

The American Academy of Sleep Medicine Inter-scorer Reliability Program: Respiratory Events

Richard S. Rosenberg, Ph.D., F.A.A.S.M.¹; Steven Van Hout, B.S.²

¹California State University, Long Beach, CA; ²American Academy of Sleep Medicine, Darien, IL

Study Objectives: The American Academy of Sleep Medicine (AASM) Inter-scorer Reliability program provides a unique opportunity to compare a large number of scorers with varied levels of experience to determine agreement in the scoring of respiratory events. The objective of this paper is to examine areas of disagreement to inform future revisions of the *AASM Manual for the Scoring of Sleep and Associated Events*.

Methods: The sample included 15 monthly records, 200 epochs each. The number of scorers increased steadily during the period of data collection, reaching more than 3,600 scorers by the final record. Scorers were asked to identify whether an obstructive, mixed, or central apnea; a hypopnea; or no event was seen in each of the 200 epochs. The "correct" respiratory event score was defined as the score endorsed by the most scorers. Percentage agreement with the majority score was determined for each epoch and the mean agreement determined.

Results: The overall agreement for scoring of respiratory events was 93.9% ($\kappa = 0.92$). There was very high agreement on epochs without respiratory events (97.4%), and the majority score for most of the epochs (87.8%) was no event. For the 364 epochs scored as having a respiratory event, overall

agreement that some type of respiratory event occurred was 88.4% ($\kappa = 0.77$). The agreement for epochs scored as obstructive apnea by the majority was 77.1% ($\kappa = 0.71$), and the most common disagreement was hypopnea rather than obstructive apnea (14.4%). The agreement for hypopnea was 65.4% ($\kappa = 0.57$), with 16.4% scoring no event and 14.8% scoring obstructive apnea. The agreement for central apnea was 52.4% ($\kappa = 0.41$). A single epoch was scored as a mixed apnea by a plurality of scorers.

Conclusions: The study demonstrated excellent agreement among a large sample of scorers for epochs with no respiratory events. Agreement for some type of event was good, but disagreements in scoring of apnea vs. hypopnea and type of apnea were common. A limitation of the analysis is that most of the records had normal breathing. A review of controversial events yielded no consistent bias that might be resolved by a change of scoring rules.

Keywords: Scoring, reliability, apnea, hypopnea, respiratory events

Citation: Rosenberg RS; Van Hout S. The American Academy of Sleep Medicine inter-scorer reliability program: respiratory events. *J Clin Sleep Med* 2014;10(4):447-454.

Identification of respiratory events during sleep is critical to the diagnosis of sleep related breathing disorders, especially obstructive sleep apnea.¹ Rules are provided by the *AASM Manual for the Scoring of Sleep and Associated Events*² (the *Manual*) and require not only identification of events but classification of types. Apnea types include obstructive, mixed, and central events. In addition, apneas are distinguished from hypopneas, defined as a partial decrease of air flow as measured using a nasal pressure transducer.

Studies have attempted to assess scoring reliability for respiratory events, often employing non-standard techniques. A widely cited study published before the *Manual* used data from home sleep testing relying on a single nasal thermistor.³ The study defined apneas as a 75% or greater reduction of flow amplitude and hypopneas as a 30% or greater reduction of flow amplitude. A sample of 60 recordings was scored by 3 independent scorers and interclass correlations determined. A unique feature of this study was that recordings were scored and rescored with different hypopnea definitions with respect to oxygen desaturation and arousal. Agreement was based on the respiratory disturbance index (RDI). Extremely high interclass correlations were seen when desaturations were required, rising

to 0.99 when 4% or 5% desaturation cutoffs were used. These data were subsequently used in the construction of the *Manual* scoring rules.⁴

Scoring agreement may be based on a variety of measures. At one end of this spectrum, agreement based on classification of normal, mild, moderate, or severe obstructive sleep apnea (OSA) provides a crude measure of agreement. Use of the RDI or apnea-hypopnea index (AHI) provides a clinically relevant measure in that these scores are used in the diagnostic criteria for apnea. However, perfect agreement for classification or index can be achieved with complete disagreement on specific events. Assessment of agreement on the duration of specific events, at the other end of the spectrum, may indicate disagreement when the events identified by scorers overlap on all but a few seconds, depending on the algorithm used to define agreement. These minor variations have no clinical significance except when the event duration is at or near the criterion of 10 seconds.

The AASM Inter-scorer Reliability program uses an intermediate measure of agreement based on an epoch-by-epoch analysis. Scorers are asked to identify whether or not an epoch contains a respiratory event. Events were scored in each of the

epochs in which they occurred, providing a gross measure of the scorer's ability to evaluate the duration of the event. However, this precludes the calculation of a precise AHI as it does not provide the number of respiratory events in a record; events that are seen in more than one epoch would be scored more than once. Scorers were also asked to classify the events seen in each epoch as obstructive, central, or mixed apnea. Scorers were asked to identify the presence of hypopnea but were not asked to discriminate between obstructive and central hypopnea.

A report by Pittman and colleagues allows comparison of these assessment methods.⁵ Their analysis was based on 31 sleep studies scored by 2 independent sleep technologists. There was 93.5% agreement on OSA severity with all disagreements occurring in the mild and moderate categories. The intraclass correlation coefficient (ICC) for RDI was 0.99 between the 2 manual scorers. An epoch-by-epoch analysis of respiratory scoring resulted in agreement on 94.9% of epochs with a κ value of 0.82. This suggests that there is good agreement between measures of scoring reliability; the epoch-by-epoch method provides a measure of agreement that is similar to severity classification and AHI or RDI comparison. However, it should be noted that this agreement does not reflect differences in apnea type or differentiate apneas from hypopneas.

Apnea type and hypopnea agreement were measured in a recent international study.⁶ This study compared scorers at 9 sites using 15 full-night, attended sleep records. The ICC for the AHI was 0.95. This value fell to 0.73 for all apneas and 0.80 for hypopneas, indicating some disagreement in distinguishing apneas from hypopneas. Agreement on classification of apnea type was much less robust, with ICC values of 0.70 for obstructive events, 0.42 for mixed apneas, and 0.46 for central apneas.

The AASM Inter-scorer Reliability program was described in our previous report on sleep stage scoring.⁷ It provides standard recordings each month in a web-based format that currently includes more than 3,500 regular users. Users are given immediate feedback based on the majority score of a 3-person panel of experts. The "correct" score for the present analysis was based on the largest number of users rather than the panel score. According to a survey taken in July 2012, the majority of users were experienced sleep technologists, with few physicians and few novice scorers participating in the program.

We hypothesized that overall scoring agreement for sleep related breathing events would be similar to that found in previous studies. We anticipated that scorers would have some difficulty discriminating between apnea and hypopnea, and considerable difficulty differentiating apnea type. Events with poor agreement were reviewed to determine if there were elements of the scoring rules that were particularly troublesome. This was used to suggest changes to the scoring rules that might result in higher levels of agreement.

METHODS

Analyses were based on scoring of the 200 epoch monthly record fragments provided by the AASM Inter-scorer Reliability program between January 2012 and March 2013. Several improvements in the display and scoring methodology were implemented during that time, but these had no effect on the agreement in stage scoring⁷ and were unlikely to have

an effect on respiratory scoring. Instructions for users of the AASM Inter-scorer Reliability program for scoring respiratory events were as follows:

Score any event that appears in the epoch by using the Scoring Answer Menu and selecting up to two of the following choices in the "Respiratory" category: "Obstructive Apnea", "Central Apnea", "Mixed Apnea" or "Hypopnea". Because not every epoch has a respiratory event, the default answer is none. Score any event that appears in an epoch, even if the event appears across multiple epochs, and only a portion of the event occurs in the current epoch. For example, if an obstructive apnea starts in epoch 97 and ends in epoch 98, score OA for epochs 97 and 98. If there is an obstructive apnea and a central apnea in epoch 125, score OA and CA. If there are 2 obstructive apneas in epoch 36, score OA for the epoch. Use Rule 4.A. for scoring hypopneas. The criteria are a drop of 30% of the nasal pressure transducer signal and an oxygen desaturation of at least 4%. The pre-event baseline should be defined as the closest stable oximetry reading prior to the event or in the case where events are sufficiently frequent such that recovery of oxygenation does not occur, the highest value prior to the event. Although it is recognized that apneas and hypopneas can occur during drowsiness preceding stage N1 sleep, these should not be scored because of the difficulty of defining the denominator to calculate an apnea hypopnea index. Therefore, do not enter any events for all epochs scored as W.

These rules allowed for the possibility of more than one type of event in a particular epoch, but the majority score never included more than one type of event for any of the epochs in this analysis. The rules do not allow for a precise calculation of an AHI, but the number of epochs with respiratory events per record provides an estimate of severity. This resulted in the categorization of 1 record as severe OSA, 2 moderate OSA, 1 mild OSA, and the remaining 11 record fragments as within normal limits. The majority score identified no respiratory events at all in 6 of the recordings. Admittedly this is not a representative sample of patients from a typical sleep disorders center. However, the purpose of the AASM Inter-scorer Reliability program is to "satisfy the AASM Standards for Accreditation item F-6: Inter-scorer Reliability by using the AASM Inter-scorer Reliability Assessment System. This standard requires sleep technologists at each site to demonstrate competency in scoring."⁸ In keeping with this, records were chosen that had minimal artifact and robust signals. Most of these records were normal or nearly normal. Patients ranged from 26 to 78 years of age.

Records used in this study are listed in **Table 1** and were recorded in accordance with AASM standards.² Eleven of the studies were diagnostic with flow recorded from an oro-nasal thermistor and a nasal pressure transducer. Four of the studies were CPAP titrations with flow recorded from the PAP device. Respiratory inductance plethysmography effort belts were used in all studies.

The majority score was determined by counting the number of users who scored no event, obstructive apnea, mixed apnea,

Table 1—Records Used in the Analysis

Record Number	Date	Age	Sex	Type of Study	Epochs with Events	Estimated AHI
1	Jan-12	36	M	Diagnostic	0	0
2	Feb-12	78	F	CPAP	5	1.5
3	Mar-12	40	M	Diagnostic	51	15.3
4	Apr-12	26	M	Diagnostic	1	0.3
5	May-12	37	F	Diagnostic	0	0
6	Jun-12	45	M	Diagnostic	187	56.1
7	Jul-12	66	M	CPAP	2	0.6
8	Aug-12	54	M	CPAP	5	1.5
9	Sep-12	46	M	CPAP	0	0
10	Oct-12	30	F	Diagnostic	0	0
11	Nov-12	26	F	Diagnostic	37	11.1
12	Dec-12	66	M	Diagnostic	0	0
13	Jan-13	40	M	Diagnostic	68	20.4
14	Feb-13	38	M	Diagnostic	8	2.4
15	Mar-13	26	F	Diagnostic	0	0

central apnea, or hypopnea. The category with the most scores was used as the “correct” score. This meant that if scores for apnea and hypopnea were somewhat evenly split among the categories, it would be possible for the majority score to be no event even if most scorers scored a respiratory event. Epochs scored as having more than one type of respiratory event were rare and never achieved a majority score. For purposes of categorizing “incorrect” responses, these were included in the type of event chosen first by the scorer.

The number of scorers increased month-by-month, with 2,630 scorers for the January 2012 record and 3,692 for the March 2013 record. A total of 9,108,197 scoring decisions were recorded and used in this analysis. Data for individual scorers is confidential, and it was therefore impossible to track individual users to see if there was consistency across records. However, there was no reason to assume that the users at the beginning of the sample were any different from those at the end of the sample.

RESULTS

Of the 3,000 epochs used in this analysis, 364 (12.2%) were scored as having a respiratory event by the majority of scorers. Of these, 47.3% were hypopnea, 41.2% obstructive apnea, 11.3% central apnea, and a single event (0.3%) was scored as mixed apnea by a majority of scorers. The overall agreement matrix is shown in **Table 2**.

The overall agreement for scoring of respiratory events was 93.9% ($\kappa = 0.92$). This was primarily due to very high agreement on epochs without respiratory events at 97.4% and the fact that the majority score for most of the epochs (87.8%) was no respiratory event. The most common disagreement for these epochs was a score of hypopnea, which was endorsed by 1.6% of scorers.

For the 364 epochs scored as having a respiratory event by the majority of scorers, the overall agreement that some type of respiratory event occurred during the epoch was 88.4% ($\kappa = 0.77$). The agreement for epochs scored as obstructive apnea by the majority was 77.1% ($\kappa = 0.71$), and the most common disagreement was a score of hypopnea rather than obstructive

Table 2—Scoring Agreement for the 15 Records Used

		None	OA	MA	CA	HY
Majority Score	None	97.4%	0.5%	0.1%	0.4%	1.6%
	OA	5.5%	77.1%	1.5%	1.4%	14.4%
	MA*	17.7%	17.8%	39.8%	22.1%	2.5%
	CA	11.1%	20.3%	5.5%	52.4%	10.7%
	HY	16.4%	14.8%	0.9%	2.4%	65.4%

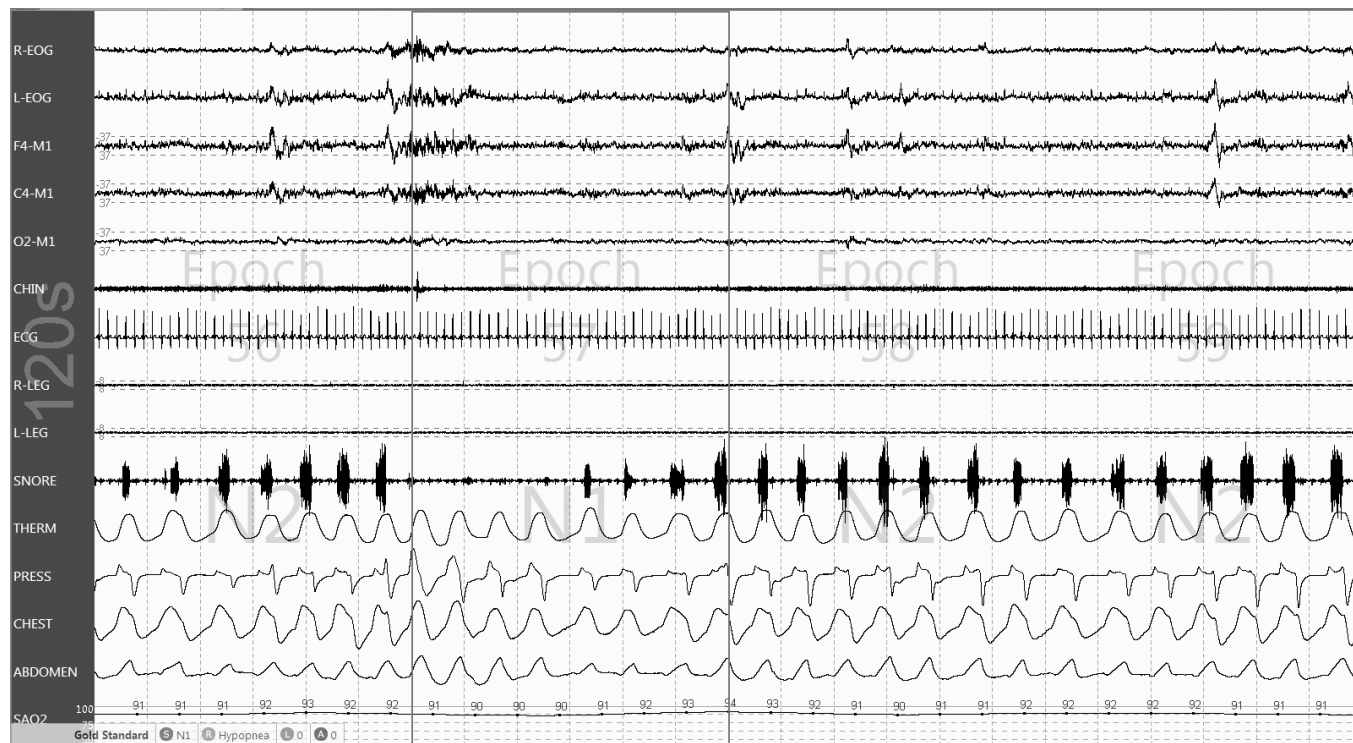
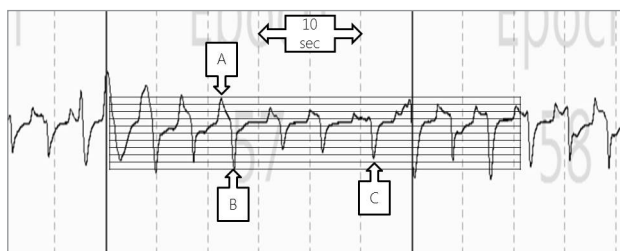
*A single epoch was scored as a mixed apnea by the majority of scorers. OA, obstructive apnea; MA, mixed apnea; CA, central apnea; HY, hypopnea.

apnea by 14.4% of scorers. The agreement for hypopnea was 65.4% ($\kappa = 0.57$), with 16.4% scoring no event in those epochs and 14.8% scoring obstructive apnea. The agreement for central apnea was 52.4% ($\kappa = 0.41$). A single epoch was scored as a mixed apnea by a plurality of scorers.

Event vs. No Event

Figure 1 includes epoch 57 from record #14 (second epoch from the right). This is an example of an epoch where a slim majority of scorers scored no event. Loud snoring is seen throughout the 2-minute sample. There is no change in the thermal signal, but midway through the epoch a decrease in the amplitude of the pressure signal occurs with flattening of the waveform. Chest and abdominal signals persist through the epoch with a tendency toward paradoxical effort. The highest oxygen saturation was 94% at the end of the epoch, and over the course of the next 20 seconds the saturation reached a low of 90%. This appears to meet the scoring criteria for hypopnea, as the event lasted for more than 10 seconds, included a decrease of the pressure signal amplitude of more than 30%, and was accompanied by a 4% oxygen desaturation. The panel of experts scored a hypopnea in this epoch, but only 49.5% of scorers agreed.

The nasal pressure signal from this epoch and a portion of the epochs before and after is shown in **Figure 2**. The amplitude of the signal varied across this sample making it difficult to choose

Figure 1—An epoch scored as having no respiratory event by a majority (50.2%) of scorers.**Figure 2**—Nasal pressure signal from Figure 1.

See text for details.

a baseline breath for comparison. Using the breath marked A as the baseline, a grid can be applied to measure decreases during subsequent breaths. Each line on the grid represents 10% of the baseline breath amplitude. According to the scoring rules, an event begins at the nadir preceding the first reduced breath (marked as B in **Figure 2**). The end of the event is the nadir preceding the first normal breath (marked as C). Therefore, the duration of the event clearly exceeds the 10 second requirement. The rule for amplitude measurement requires that at least 9 seconds of the event have an amplitude decrease of at least 30% from baseline (3 lines on the grid). All 3 of the breaths between arrows B and C appear to meet this criterion.

The highest oxygen saturation, presumably associated with breath A and the 3 breaths preceding it, was 94%. This appears at the very end of the epoch and is therefore delayed by about 25 seconds. The low point for oxygen saturation of 90% occurs approximately 17 seconds after the peak. One report⁹ recommends scoring of low saturation within 30 seconds of the end of the event, and these measurements fall within that window.

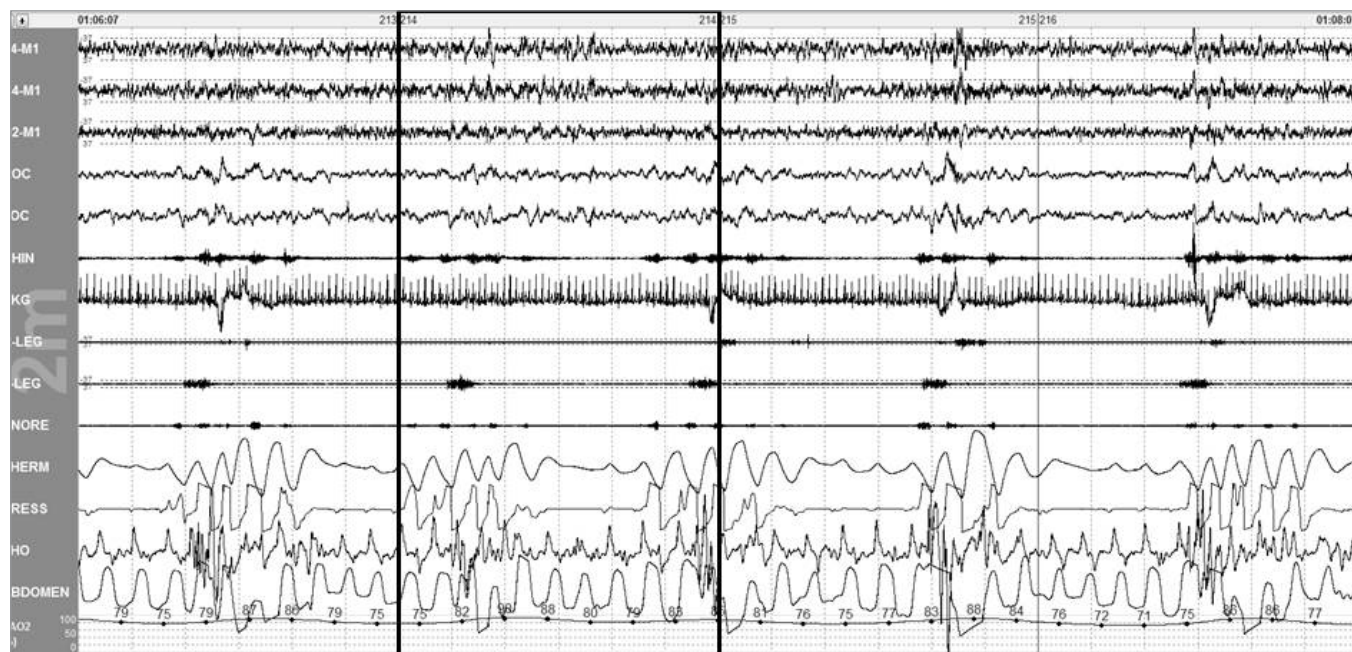
Apnea vs. Hypopnea

Figure 3 is from record #6, the patient with severe OSA. The majority of scorers (50.1%) scored an obstructive apnea in this epoch. The expert scorers scored hypopnea, as did 44.7% of the program users. Scoring of obstructive apnea required an event lasting at least 10 seconds, with a decrease of the thermal signal by 90% or more lasting at least 9 seconds. Oxygen desaturation is irrelevant for the scoring of apnea but was approximately 10% for this event. Respiratory effort was clearly present in the thoracic and especially abdominal tracings. The duration of the event clearly exceeds the 10-second requirement. Disagreement on this epoch clearly relies on differences in the scoring of the thermal signal amplitude.

Figure 4 provides the oro-nasal thermal sensor for this event. Selection of a baseline breath is more difficult than in the previous tracing as the patient is having repetitive events throughout this portion of the recording. Breath A is chosen as the baseline and the 10% grid is placed over the tracing. The event is scored as beginning at point B, the nadir preceding the first reduced breath. The event ends at point C, the nadir preceding the first normal breath, and exceeds the 10-second requirement for a respiratory event. The 90% reduction in amplitude means that the signal cannot exceed the distance between 2 of the horizontal bars on the grid.

The 2 partial breaths in the second half of the event do not meet the 90% reduction in amplitude. Therefore, the portion of the event with no breaths must last at least 9 seconds to meet the duration of the amplitude reduction criterion. This is an especially difficult decision to make, especially with the drift of thermal signal amplitude.

Figure 3—An epoch scored as having an obstructive apnea by a majority (50.1%) of scorers.



Obstructive vs. Central Apnea

Scoring for the event in **Figure 5** was more evenly distributed than for the epochs previously discussed. For this epoch (record #13, epoch 169), 33.4% scored central apnea and this became the “correct” answer, 33.0% scored obstructive apnea, 7.2% scored mixed apnea, 21.9% scored hypopnea, and 4.5% scored no event. The expert scorers scored an obstructive apnea for this epoch.

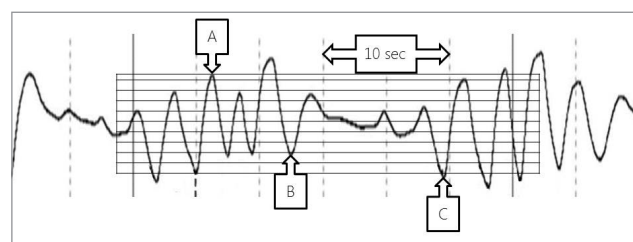
Most scorers (73.6%) scored this event as an apnea rather than hypopnea requiring that a 9-second portion of the event meet the 90% reduction in oro-nasal thermal signal amplitude. In **Figure 6** the baseline breath is chosen at A and the 10% grid is placed. This is somewhat arbitrary in that it follows a respiratory event and may represent hyperpnea. The event begins at B with the nadir preceding the first reduced breath and ends at C with the nadir preceding the first normal breath. Partial breaths are seen at D and E, and the assumption behind a score of apnea is that there is a 9-second period with a 90% reduction in amplitude between these arrows.

The scoring rules require that mixed apnea must have an absence of respiratory effort in the first half of the event and resumption of effort in the second half of the event.² Obstructive apnea must have continued or increased respiratory effort evident throughout the entire period of absent air flow. Central apnea must have absent respiratory effort throughout the entire period of absent air flow. Turning to the effort channels in **Figure 6**, clear paradoxical effort is seen at arrow F. This would seem to preclude the scoring of a mixed apnea as the effort occurs in the first half of the event. The scoring decision rests, therefore, on the definition of the “period of absent airflow” and the interpretation of the signal in the effort channels between arrow F and the end of the event.

Mixed Apnea

The event in **Figure 7**, from record #11, was scored as a mixed apnea in epoch 133 by 39.8% of scorers. It was scored

Figure 4—Nasal-oral thermal signal from Figure 3.

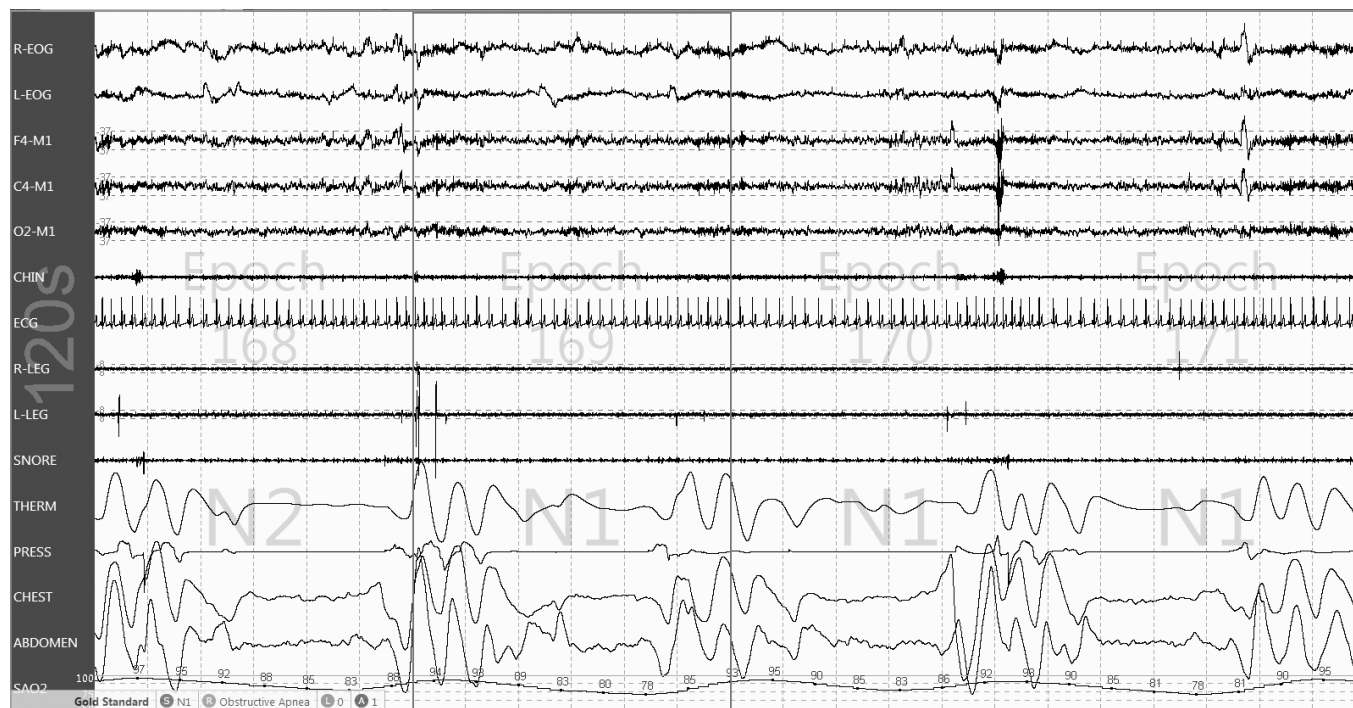
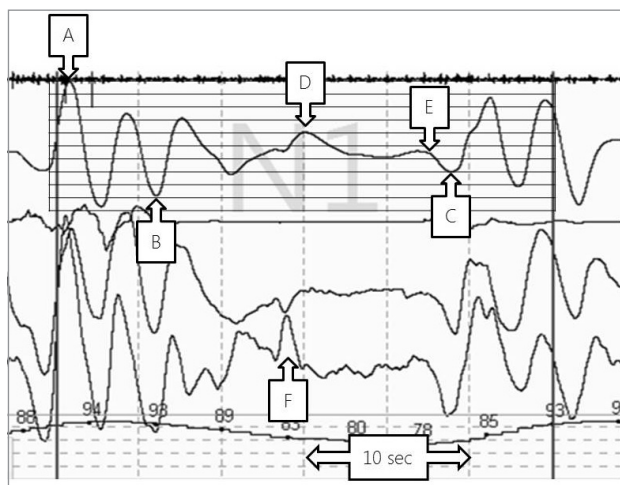


See text for details.

as an obstructive apnea by 17.7% of scorers and a central apnea by 22.1% of scorers. No event was scored for the epoch by 17.8% of scorers. The event begins with an apparent central component. Paradoxical movement of the chest and abdomen is seen in the second half of the event. The oro-nasal thermal channel appears to decrease by more than 90% for almost the entire epoch. Interestingly, a clear majority of scorers scored no event in epoch 132, even though nearly all scorers (90.5%) scored epoch 132 as stage N2. As mentioned, the scoring rules clearly state that events should be scored in all of the epochs in which they occur. The first abnormal breath for the event in epoch 133 occurs in epoch 132, and therefore the event should be scored in both epochs.

DISCUSSION

The results of this study indicate high agreement among scorers in the AASM Inter-scorer Reliability program in identifying epochs with normal breathing during sleep. Agreement for epochs scored by the majority of users as having no respiratory events was 97.4%. Agreement for epochs where the majority scored an event was 88.4%, leading to an overall

Figure 5—An epoch scored as having an obstructive apnea by a majority (50.1%) of scorers.**Figure 6**—Respiratory signals from Figure 5.

See text for details.

agreement of 93.9%. Agreement was significantly reduced for epochs scored by the majority as having hypopnea (65.4%). Agreement was better for obstructive apnea (77.1%) than for epochs where the majority scored a central apnea (52.4%). Cohen's κ values for agreement was excellent for identifying the absence of an event but more in the moderate range for the type of event scored.

The study has some major limitations, chief among them that most of the records had normal breathing resulting in a small number of epochs scored as having respiratory events by the majority of scorers. For example, a single epoch was scored as a mixed apnea by the majority (only 4 epochs were identified as containing a mixed apnea by the expert scorers). The prevalence

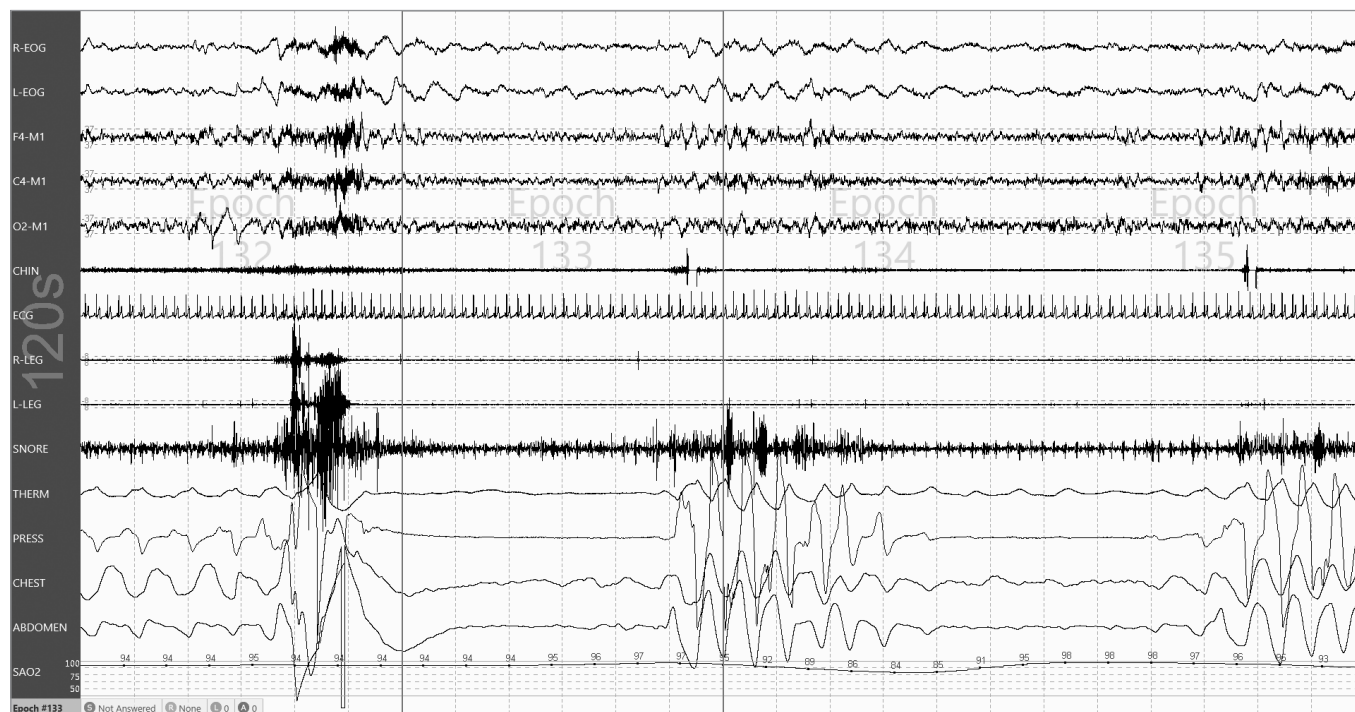
of normal breathing during sleep raises concerns regarding the validity of our sample as an indicator of scoring agreement in sleep centers. However, the high number of scorers participating in the program presents a unique opportunity to provide “real world” data on scoring agreement. This contrasts to other published estimates of scoring agreement which relied on a small number of scorers but typically used a more representative set of recordings.

Epochs of high disagreement provided an opportunity to explore the scoring rules and speculate on the causes of poor agreement. The epoch in **Figure 1** had a modest reduction of the amplitude of the nasal pressure signal. Scorer disagreement in the percentage reduction of amplitude probably accounts for much of the disagreement in scoring of the event. This event (or non-event) is notable for the absence of an arousal, which, although not part of the scoring rule for hypopnea in effect at the time this record was scored, may have influenced some scorers not to score the event.

Similarly, the amplitude of the oro-nasal thermal sensor decreased in the epoch in **Figure 3**. Here the question is not whether the amplitude reached the 90% reduction required for scoring of apnea, but whether the amplitude remained below this level for the 9 seconds required by the scoring manual.

An epoch illustrating the difficulty in apnea classification is presented in **Figure 5**. A portion of the event has clear respiratory effort, but is followed by a portion with little or no effort. This cannot be called a mixed apnea due to the obstructive part preceding the central part. Another complication is the requirement for central apneas to have absent effort during the portion of the event with absent air flow. Is absent air flow the same as signal amplitude reduced by 90% or more? Or is the portion of the event with paradoxical effort and limited air flow sufficient to call the event obstructive? The epoch scored as having

Figure 7—The single epoch scored as a mixed apnea by the largest number of scorers.



a mixed apnea (**Figure 7**) was in fact scored as something else by 60.2% of scorers.

Modifications of the rules for scoring respiratory events were published in 2012,¹⁰ but were not incorporated into the AASM Inter-scorer Reliability program until June 2013. All of the records included in this analysis were scored prior to this change. The changes were made as a result of deliberations by the Sleep Apnea Definitions Task Force which included discussion of technical requirements and scoring rules.¹¹ However, the most recent version of the rules makes the 3% or arousal rule “recommended” and the 4% rule “acceptable.”¹² The AASM Inter-scorer Reliability program has since reverted to the “acceptable” rule for respiratory event scoring.

Several changes to the scoring rules remain in effect. The Task Force modified the requirement for duration of the specified flow decrease—a 10-second event is required, and a minimum of 10 seconds must meet the required decrease of flow amplitude. Events that meet criteria for apnea and hypopnea must now be scored as apnea. The requirement for apnea to have an amplitude reduction of 90% or more in the oro-nasal thermal channel remains the same.

The amplitude reduction criterion is difficult to operationalize for many reasons. Thermal flow signal amplitude is affected by changes in flow volume, distance of the probe from the nose and mouth, ambient temperature, and accumulation of fluids on the probe tips. The frequent and cyclical events such as that in patient #6 (**Figure 3**) suggests that in patients with severe disease periods of apnea alternate with periods of hyperpnea and a baseline amplitude may not be evident. The scoring rule notes provide guidance on determining the end of an event based on a 2% resaturation, but no guidance on estimating baseline flow amplitude is provided.¹⁰

The rule changes for duration measurement may help improve agreement, and scoring apneas whenever a 10-second portion of an event reaches the 90% reduction criterion will likely help as well. In the examples shown here, a grid was used to help estimate a 90% reduction. We are not aware of this feature in any available scoring systems or any other aid for scoring changes in flow amplitude. Baseline drift is common in flow signals, and it is uncertain whether this represents slow exhalation or is an artifact of filtering. The purpose of the 90% reduction criterion is to enable scoring of an apnea even when there are minor fluctuations in the amplitude of the flow signal that are not breaths. But the 90% figure is arbitrary and difficult to score at times. One feature that separates physiological flow from artifact is the concordance of flow and effort. A more reliable score might result if the rule was to score apnea if there is 10 seconds without concordant flow and effort.

A review of the epochs in this sample appears to indicate that a 30% reduction is at the limit of scorers to detect. For epochs scored as containing a hypopnea, 16.4% scored no event. The “recommended” rules in the *Manual* suggest that the authors expect scorers to identify periods of decreased flow amplitude and then confirm the scoring of an event by identifying a desaturation or arousal. However, some scorers may use a reverse strategy because identification of flow amplitude changes is difficult. In many patients a 3% desaturation may occur from time to time during a recording with no apparent trigger. It may be that even prior to the rule change some scorers were already using an arousal to signal the presence of a preceding event, and then look for an associated desaturation. In the absence of an arousal, no event may be scored.

Widespread adoption of the “recommended” rule will result in a significant increase in the number of hypopneas scored.

The effect of this change on scoring agreement is unknown at this point and the AASM Inter-scorer Reliability program does not have sufficient data to contribute to the discussion. Moving forward, the program will provide data to assess agreement for the “acceptable” rule and allow comparison with the agreement described here.

REFERENCES

1. American Academy of Sleep Medicine. International Classification of Sleep Disorders, Second Edition: Diagnostic and Coding Manual. Westchester, IL: American Academy of Sleep Medicine, 2005.
2. Iber C, Ancoli-Israel S, Chesson AL Jr., Quan SF. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications, 1st ed. Westchester, IL: American Academy of Sleep Medicine, 2007.
3. Whitney CW, Gottlieb DJ, Redline S, et al. Reliability of scoring respiratory disturbance indices and sleep staging. *Sleep* 1998;21:749-57.
4. Redline S, Budhiraja R, Kapur V, et al. The scoring of respiratory events in sleep: reliability and validity. *J Clin Sleep Med* 2007;3:169-200.
5. Pittman SD, MacDonald MM, Fogel RB, et al. Assessment of automated scoring of polysomnographic recordings in a population with suspected sleep-disordered breathing. *Sleep* 2004;27:1394-403.
6. Magalang UJ, Chen NH, Cistulli PA, et al. Agreement in the scoring of respiratory events and sleep among international sleep centers. *Sleep* 2013;36:591-6.
7. Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring. *J Clin Sleep Med* 2013;9:81-7.
8. American Academy of Sleep Medicine. Inter-scorer Reliability. Accessed on July 29, 2013. <http://www.aasmnet.org/isrl/>.
9. Zafar S, Ayappa I, Norman RG, Krieger AC, Walsleben JA, Rapoport DM. Choice of oximeter affects apnea-hypopnea index. *Chest* 2005;127:80-8.

10. Berry R, Brooks R, Gamaldo C, et al.; for the American Academy of Sleep Medicine The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.0. www.aasmnet.org. Darien, IL: American Academy of Sleep Medicine, 2012.
11. 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;8:597-619.
12. Berry RB, Brooks R, Gamaldo CE, et al.; for the American Academy of Sleep Medicine. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.0.2. www.aasmnet.org, Darien, IL: American Academy of Sleep Medicine, 2013.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication January, 2014

Accepted for publication January, 2014

Address correspondence to: Richard S. Rosenberg, Ph.D., American Academy of Sleep Medicine, 2510 N Frontage Road, Darien, IL 60561; Tel: (630) 737-9700; Fax: (630) 737-9790; E-mail: rrosenberg@aasmnet.org

DISCLOSURE STATEMENT

This is not an industry supported study. Dr. Rosenberg has provided consulting services to the American Academy of Sleep Medicine, the American Association of Sleep Technologists and Natus Neurology, Inc. He has received a speaker's honorarium from the Louisiana Association of Sleep Medicine. Mr. Van Hout is a full time employee of the American Academy of Sleep Medicine. The American Academy of Sleep Medicine developed the inter-scorer reliability program.