The 2012 AASM Respiratory Event Criteria Increase the Incidence of Hypopneas in an Adult Sleep Center Population

Brett Duce, BSc (Hons), RPSGT[1,2]; Jasmina Milosavljevic, MBBS, MPH[1]; Craig Hukins, MBBS FRACP[1]

[1]Sleep Disorders Centre, Department of Respiratory & Sleep Medicine, Princess Alexandra Hospital, Ipswich Rd, Woolloongabba, Australia; [2]Faculty of Science and Engineering, Queensland University of Technology, Brisbane, Australia

Study Objectives: To investigate the effect of the 2012 American Academy of Sleep Medicine (AASM) respiratory event criteria on severity and prevalence of obstructive sleep apnea (OSA) relative to previous respiratory event criteria.

Methods: A retrospective, randomized comparison was conducted in an Australian clinical sleep laboratory in a tertiary hospital. The polysomnograms (PSG) of 112 consecutive patients undertaking polysomnography (PSG) for suspected OSA were re-scored for respiratory events using either 2007 AASM recommended (AASM$_{2007Rec}$), 2007 AASM alternate (AASM$_{2007Alt}$), Chicago criteria (AASM$_{1999}$), or 2012 AASM recommended (AASM$_{2012}$) respiratory event criteria.

Results: The median AHI using AASM$_{2012}$ was approximately 90% greater than the AASM$_{2007Rec}$ AHI, approximately 25% greater than the AASM$_{2007Alt}$ AHI, and approximately 15% lower than the AASM$_{1999}$ AHI. These changes increased OSA diagnoses by approximately 20% and 5% for AASM$_{2007Rec}$ and AASM$_{2007Alt}$, respectively. Minimal changes in OSA diagnoses were observed between AASM$_{1999}$ and AASM$_{2012}$ criteria. To achieve the same OSA prevalence as AASM$_{2012}$, the threshold for previous criteria would have to shift to 2.6/h, 3.6/h, and 7.3/h for AASM$_{2007Rec}$, AASM$_{2007Alt}$, and AASM$_{1999}$, respectively. Differences between the AASM$_{2007Rec}$ and AASM$_{2012}$ hypopnea indices (HI) were predominantly due to the change in desaturation levels required. Alterations to respiratory event duration rules had no effect on the HI.

Conclusions: This study demonstrates that implementation of the 2012 AASM respiratory event criteria will increase the AHI in patients undergoing PSG, and more patients are likely to be diagnosed with OSA.

Commentary: A commentary on this article appears in this issue on page 1357.

Keywords: obstructive sleep apnea, hypopnea definition, respiratory event scoring, AASM manual, methodology

Citation: 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.

T he identification of obstructive apneas and hypopneas in a polysomnogram (PSG) is a key component in the diagnosis, classification, and prevalence of obstructive sleep apnea (OSA). This contribution has been recognized with multiple efforts over time to standardize definitions of these events. The first publication to standardize definitions of obstructive apneas and hypopneas was the "Chicago Criteria" (AASM$_{1999}$).[1] Although this document stressed that these criteria were for research purposes, the authors did encourage clinicians to consider these criteria with respect to their day-to-day practice. Of particular note were the two hypopnea definitions that could be used together in the scoring of events. The first hypopnea definition required > 50% decrease in a valid measure of breathing during sleep without any consequence (either SpO$_2$ desaturation or EEG arousal). The second definition required a "clear amplitude reduction" (the amplitude reduction was not specified) with an associated 3% SpO$_2$ desaturation or EEG arousal. The implementation of AASM$_{1999}$ by sleep laboratories was not uniform.[2]

In 2007, the AASM published the Manual for the Scoring of Sleep and Associated Events (hereafter known as the 2007 Manual).[3] The 2007 Manual tried to improve standardization of clinical practice and incorporate the evidence that had accrued.

BRIEF SUMMARY

Current Knowledge/Study Rationale: The rules for the scoring of hypopneas have been modified with the release of the American Academy of Sleep Medicine's 2012 version of the Manual for the Scoring Sleep and Associated Events. The clinical impact of these new recommended criteria with respect to the prevalence of OSA and OSA severity classification have not been clarified.

Study Impact: This study demonstrated that the prevalence of OSA will increase with the implementation of the AASM 2012 hypopnea criteria. Changes to the SpO$_2$ desaturation criteria as well as the addition of the EEG arousal to the hypopnea rules contributed almost equally to the increased AHI observed in this study.

The rules for scoring respiratory events in the 2007 Manual included two hypopnea definitions. Unlike the AASM$_{1999}$, the definitions could not be used together in the scoring of respiratory events. The recommended hypopnea definition (AASM$_{2007Rec}$) required ≥ 30% decrease in nasal pressure amplitude with at least 4% SpO$_2$ desaturation. The alternative hypopnea definition (AASM$_{2007Alt}$) required ≥ 50% decrease in nasal pressure amplitude with at least 3% SpO$_2$ desaturation or an EEG arousal. The introduction of two hypopnea definitions in the 2007 Manual has attracted some criticism,[4,5] not just from

a standardization point of view but also from the contrasting results in the AHI. The work of Ruehland et al.[5] best demonstrated the differences between the hypopnea definitions as well as their respective differences to the AASM$_{1999}$. After the release of the 2007 Manual, sleep societies outside the United States had to determine which hypopnea definition to recommend. While most sleep societies chose to use the AASM$_{2007Rec}$ definition, others such as Australia and New Zealand prescribed the use of AASM$_{2007Alt}$ as their recommended clinical hypopnea definition.[6]

In 2012 the AASM introduced an update of the 2007 Manual (hereafter known as the 2012 Manual).[7] The changes to the 2012 Manual are mostly in the respiratory event scoring rules, with changes to event definitions and sensor recommendations. Of particular interest was the 2012 Manual reverting to a single hypopnea definition (AASM$_{2012}$). The new hypopnea definition appears to be an amalgamation of the AASM$_{2007Rec}$ and AASM$_{2007Alt}$, requiring $\geq 30\%$ decrease in nasal pressure amplitude with at least 3% SpO$_2$ desaturation or an EEG arousal. This change in definition by the Sleep Apnea Definitions Task Force was recognition that clinicians treat OSA in their patients for more than just cardiovascular outcomes and thus the definitions should have a link to these other outcomes. Another, less heralded change to the 2012 Manual lies with the definition of respiratory event duration. The 2012 Manual is more strict in its requirement for all 10 seconds of the minimum duration to meet the amplitude criterion, whereas the 2007 Manual mandated that at least 9 seconds met the strict amplitude criterion while the other second of the 10-second duration to be "clearly decreased."

Recently, BaHammam et al.[8] demonstrated that the AASM$_{2012}$ criteria was generally associated with higher AHIs when compared to the AASM$_{2007Rec}$ and AASM$_{2007Alt}$ criteria. However, the full impact of the AASM$_{2012}$ criteria has not been clarified; most notably the equivalent thresholds with respect to previous criteria and the contributions of specific rule changes to AHI differences. Furthermore, since the AASM$_{2012}$ criteria seem to resemble the second of the two AASM$_{1999}$ hypopnea definitions, we believed it was prudent to compare these criteria as well. Thus, our study had a number of aims. Our first aim was to confirm the findings of BaHammam et al. with respect to differences between the most recent AASM hypopnea definitions. Our second aim was to assist clinicians with interpreting the new criteria with respect to previous AASM hypopnea criteria by calculating equivalent thresholds and determining the contribution of specific rule changes to differences in AHI.

METHODS

Patient Selection

Diagnostic PSGs from consecutive patients during the period of January 2012 to March 2012 were re-analyzed for this study. Only patients examined for suspected OSA were included in this study. PSGs were not included if a split-night treatment protocol (diagnostic to PAP therapy) was implemented or if a primary PSG channel (nasal pressure, pulse oximetry, all EEG, respiratory effort) contained too much artifact for reliable analysis. The institutional Human Research Ethics Committee approved this study.

Polysomnography

PSGs were recorded with the E-series acquisition system (Compumedics, Abbotsford, Australia). The recording montage comprised of EEG (F4-M1, C4-M1, O2-M1), left and right EOG (recommended derivation: E1-M2, E2-M2), chin electromyogram (EMG, mental/submental positioning), modified lead II ECG, nasal pressure (DC amplified), oronasal thermocouple, body position, thoracic and abdominal effort (inductive plethysmography), pulse oximetry (Nonin Xpod 3011), left and right leg movement (anterior tibialis EMG), and sound pressure (dBA meter: Tecpel 332).

Polysomnogram Scoring Protocol

PSGs were de-identified and all previous respiratory event scoring was removed. PSGs were then presented to 2 scorers (BD, JM) in random order and scored using either AASM$_{2012}$, AASM$_{2007Rec}$, AASM$_{2007Alt}$, or AASM$_{1999}$ criteria. The same scorer therefore scored each PSG 4 times for respiratory events. Randomization of PSG's was performed using the freely accessible Randomizer website.[9] To calculate inter-scorer reliability, a subset of 20 randomly selected PSG's were scored with each criteria by both scorers.

PSGs were scored with Compumedics Profusion 3.4 (Build 365) software while viewed on Planar PX212M (1600 × 1200 resolution) LCD monitors via ATI Radeon HD2400 Pro 256 Mb graphics cards. Each of the PSG scorers (BD, JM) have over 10 years' experience in scoring PSGs and participate regularly in intra- and inter-laboratory scoring concordance activities.

Statistical Analysis

Statistical analyses were performed using GraphPad Prism 6.02 (GraphPad Software, La Jolla, CA). The normality of data was determined by applying the DÁgostino-Pearson omnibus K2.[10] Data are presented as mean and standard error or median and interquartile range where appropriate. Group data were compared using the Friedman test with Dunn posttest for multiple comparisons.[11] Inter-scorer reliability for respiratory events was assessed using the percentage of AHI agreement between the scorers. AHI and hypopnea index (HI) agreement between the hypopnea criteria were determined according to the method of Bland and Altman.[12] Detailed analysis of hypopnea scoring was undertaken in all PSGs to determine the contribution of each rule change of the scoring criteria to differences in AHI between AASM$_{2012}$ and AASM$_{2007Rec}$, AASM$_{2007Alt}$, and AASM$_{1999}$, respectively. These contributions were calculated as the proportion of total HI differences and the median (interquartile range) change in AHI.

Contingency tables for prevalence of OSA diagnoses at each of the standard levels of OSA severity (mild: ≥ 5/h, moderate: ≥ 15/h, and severe: ≥ 30/h) were tabulated for each criterion. These criteria were then compared to AASM$_{2012}$ at each severity level using the McNemar test. Equivalent AHIs at these severity levels were calculated using receiver operator curves (ROC) to compare the AASM$_{2012}$ definition to each of the previous hypopnea definitions (AASM$_{2007Rec}$, AASM$_{2007Alt}$,

and AASM$_{1999}$). Threshold values were chosen to give the best combination of sensitivity and specificity.

To determine if AHI data calculated from the previous hypopnea definitions (AASM$_{2007Rec}$, AASM$_{2007Alt}$, and AASM$_{1999}$) can be reliably converted to AASM$_{2012}$, we fitted a regression line (with 95% confidence interval) and then calculated prediction intervals for each of the previous hypopnea definitions to the AASM$_{2012}$ definition. The prediction interval provides a 95% confidence estimate of the interval in which future observations will lie with respect to the different AHI criteria.[13] Values were plotted with the gold standard values (previous hypopnea definitions) as the abscissa and the AASM$_{2012}$ values as the ordinate. CUSUM linearity criteria was utilized to consider suitability for Passing-Bablok linear regression.[14] Where data did not meet the linearity criteria, the Emancipator-Kroll stepwise polynomial regression method[15] was employed. In all aspects of analysis, a $p < 0.05$ was considered statistically significant.

RESULTS

Patient Cohort

Patient demographic and sleep data are presented in **Table 1**. Patients were middle-aged and obese. There was a male gender bias in this cohort, reflecting the reported gender-based prevalence in OSA. The comorbidity profile of this patient cohort was also representative of the OSA population. The sleep of the study cohort was characterized by reduced sleep efficiency, reduced proportions of N3 and REM sleep, as well as an increase in the proportion of N1 sleep. The EEG arousal index (ArI) was also increased above the normal range. Patients generally identified themselves as being mildly somnolent.

AHI Agreement

The median AHI and HI for each scoring criteria are presented in **Table 2**. All AHI and HI values were statistically different from each other. The AASM$_{2012}$ criteria resulted in a median (interquartile range) percentage increase in AHI of 77% (42, 148) and 17% (6, 33) compared to AASM$_{2007Rec}$ and AASM$_{2007Alt}$, respectively. In comparison to AASM$_{1999}$, the AASM$_{2012}$ criteria resulted in a 9% (3, 21) decrease in the AHI. The relative contribution of hypopneas to the AHI was 96% (86, 99), 89% (65, 97), 94% (82, 99), and 96% (87, 99) for AASM$_{2012}$, AASM$_{2007Rec}$, AASM$_{2007Alt}$, and AASM$_{1999}$, respectively. Inter-scorer reliability for the scoring of respiratory

events was 0.89 ± 0.03 (mean ± SEM), 0.91 ± 0.04, 0.90 ± 0.03, and 0.87 ± 0.04 for AASM$_{2012}$, AASM$_{2007rec}$, AASM$_{2007alt}$, and AASM$_{1999}$, respectively. There was no significant difference in inter-scorer reliability between each scoring criteria.

Bland-Altman plots (as shown in **Figure 1**) were created to determine the level of agreement between AASM$_{2012}$ criteria and AASM$_{2007Rec}$, AASM$_{2007Alt}$, and AASM$_{1999}$, respectively for AHI. There was a median (IQR) AHI bias of 9.1 (3.6, 13.8) events/h and 2.6 (0.8, 6.8) events/h towards the AASM$_{2012}$ criteria when compared to AASM$_{2007Rec}$ and AASM$_{2007Alt}$ respectively. Considerable variability in bias between AASM$_{2007Rec}$ and AASM$_{2012}$ was observed. There was a median (IQR) AHI bias of 1.9 (0.8, 3.7) towards AASM$_{1999}$ compared to AASM$_{2012}$ criteria.

Table 1—Patient characteristics of the study cohort.

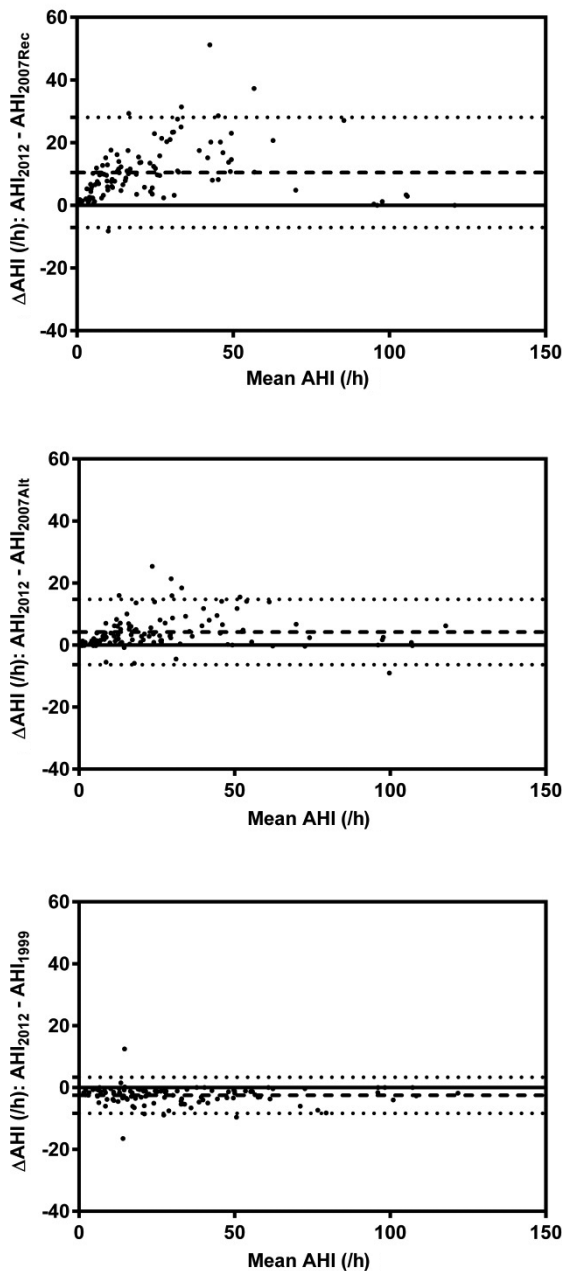| Parameter	
Number	112
Age, years	55 (41, 68)
BMI, kg/m^2	31.8 (28.6, 35.9)
ESS	11 (5, 16)
Gender, M:F	68:44
Comorbidities, number of patients	
Hypertension	52
Type 2 diabetes	26
GERD	26
Hyperlipidemia	22
Ischemic heart disease	20
Depression	18
TST, min	289 (240.5, 346.9)
Sleep latency, min	19.8 (9.4, 47.3)
Sleep efficiency, %	65.8 (53.3, 78.9)
Sleep stage proportions, % of TST	
NREM Total	82.2 (77.1, 88.8)
N1	16.1 (11.1, 25.5)
N2	51.2 (42.3, 58.4)
N3	10.4 (3.9, 17.8)
R	17.9 (11.4, 22.9)
ArI, events/h	21.1 (12.5, 31.8)
Mean SpO$_2$, %	92 (90, 94)
% of TST with SpO$_2$ < 90%	3.7 (0.6, 29.9)

Values are presented as median (interquartile range). BMI; body mass index, ESS; Epworth sleepiness scale, GERD; gastroesophageal reflux disease, TST; total sleep time, N1; stage 1 sleep, N2; stage 2 sleep, N3; stage 3 sleep, R; REM sleep, ArI; arousal index.

Table 2—Apnea-hypopnea and hypopnea indices according to hypopnea criteria.

Parameter	AASM$_{2012}$	AASM$_{2007Rec}$	AASM$_{2007Alt}$	AASM$_{1999}$	p value
AHI (/h)	21.5 (11.4, 42.4)	11.4 (3.4, 22.0)	17.1 (7.3, 34.5)	24.5 (13.1, 45.7)	< 0.0001
NREM AHI (/h)	18.9 (8.1, 42.6)	8.3 (2.1, 24.4)	14.0 (5.5, 37.9)	22.2 (9.2, 44.83)	< 0.0001
REM AHI (/h)	29.9 (14.2, 54.0)	16.5 (5.2, 39.0)	22.9 (10.2, 50.4)	34.0 (16.3, 59.3)	< 0.0001
HI (/h)	19.1 (10.0, 37.5)	8.1 (2.6, 18.3)	14.4 (7.1, 28.2)	21.6 (12.3, 39.7)	< 0.0001
NREM HI (/h)	16.8 (7.5, 36.5)	7.1 (1.7, 18.0)	12.6 (4.6, 28.5)	18.4 (8.5, 39.5)	< 0.0001
REM HI (/h)	24.9 (11.6, 40.1)	11.8 (3.1, 26.4)	19.7 (8.1, 34.5)	27.7 (14.6, 45.6)	< 0.0001

Values expressed as median (interquartile range). AHI; apnea-hypopnea index, HI; hypopnea index.

Figure 1—Bland-Altman plots demonstrating the level of agreement between AASM$_{2012}$ and AASM$_{2007Rec}$, AASM$_{2007Alt}$, and AASM$_{1999}$, respectively.

The dashed line represents the mean difference and the dotted lines represent the 95th percentile confidence limits.

OSA Prevalence and Diagnosis

The diagnosis and OSA severity of the patient cohort for each hypopnea criteria is presented in **Table 3**. Significant differences were found in the diagnosis of OSA when AASM$_{2012}$ criteria were compared to AASM$_{2007Rec}$ and AASM$_{2007Alt}$. No significance was found between AASM$_{2012}$ and AASM$_{1999}$. The proportion of the patient cohort to reach the OSA threshold of 5/h (assuming that each patient was symptomatic) were 90.2%, 69.6%, 83.8%, and 92.5% for AASM$_{2012}$, AASM$_{2007Rec}$, AASM$_{2007Alt}$, and AASM$_{1999}$, respectively The AASM$_{2007Rec}$ hypopnea criteria was associated with the least number of patients diagnosed with OSA.

Equivalent AHIs and Prediction Intervals

The equivalent AHIs for previous hypopnea definitions to achieve the same OSA prevalence rates as AASM$_{2012}$ at thresholds of 5/h, 15/h, and 30/h are shown in **Table 4**. To achieve a similar level of OSA prevalence as AASM$_{2012}$, AASM$_{2007Rec}$ would have to shift its OSA threshold down to approximately 2.6/h. Similarly, for AASM$_{2007Alt}$ and AASM$_{1999}$, the thresholds would have to shift to approximately 3.6/h and 7.3/h, respectively. Regression curves were fitted to determine the relationship between each of the previous hypopnea criteria and AASM$_{2012}$. These are shown in **Figure 2**. The relationships between AASM$_{2007Alt}$ and AASM$_{1999}$ with AASM$_{2012}$ were successfully represented by linear regression. The relationship between AASM$_{2007Rec}$ and AASM$_{1999}$ was nonlinear and hence fitted to 2nd order polynomial regression. The calculated prediction interval between AASM$_{2012}$ and AASM$_{2007Rec}$ was much wider than the calculated prediction intervals for AASM$_{2012}$ versus AASM$_{2007Alt}$ and AASM$_{2012}$ versus AASM$_{1999}$.

DISCUSSION

In this study, we investigated the consequences of implementing the 2012 AASM hypopnea definition on measured AHI and OSA prevalence compared to other hypopnea definitions such as AASM$_{2007Rec}$, AASM$_{2007Alt}$, and AASM$_{1999}$. Our results confirm that the AHI and proportion of patients diagnosed with OSA will increase if a sleep laboratory implements the AASM$_{2012}$ criteria. This increase is more marked if the transition to AASM$_{2012}$ is implemented from AASM$_{2007Rec}$ rather than AASM$_{2007Alt}$. Our results also demonstrate that implementation of AASM$_{2012}$ produces similar AHI and diagnostic outcomes as AASM$_{1999}$.

We are only aware of one other study which has also examined the consequences of implementing AASM$_{2012}$ with respect to the AHI. Similar to our findings, BaHammam et

Table 3—Patient diagnoses according to hypopnea criteria.

Diagnosis	AASM$_{2012}$	AASM$_{2007Rec}$	AASM$_{2007Alt}$	AASM$_{1999}$
No OSA	11 (9.8%)	34 (30.4%)***	18 (16.2%)**	8 (7.5%)[NS]
Mild OSA	29 (25.9%)	31 (27.6%)***	33 (29.1%)**	26 (23.2%)[NS]
Moderate OSA	29 (25.9%)	24 (21.5%)***	29 (25.9%)**	32 (28.8%)[NS]
Severe OSA	43 (38.4%)	23 (20.5%)***	32 (28.8%)**	46 (40.5%)[NS]

Values presented as absolute number of diagnoses (% of total number of patients). McNemar's test was used to compare the prevalence of OSA severity levels to AASM$_{2012}$. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$ compared to AASM$_{2012}$. [NS] Not significant.

Table 4—Equivalent AHIs of previous hypopnea criteria ($AASM_{2007Rec}$, $AASM_{2007Alt}$, and $AASM_{1999}$) to achieve the same $AASM_{2012}$ prevalence at the severity thresholds of 5 events/h, 15 events/h, and 30 events/h.

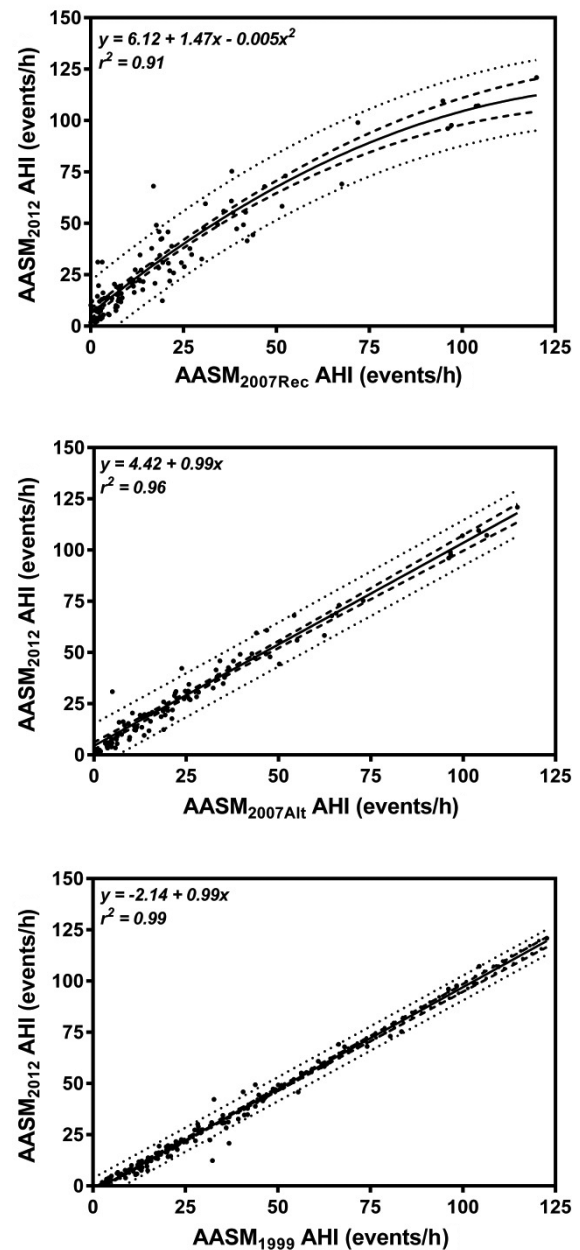| $AASM_{2012}$	$AASM_{2007Rec}$	$AASM_{2007Alt}$	$AASM_{1999}$
5/h	2.6 (87.1, 81.8)	3.6 (100, 90.9)	7.3 (96.0, 90.9)
15/h	7.2 (90.4, 92.3)	11.5 (87.7, 87.2)	17.4 (97.3, 94.9)
30/h	14.1 (93.2, 91.2)	21.1 (90.9, 89.7)	31.7 (97.7, 98.5)

Data presented as threshold (sensitivity, specificity).

al.[8] found that the $AASM_{2012}$ increases the AHI relative to the $AASM_{2007Rec}$ and $AASM_{2007Alt}$ criteria. However, the study of BaHammam demonstrated somewhat more distinct differences between criteria than we have shown in this study. We believe these differences are likely explained by variations in the cohorts that were examined in each study. Our cohort appeared be older, more somnolent, spent more time in REM sleep during their PSG, and displayed more severe SpO_2 desaturations. All these factors may contribute to differences in the type of events presented during overnight PSG.

A previous study by Ruehland et al.[5] had explored the differences between $AASM_{2007Rec}$, $AASM_{2007Alt}$, and $AASM_{1999}$. They found that implementing $AASM_{2007Rec}$ and $AASM_{2007Alt}$ resulted in median AHI reductions of 10.9/h and 6.1/h compared to $AASM_{1999}$. Our results displayed a similar trend to Ruehland, with median AHI reductions of 11.1/h and 5.7/h for $AASM_{2007Rec}$ and $AASM_{2007Alt}$ when compared to $AASM_{1999}$. Surprisingly, we were only able to find one other study to confirm these findings of Ruehland et al. Ward et al.[16] compared the $AASM_{2007Rec}$ and $AASM_{2007Alt}$ criteria in a cohort of stable heart failure patients. In similar fashion to this study and that of Ruehland, they also demonstrate a small but nonetheless significant difference between $AASM_{2007Rec}$ and $AASM_{2007Alt}$. In our study, not only do we confirm the findings of Ruehland, we also demonstrate that $AASM_{2012}$ brings the AHI towards the previous $AASM_{1999}$.

Of particular interest in this study are the 23 patients who met the AHI threshold of 5/h using the $AASM_{2012}$ criteria but not the $AASM_{2007Rec}$. It is too simplistic to conclude that the $AASM_{2012}$ criteria are more sensitive than $AASM_{2007Rec}$ in diagnosing patients with OSA, as diagnosis involves more than assessment of numerical thresholds. For the diagnosis of OSA, the International Classification of Sleep Disorders, Third Edition[17] (ICSD-3) has a two-tiered system. OSA can be diagnosed in the absence of clinical symptoms if the number of events per hour is greater than 15. On the other hand, if the number of events lies between 5/h and 15/h, clinical symptoms are required in order to satisfy the diagnostic criteria for OSA. Applying the ICSD-3 diagnostic criteria to our cohort, only three of these 23 patients did not reach the threshold of 5/h using $AASM_{2007Rec}$ but achieved the threshold of 15/h using $AASM_{2012}$ criteria. These three did not require the presence of symptoms to have a diagnosis of OSA. Of the other 20 patients, five had an Epworth Sleepiness Scale above the normal range and thus could be considered symptomatic. This highlights the need for clinical interpretation associated with

Figure 2—Least squares regression demonstrating the apnea-hypopnea index relationships between $AASM_{2012}$ and $AASM_{2007Rec}$, $AASM_{2007Alt}$ and $AASM_{1999}$, respectively.

$$y = 6.12 + 1.47x - 0.005x^2$$
$$r^2 = 0.91$$

$$y = 4.42 + 0.99x$$
$$r^2 = 0.96$$

$$y = -2.14 + 0.99x$$
$$r^2 = 0.99$$

The unbroken line represents the line of best fit while the dashed lines represent the 95th percentile confidence intervals. The dotted lines represent the upper and lower bounds of the prediction interval.

these studies rather than relying purely on the numbers produced by sleep studies.

Our decision to compare $AASM_{2012}$ with both $AASM_{rec}$ and $AASM_{alt}$ was recognition that either of these criteria are used by various clinical sleep laboratories. While $AASM_{2007Rec}$ criteria are the predominant criteria in use, other countries, such as Australia and New Zealand have adopted the $AASM_{2007Alt}$ definition. Thus to increase the utility of this study we felt it was necessary to provide comparisons to both criteria. Based on our findings, the transition from $AASM_{2007Rec}$ to $AASM_{2012}$

Table 5—The contribution of specific rule changes to the AHI between previous hypopnea criteria and AASM$_{2012}$.

Criteria Conversion	Rule Change	Contribution	Δ AHI (/h)
AASM$_{2007Rec}$ to AASM$_{2012}$	≥ 4% to ≥ 3% desaturation	59%	4.4 (1.8, 9.6)
	EEG Arousal without ≥ 3% desaturation	39%	3.8 (2.0, 6.2)
	9+1 to 10 sec duration	2%	0.1 (0.0, 0.3)
AASM$_{2007Alt}$ to AASM$_{2012}$	50% to 30% flow reduction	98%	2.1 (0.9, 6.2)
	9+1 to 10 sec duration	2%	0.0 (0.0, 0.4)
AASM$_{1999}$ to AASM$_{2012}$	50% flow reduction without consequence	100%	1.9 (0.8, 3.7)

Apnea-hypopnea index values expressed as median (interquartile range). AHI; apnea-hypopnea index.

will have a larger impact on sleep laboratories and health systems. A crude extrapolation from our small cohort points to a further 18% and 5% of the referred population will be diagnosed with OSA in laboratories currently using AASM$_{2007Rec}$ and AASM$_{2007Alt}$, respectively. This resulting increase in OSA diagnoses will lead to increases in the demand for sleep medicine services. While this may increase the initial healthcare costs associated with OSA diagnosis, there is good evidence to show that these costs are compensated by reducing health care utilization in other disorders linked to OSA.[18]

The implementation of new hypopnea rules, however, presents a number of difficulties for a clinical sleep laboratory. Perhaps the most difficult problem is how does a clinician interpret the PSG results under the new rules with respect to the patients past PSGs scored under older rules? To clarify this problem we explored if a correction factor could be applied to "translate" older results to AASM$_{2012}$. From our calculated prediction intervals (see **Figure 2**) in this study, we believe that a correction factor could be used reliably to translate AASM$_{2007Alt}$ and AASM$_{1999}$ to AASM$_{2012}$. However a large prediction interval between AASM$_{2007Rec}$ and AASM$_{2012}$ (see **Figure 2**) would indicate that translation between these two criteria is not recommended.

In this study we also examined the contribution of rule changes to differences in the hypopnea index (**Table 5**). When we examined the implementation of AASM$_{2012}$ from AASM$_{2007Rec}$, we found that the inclusion of EEG arousals contributed almost equally to hypopnea index differences as the decrease in SpO$_2$ desaturation requirement. This has wider implications for quality assurance processes should a laboratory decide to the AASM$_{2012}$ hypopnea criteria. Since the scoring of EEG arousals typically shows poorer inter-scorer reliability,[19] irrespective of the AASM montage utilized,[20] the reliability of scoring hypopneas may decrease with AASM$_{2012}$. Nonetheless, we believe that the inclusion of EEG arousals in the hypopnea criteria is a positive step forward. Patients are often referred by primary care physicians because of neurocognitive deficits (excessive daytime somnolence, poor concentration, and changes in mood) often related to OSA. These symptoms and subsequent resolution with therapy can occur in the absence of SpO$_2$ desaturations as demonstrated by the study of Guilleminault et al.[21] Thus the inclusion of EEG arousals, while controversial in some quarters, would appear to provide a direct link between OSA and its associated neurocognitive deficits.

Despite the relative importance of our findings, our study is not without its limitations. First, we have used a relatively homogenous cohort to examine these scoring differences. This has the advantage of examining the different scoring rules in the likely absence of other sleep disorders which may potentially confound the results. On the other hand, this reduces the utility in groups where comorbid disorders such as sleep hypoventilation or periodic limb movement disorder are prominent. These disorders are particularly important as they can create difficulties in discerning the presence or absence of hypopneas. Another limitation of our study was that the two scorers used to score PSGs in this study were from the same laboratory. This may limit the generalizability of the study to other laboratories since, despite the improvements in defining the scoring criteria, each laboratory will still have their own way of dealing with specific PSG scenarios that has not been adequately clarified by the AASM Manual. We believe that this limitation is partially mitigated by the scorers' participation in inter-laboratory concordance programs.

CONCLUSION

Our study has shown that implementation of AASM$_{2012}$ criteria will lead to increases in the AHI of patients suspected of OSA with subsequent increases in the diagnosis of the disorder. This increase in the AHI will be more pronounced for sleep laboratories currently using AASM$_{2007Rec}$. The lowering of the SpO$_2$ desaturation threshold and the inclusion of EEG arousals contribute almost equally to this AHI change.

ABBREVIATIONS

AASM, American Academy of Sleep Medicine
AASM$_{1999}$, "Chicago Criteria" hypopnea definitions
AASM$_{2007Alt}$, 2007 AASM alternate hypopnea definitions
AASM$_{2007Rec}$, 2007 AASM recommended hypopnea definitions
AASM$_{2012}$, 2012 AASM recommended hypopnea definitions
AHI, apnea-hypopnea index
ArI, EEG arousal index
ECG, electrocardiogram
EEG, electroencephalogram
EOG, electrooculogram
EMG, electromyogram

HI, hypopnea index
ICSD, International Classification of Sleep Disorders
LCD, liquid crystal display
NREM, non-rapid eye movement sleep
N1, stage 1 sleep
N2, stage 2 sleep
N3, stage 3 sleep
OSA, obstructive sleep apnea
PSA, proportion of specific agreement
PSG, polysomnography
R, rapid eye movement sleep

REFERENCES

1. American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. Sleep 1999;22:667–89.
2. Kushida CA, Littner MR, Morgenthaler T, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 2005;28:499–521.
3. Iber C, Ancoli-Israel S, Chesson AL, Quan, S. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specification. 1st ed. Westchester, IL: American Academy of Sleep Medicine, 2007.
4. Grigg-Damberger MM. The AASM scoring manual: a critical appraisal. Curr Opin Pulm Med 2009;15:540–49
5. Ruehland WR, Rochford PD, O'Donoghue FJ, et al. The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. Sleep 2009;32:150–7.
6. Thornton A, Ruehland W, Duce B, et al. ASTA/ASA Commentary on AASM Manual for the Scoring of Sleep and Associated Events. Available at: http://www.sleep.org.au/documents/item/217. Accessed: January 27, 2014.
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. J Clin Sleep Med 2012;8:597–619.
8. BaHammam A, Obeidat A, Barataman K, et al. A comparison between the AASM 2012 and 2007 definitions for detecting hypopnea. Sleep Breath 2014;18:767–73.
9. Urbaniak GC, Plous S. Research Randomizer (Version 3.0) [Computer software]. Retrieved on February 11, 2013, from http://www.randomizer.org/
10. D'Agostino RB, Pearson ES. Tests for departure from normality. Empirical results for the Distributions of b2 and √ b1. Biometrika 1973;60:613–22
11. Dunn OJ. Multiple comparisons among means. J Am Stat Assoc 1961;56:54–64.
12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
13. Olive DJ. Prediction intervals for regression models. Comput Stat Data Anal 2007;51:3115–22.
14. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. J Clin Chem Clin Biochem 1983;21:709–20.
15. Kroll MH, Emancipator K. A theoretical evaluation of linearity. Clin Chem 1993;39:405–13.
16. Ward NR, Roldao V, Cowie MR, et al. The effect of respiratory scoring on the diagnosis and classification of sleep disordered breathing in chronic heart failure. Sleep 2013;36:1341–8.
17. American Academy of Sleep Medicine. International classification of sleep disorders, 3rd ed. Darien, IL: American Academy of Sleep Medicine, 2014.
18. Albarrak M, Banno K, Sabbagh AA, et al. Utilization of healthcare resources in obstructive sleep apnea syndrome: a 5-year follow-up study in men using CPAP. Sleep 2005;28:1306–11.
19. Loredo JS, Clausen JL, Ancoli-Israel S, et al. Night-to-night arousal variability and interscorer reliability of arousal measurements. Sleep 1999;22:916–20.
20. Duce B, Rego C, Milosavljevic J, et al. The AASM recommended and acceptable EEG montages are comparable for the staging of sleep and scoring of EEG arousals. J Clin Sleep Med 2014;10:803–9.
21. Guilleminault C, Hagen CC, Huynh NT. Comparison of hypopnea definitions in lean patients with known obstructive sleep apnea hypopnea syndrome (OSAHS). Sleep Breath 2009;13:341–7.

SUBMISSION & CORRESPONDENCE INFORMATION

DISCLOSURE STATEMENT