable and persistent, more negative inspiratory waveform.	60 s	An abrupt drop in respiratory effort, indicated by a less negative end inspiratory waveform.	Guilleminault et al., September 2001.[18]

P_{es} is short for esophageal pressure, usually referring to the peak negative pressure during an attempted inspiration. Question marks mark information that was not clearly stated in the respective paper or essay.
Most of the cited papers were discovered by recursively following the references of Vandenbussche et al.,] 2015;[36] Exar and Collop, 1999;[11] and Marta Serwatko, 2016.[5] Cited references were restricted to articles available online in HTML or PDF.

Appendix 2. *The overlap coefficient of pairs of respiratory effort scoring rules as measured on the 26 examined polysomnograms.*

Prevalence		12%	12%	3%	8%	17%	33%
		technologist 1	**technologist 2**	**Cres5**	**Cres4**	**Cres3**	**Cres2**
18%	**cres10sec**	36%	34%	100%	99%	89%	100%
33%	**Cres2**	52%	47%	100%	100%	100%	
17%	**Cres3**	37%	34%	100%	100%		
8%	**Cres4**	33%	29%	100%			
3%	**Cres5**	40%	37%				
12%	**Technologist 2**	**34%**					

Prevalence		12%	12%	1%	2%	5%
		Technologist 1	**Technologist 2**	**Cresrev5**	**Cresrev4**	**Cresrev3**
11%	**Cresrev2**	14%	10%	51%	63%	76%
5%	**Cresrev3**	22%	14%	63%	80%	
2%	**Cresrev4**	31%	20%	77%		
1%	**Cresrev5**	39%	25%			
12%	**technologist 2**	**34%**				

Prevalence		12%	12%	0%	0%	1%
		Technologist 1	**Technologist 2**	**Cresbaserev5**	**Cresbaserev4**	**Cresbaserev3**
4%	**Cresbaserev2**	20%	18%	68%	71%	75%
1%	**Cresbaserev3**	38%	34%	68%	72%	
0%	**Cresbaserev4**	57%	53%	77%		
0%	**Cresbaserev5**	62%	53%			
12%	**Technologist 2**	**34%**				

Prevalence		4%	10%
		Cresbaserev2	**Cresrev2**
33%	**Cres2**	100%	100%
10%	**Cresrev2**	85%	

Prevalence		12%	1%	5%
		Technologist 2	**Cresbaserev3**	**Cresrev3**
17%	**Cres3**	35%	100%	100%
5%	**Cresrev3**	14%	75%	
1%	**Cresbaserev3**	40%		

Prevalence		0%	2%
		Cresbaserev4	**Cresrev4**
8%	**Cres4**	100%	100%
2%	**Cresrev4**	78%	

The border percentages state what proportion of the total sleep time matched the corresponding scoring rule. Each internal cell contains the overlap coefficient between the scoring rules corresponding to the row and the column, respectively. The expected value of the overlap coefficient, for uncorrelated scoring rules, is the prevalence of events identified in accordance with the scoring rule with the lower prevalence. The only overlap coefficient not greater than this expected value was the 10% overlap coefficient between Cresrev2 and technologist 2. The expected value of would have been 11% since both Cresrev2 and the technologist identified 11% or more of the total sleep time as IUAR events.

Appendix 3. The overlap coefficient is equal to precision or recall, whichever is greater.

Proof.

By definition,

$$\text{overlap}(X, Y) = \frac{|X \cap Y|}{\min(|X|, |Y|)}$$

where X and Y are scorings, and $|X \cap Y|$ is the number of seconds which they agree belong to an IUAR event. If one scoring is considered an approximation to the other, $|X \cap Y|$ is the number of seconds belonging to true IUAR events; the number of true positive seconds. If X is considered as an approximation to Y, then $|X|$ is the sum of false positive and true positive seconds. Then $|Y|$ likewise becomes the sum of false negative and true positive seconds.

$$\text{overlap}(X, Y) = \frac{TP}{\min(FP + TP, FN + TP)}$$

The same formula can be obtained by defining Y as an approximation to X. Since none of the second counts is negative, picking a small denominator is choosing the denominator that maximizes the fraction.

$$\text{overlap}(X, Y) = \max\left(\frac{|X \cap Y|}{X}, \frac{|X \cap Y|}{Y}\right)$$

$$\text{overlap}(X, Y) = \max\left(\frac{TP}{FP + TP}, \frac{TP}{FN + TP}\right)$$

But $\frac{|X \cap Y|}{X}$ and $\frac{|X \cap Y|}{Y}$ are the precision and recall of a scoring that is considered an approximation of another. Thus,

$$\text{overlap}(X, Y) = \max(precision, recall)$$

and note that $\text{overlap}(X, Y) = \text{overlap}(Y, X)$.

■