The New AASM Criteria for Scoring Hypopneas: Impact on the Apnea Hypopnea Index

Warren R. Ruehland, BSc(Hons)[1]; Peter D. Rochford, BAppSc, Grad Dip Bio Instr[1]; Fergal J. O'Donoghue, MD, PhD[1]; Robert J. Pierce, MD[1]; Parmjit Singh, BSc(Hons), MBA[2]; Andrew T. Thornton, PhD[2]

[1]Institute for Breathing and Sleep, Austin Health, Heidelberg, Victoria, Australia; [2]Royal Adelaide Hospital, Adelaide, South Australia, Australia

Study Objectives: To compare apnea-hypopnea indices (AHIs) derived using 3 standard hypopnea definitions published by the American Academy of Sleep Medicine (AASM); and to examine the impact of hypopnea definition differences on the measured prevalence of obstructive sleep apnea (OSA).

Design: Retrospective review of previously scored in-laboratory polysomnography (PSG).

Setting: Two tertiary-hospital clinical sleep laboratories.

Patients or Participants: 328 consecutive patients investigated for OSA during a 3-month period.

Interventions: N/A

Measurements and Results: AHIs were originally calculated using previous AASM hypopnea scoring criteria ($AHI_{Chicago}$), requiring either > 50% airflow reduction or a lesser airflow reduction with associated > 3% oxygen desaturation or arousal. AHIs using the "recommended" (AHI_{Rec}) and the "alternative" (AHI_{Alt}) hypopnea definitions of the *AASM Manual for Scoring of Sleep and Associated Events* were then derived in separate passes of the previously scored data. In this process, hypopneas that did not satisfy the stricter hypopnea definition criteria were removed. For AHI_{Rec}, hypopneas were required to have ≥ 30% airflow reduction and ≥ 4% desaturation; and for AHI_{Alt}, hypopneas were required to have ≥ 50% airflow reduction and ≥ 3% desaturation or arousal. The median AHI_{Rec} was approximately 30% of the median $AHI_{Chicago}$, whereas the median AHI_{Alt} was approximately 60% of the $AHI_{Chicago}$, with large, AHI-dependent, patient-specific differences observed. Equivalent cut-points for AHI_{Rec} and AHI_{Alt} compared to $AHI_{Chicago}$ cut-points of 5, 15, and 30/h were established with receiver operator curves (ROC). These cut-points were also approximately 30% of $AHI_{Chicago}$ using AHI_{Rec} and 60% of $AHI_{Chicago}$ using AHI_{Alt}. Failure to adjust cut-points for the new criteria would result in approximately 40% of patients previously classified as positive for OSA using $AHI_{Chicago}$ being negative using AHI_{Rec} and 25% being negative using AHI_{Alt}.

Conclusions: This study demonstrates that using different published standard hypopnea definitions leads to marked differences in AHI. These results provide insight to clinicians and researchers in interpreting results obtained using different published standard hypopnea definitions, and they suggest that consideration should be given to revising the current scoring recommendations to include a single standardized hypopnea definition.

Keywords: Hypopnea, obstructive sleep apnea, sleep disordered breathing, polysomnography, methodology, scoring, diagnosis.

Citation: Ruehland WR; Rochford PD; 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.

POLYSOMNOGRAPHY (PSG) IS PERFORMED FOR A WIDE VARIETY OF INDICATIONS, MOST COMMONLY FOR INVESTIGATION OF OBSTRUCTIVE SLEEP APNEA (OSA). OSA is characterized by repeated episodes of upper airway obstruction resulting in cessation (apnea) or reduction (hypopnea) in airflow during sleep. The apnea hypopnea index (AHI), a count of the number of apneas and hypopneas per hour of sleep, is the key measure used for case identification, for quantifying disease severity, and for defining disease prevalence in normal and clinical populations.

Despite the importance of this measure, inter-laboratory variations in apnea and, in particular, hypopnea definition have been reported.[1,2] Differences in hypopnea definition relate to the degree of airflow reduction and/or oxygen desaturation required and the requirement for associated EEG arousal. The effect of varying definitions of hypopnea on AHI has been examined in a number of studies[3-6] and the importance of standardizing the hypopnea definition, and thereby reducing inter-laboratory variability in AHI, has been recognized.[3,5,7,8]

In 1999, the American Academy of Sleep Medicine (AASM) produced a consensus report,[9] targeted at clinical research rather than clinical practice,[10] recommending standardized scoring criteria for a range of respiratory events. These guidelines (also known as "Chicago Criteria") described 2 types of hypopneas: (i) Those with a > 50% decrease in a valid measure of airflow without a requirement for associated oxygen desaturation or arousal, and (ii) Those with a lesser airflow reduction in association with oxygen desaturation of > 3% or an arousal.

The lack of clinical practice guidelines was addressed in 2001 when the AASM, via the Clinical Practices Review Committee, published a position paper[11] which described a hypopnea as an abnormal respiratory event lasting ≥ 10 sec with ≥ 30% reduction in thoracoabdominal movement or airflow, and with ≥ 4% oxygen desaturation. This is currently the approved hypopnea definition for the Centers for Medicare and Medicaid Services in the United States to determine eligibility for treatment funding.[12] Nevertheless, in 2005 the AASM, via the Practice Parameters Committee, reported that, "Several clinical definitions of hypopnea are in clinical use and there is no clear consensus."[13]

In a further attempt to improve standardization, the AASM recently published the *Manual for the Scoring of Sleep and Associated Events*.[14] In this manual there is a "recommended" and

Submitted for publication June, 2008
Submitted in final revised form November, 2008
Accepted for publication November, 2008
Address correspondence to: Warren Ruehland, Institute for Breathing and Sleep, Austin Health, Ground Floor Bowen Centre, Studley Road, Heidelberg VIC 3084, Australia; Tel: +61 3 9496 3528; Fax: +61 3 9496 5124; E-mail address: Warren.Ruehland@austin.org.au

an "alternative" hypopnea definition; and either can be used at the discretion of the clinician or investigator. The recommended definition is the same as the definition published in the AASM 2001 position paper: hypopnea scoring requires ≥ 30% reduction in nasal pressure signal excursions from baseline and associated ≥ 4% desaturation from pre-event baseline. The alternative definition requires ≥ 50% reduction in nasal pressure signal excursions and associated ≥ 3% desaturation or arousal.

Introduction of new standards is likely to lead to a period when individual laboratories assess and change their practices and when different laboratories use different methodologies. This study aims to assist in interpretation of clinical or research results in this setting. Specifically, this study examines the impact of the 2 recently published hypopnea definitions on the AHI, compared to the previously published "Chicago" hypopnea definition, and subsequently examines the impact on the measured prevalence of OSA in a cohort of patients presenting for diagnosis or exclusion of obstructive sleep apnea. Compared to similar previous studies[3-6] it is unique in its focus on published standard hypopnea definitions.

METHODS

Patient Selection

This study utilized 328 consecutive diagnostic in-laboratory PSGs from 2 separate patient groups recorded during a 3-month period, between April and June 2007. One group consisted of 164 PSGs that were sourced from the Austin Health sleep laboratory in Melbourne, Victoria, Australia. The second group consisted of 164 PSGs that were sourced from the Royal Adelaide Hospital sleep laboratory in Adelaide, South Australia. Both laboratories are located on the southern coast of mainland Australia at altitudes < 200 meters above sea level. All patients were being investigated for clinically suspected OSA or exclusion of OSA. PSGs were not considered if they were being conducted for research purposes or investigation of respiratory failure, involved implementation or review of treatment, or involved use of supplemental oxygen. PSGs were excluded if they were technically poor, which was defined as one key signal (oxygen saturation, nasal pressure, EEG) uninterpretable ≥ 50% of sleep time.

PSG Recordings

For the Melbourne group, PSGs were recorded using Compumedics E-series or S-Series equipment (Abbotsford, Victoria, Australia), using a recording montage consisting of C3/A2 EEG, left and right EOG, ECG, submental EMG, airflow (nasal pressure), body position, thoracic and abdominal excursion (inductance plethysmography), oxygen saturation (finger pulse oximetry), left and right leg movement (piezoelectric sensors), and sound. For the Adelaide group, PSGs were recorded using Compumedics E-series equipment, using the same recording montage as above. Oximetry at Austin Health was measured using a Nellcor N-595 (Nellcor Inc, Boulder, CO, USA) set to shortest averaging time (2-4 seconds) and sampled at 5 Hz, and at the Royal Adelaide Hospital using a MasimoSET Radical (Masimo, Irvine, CA, USA) set to an averaging time of 8 seconds and sampled at 1 Hz.

PSG Scoring

PSGs were first scored in a clinical setting using Profusion PSG 2 software (Compumedics, Abbotsford, Victoria, Australia) by one of multiple scorers (12 Melbourne group; 3 Adelaide group), all of whom participated in intra- and inter-laboratory scoring concordance programs. For both groups scoring was based on published standards for sleep[15] and arousal scoring.[16] Respiratory event scoring was based on Chicago criteria[9] for determination of the AHI ($AHI_{Chicago}$), in which a hypopnea was defined by the presence of the first or second of the following criteria, plus the third:

1. A clear decrease (> 50%) from baseline in a valid measure of breathing during sleep (a valid measure was considered to be nasal pressure or summed thoracic plus abdominal respiratory inductance plethysmography). Baseline was defined as the mean amplitude of stable breathing and oxygenation in the 2 minutes preceding onset of the event (in individuals with a stable breathing pattern during sleep) or the mean amplitude of the 3 largest breaths in the 2 preceding minutes where breathing pattern was unstable.
2. A clear (discernable) amplitude reduction of a validated measure of breathing during sleep, not reaching the above criterion but associated with either an oxygen desaturation of > 3% or an arousal.
3. The event lasts ≥ 10 sec.

An apnea was defined as a complete cessation of airflow, measured using nasal pressure, for ≥ 10 sec. Respiratory effort related arousals (RERAs) were not scored.

Protocol

For each patient group, a single investigator retrospectively reviewed the PSGs previously scored using Chicago criteria, removed hypopneas that did not meet the new stricter AASM criteria[14] and recalculated the AHI. AHIs using the new recommended hypopnea definition (AHI_{Rec}) and using the alternative definition (AHI_{Alt}) were derived in separate passes of the recorded data. Derivation of AHI_{Rec} required exclusion of hypopneas with < 30% reduction in nasal pressure signal excursions or with < 4% oxygen desaturation. For AHI_{Alt} derivation, hypopneas with < 50% reduction in nasal pressure signal excursions or with < 3% desaturation and without associated arousal were excluded. Respiratory events previously classified as apneas were left unaltered.

This study was approved by the Austin Health Human Research Ethics Committee.

Analysis and Statistics

AHI is the key metric used in research and clinical practice and therefore it was the primary variable used for analysis of the differences between hypopnea definitions.

The distribution of the differences between each pair of hypopnea indices (HI) and therefore AHIs was highly skewed; therefore, nonparametric analyses were undertaken. AHIs and HIs were described according to their median and interquartile values. The significance of differences between the 3 AHIs and HIs was examined using the Friedman test; with pair-wise post

Table 1—Summary of Patient Characteristics and Polysomnography Results

Parameter	Value
n	323
Age	51 (41, 60)
Sex M/F	211/112
Body mass index (kg/m^2)	32.7 (28.7, 37.8)*
ESS	8.0 (5.0, 13.0)†
Total sleep time (min)	331.7 (280.5, 377.5)
Sleep efficiency (%)	76.6 (65.3, 84.8)
Stage 1 (min)	18.1 (12.1-34.0)
Stage 2 (min)	177.5 (140.5, 220.3)
Slow wave sleep (min)	67.0 (31.0, 93.0)
REM (min)	50.1 (30.0, 70.1)
Arousal index (/h)	22.1 (14.2, 37.0)

Values are median (interquartile range).
Explanation of symbols;*: n = 321; †: n = 305.

hoc comparisons performed using the Wilcoxon signed-rank test. Scatter plots and Bland-Altman plots[17] were constructed for inspection of the association and agreement between AHIs. The strength of association was determined with Spearman rank test. From the Bland-Altman plots agreement was reported as the median difference (bias) between scoring criteria and the 5th and 95th percentiles of the difference.

A frequency table was constructed illustrating the point prevalence of OSA for each methodology at AHI cut-points for diagnosis of 5, 15, and 30/h. For each cut-point, a 2 x 2 table compared positive and negative frequencies of each pair of AHIs (not shown) and the significance of the differences in OSA diagnosis was established using the McNemar test.

Equivalent cut-points for the 3 AHIs were established with the use of receiver operator curves (ROC) with $AHI_{Chicago}$ treated as the gold standard. Equivalent cut-points were chosen by giving equal weight to maximizing both sensitivity and specificity.

To improve understanding of which aspects of the scoring criteria contribute to the differences in AHI, more detailed analysis was undertaken of a subset of 20 PSGs. These were selected to represent a wide range of OSA severity. In these 20 PSGs, for hypopneas which were excluded based on either or both of the new AASM criteria, the degree of desaturation, arousal association, and degree of airflow reduction was manually determined. The contribution of these criteria to AHI differences was quantified as (a) a percentage relative to all hypopnea scoring differences between methods, and (b) the resultant median (interquartile range) change in AHI.

RESULTS

Patient Characteristics

Table 1 summarizes the patient characteristics and polysomnography results. Of the 328 PSGs assessed, 3 PSGs were excluded because of a poor nasal pressure signal, and 2 were excluded because of a poor oxygen saturation signal, leaving a total of 323 PSGs for analysis.

Table 2—Apnea Hypopnea Indices and Hypopnea Indices Using Different Hypopnea Scoring Criteria

Hypopnea Definition	AHI (/h)	HI (/h)
Chicago	25.1 (11.1, 48.5)	16.3 (7.8, 26.4)
Recommended	8.3 (2.1, 26.4)	2.2 (0.5, 6.6)
Alternative	14.9 (5.5, 37.4)	7.2 (2.4, 15.0)

Values are median (interquartile range). $P < 0.001$ for all pair-wise comparisons

Agreement between Scoring Criteria

Table 2 shows median AHI and HI according to the different scoring criteria. The Friedman test and post hoc comparisons revealed that all AHIs and all HIs were significantly different from each other ($P < 0.001$). The median (interquartile range) percentage contribution of hypopneas to the AHI was 82.3% (54.3, 95.5), 38.9% (12.5, 72.8), and 64.7%, (32.7, 88.9) for $AHI_{Chicago}$, AHI_{Rec}, and AHI_{Alt}, respectively.

Scatter plots (Figure 1) and Bland-Altman plots (Figure 2) were constructed for the inspection of association and agreement between each AHI pair. They demonstrated large patient-specific differences in the impact on AHI for each comparison. The strongest correlation was observed between $AHI_{Chicago}$ and AHI_{Alt} and the weakest between $AHI_{Chicago}$ and AHI_{Rec} (Figure 1).

The Bland Altman plots demonstrate a median (5th, 95th percentiles) reduction of 10.9/h (1.2, 34.3) when comparing AHI_{Rec} to $AHI_{Chicago}$, a median reduction of 6.1/h (0.5, 21.1) when comparing AHI_{Alt} to $AHI_{Chicago}$, and a median reduction of 3.5/h (0.0, 20.1) when comparing AHI_{Rec} to AHI_{Alt}. All plots show that variation of differences is dependent on the magnitude of AHI. This pattern was most pronounced when comparing AHI_{Rec} with $AHI_{Chicago}$.

OSA Diagnosis

Differences in the frequency of OSA diagnosis using different scoring methods and AHI thresholds are shown in Table 3. Significant differences in the point prevalence of diagnosis of OSA were found for all pair-wise comparisons for all thresholds examined ($P < 0.001$). Using AHI_{Rec}, 36%, 43%, and 48% of patients previously classified as positive for OSA using $AHI_{Chicago}$ would now be negative with AHI cut-offs of 5, 15, and 30/h, respectively. Using AHI_{Alt}, 17%, 26%, and 25% of patients previously classified as positive for OSA using $AHI_{Chicago}$ would now be negative.

Table 4 shows equivalent AHIs for both AHI_{Rec} and AHI_{Alt} compared to $AHI_{Chicago}$. These values were established with the use of ROC curves with $AHI_{Chicago}$ treated as the "gold standard." The equivalent cut-points were chosen by giving equal weight to maximizing both sensitivity and specificity. Using AHI_{Rec} resulted in equivalent AHI cut-points that were approximately 30% of $AHI_{Chicago}$ whereas using AHI_{Alt} resulted in AHI cut-points that were 60% of $AHI_{Chicago}$. Equivalent cut-points for AHI_{Rec} were approximately 50% of those for AHI_{Alt}. The sensitivity and specificity (Table 4) indicate the confidence in the equivalent cut-points in characterizing OSA. Sensitivity and

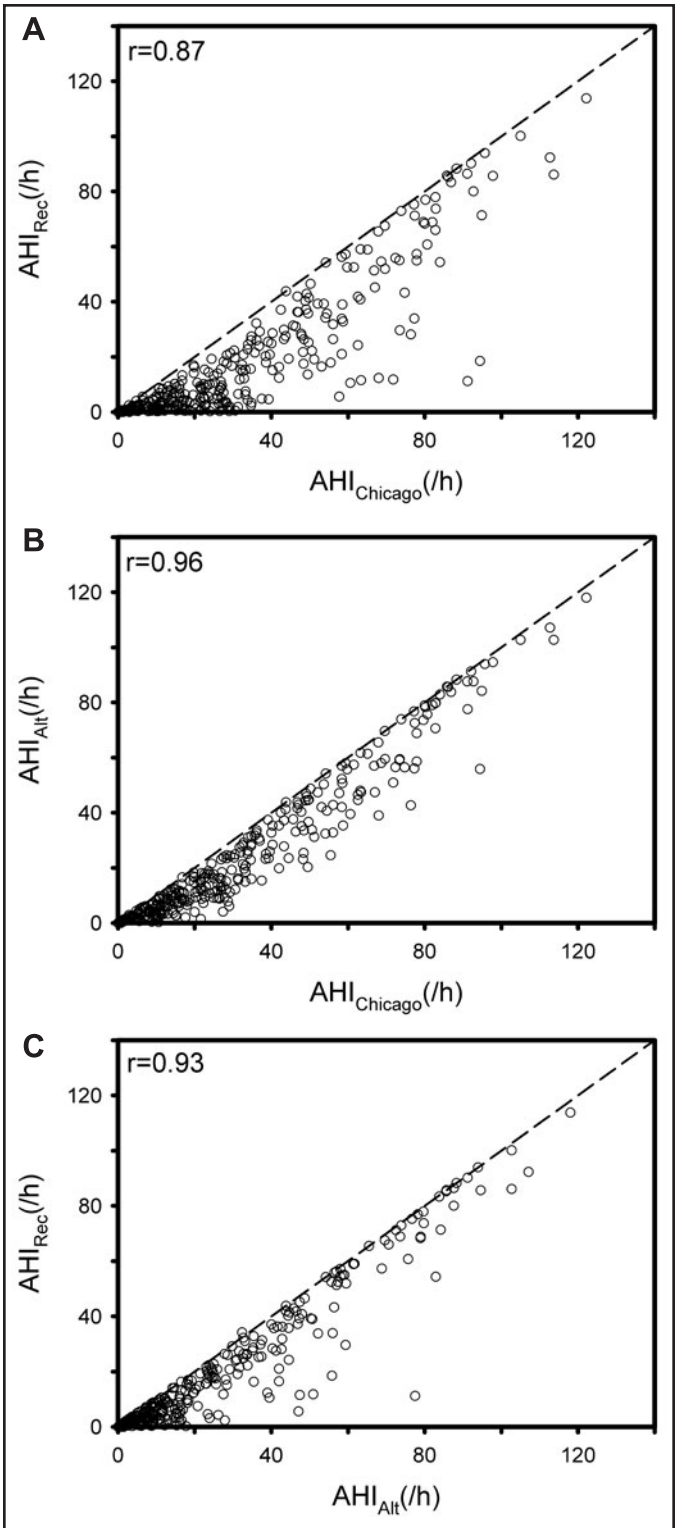

Figure 1—Scatter plots comparing: (A) AHI_{Rec} to $AHI_{Chicago}$, (B) AHI_{Alt} to $AHI_{Chicago}$, and (C) AHI_{Rec} to AHI_{Alt}. All correlations are significant at the 0.01 level (one-tailed).

specificity were greater when comparing AHI_{Alt} to $AHI_{Chicago}$ than when comparing AHI_{Rec} to $AHI_{Chicago}$.

Contribution of Scoring Criteria to HI/AHI Differences

The 20 PSGs used to examine the contribution of various scoring criteria to the differences in HI (and therefore AHI) had median (interquartile range) hypopnea indices of 27.5/h (14.9, 50.9), 6.0/h (2.5, 10.3), and 17.6/h (7.6, 33.0) for Chicago, recommended, and alternative hypopnea definitions, respectively. Table 5 illustrates the contribution of different hypopnea scoring criteria to observed differences in HI/AHI.

When comparing events scored using the alternative hypopnea definition with those scored using the recommended definition, addition of arousal criteria to the alternative definition was solely responsible for over 50% of the differences between AHI_{Alt} and AHI_{Rec}, whereas the lower desaturation requirement ($\geq 3\%$ vs. $\geq 4\%$) was responsible for approximately 25% of the differences. In addition the larger airflow reduction requirement of the alternative definition ($\geq 50\%$ vs. $\geq 30\%$) had little impact on the AHI differences.

When comparing events scored using Chicago criteria with those scored using the recommended hypopnea definition, the stricter $\geq 4\%$ desaturation requirement of the recommended definition accounted for almost all of the reduction in AHI_{Rec} compared to $AHI_{Chicago}$. There were very few events that met the $\geq 4\%$ desaturation criteria but were excluded because they had $< 30\%$ airflow reduction.

When comparing events scored using Chicago criteria with those scored using the alternative hypopnea definition, the majority were excluded because they met neither the arousal association nor $\geq 3\%$ desaturation criteria, despite having a $\geq 50\%$ airflow reduction. Although a small proportion (17%) of events did not meet the 50% airflow reduction criterion despite meeting the desaturation or arousal association criteria, these events only occurred in few patients and therefore resulted in small changes in the median AHI.

Table 3—Percentage of Patients Classified as Positive for OSA by Method and AHI Threshold for OSA Diagnosis (n = 323)

Hypopnea Definition	AHI Cut-off (events/h)		
	≥ 5	≥ 15	≥ 30
Chicago	92%	67%	42%
Recommended	59%	38%	22%
Alternative	76%	50%	31%

$P < 0.001$ for all pair-wise comparisons at all thresholds examined.

DISCUSSION

The present study examines the impact of using 3 hypopnea definitions on AHI. The study is, to the best of our knowledge, the first to provide this type of comparison using published, standard hypopnea definitions.

The data demonstrate that large differences in AHI result when the 3 different definitions are used to score hypopneas, that there are large patient-specific differences in the impact on AHI, and that variability changes systematically with OSA severity. The data also demonstrate, based on any given cut-point value of AHI, that the different approaches for measuring AHI result in substantial differences in identifying and classifying sleep disordered breathing.

These findings provide insight to clinicians and researchers when comparing results obtained using different methodolo-

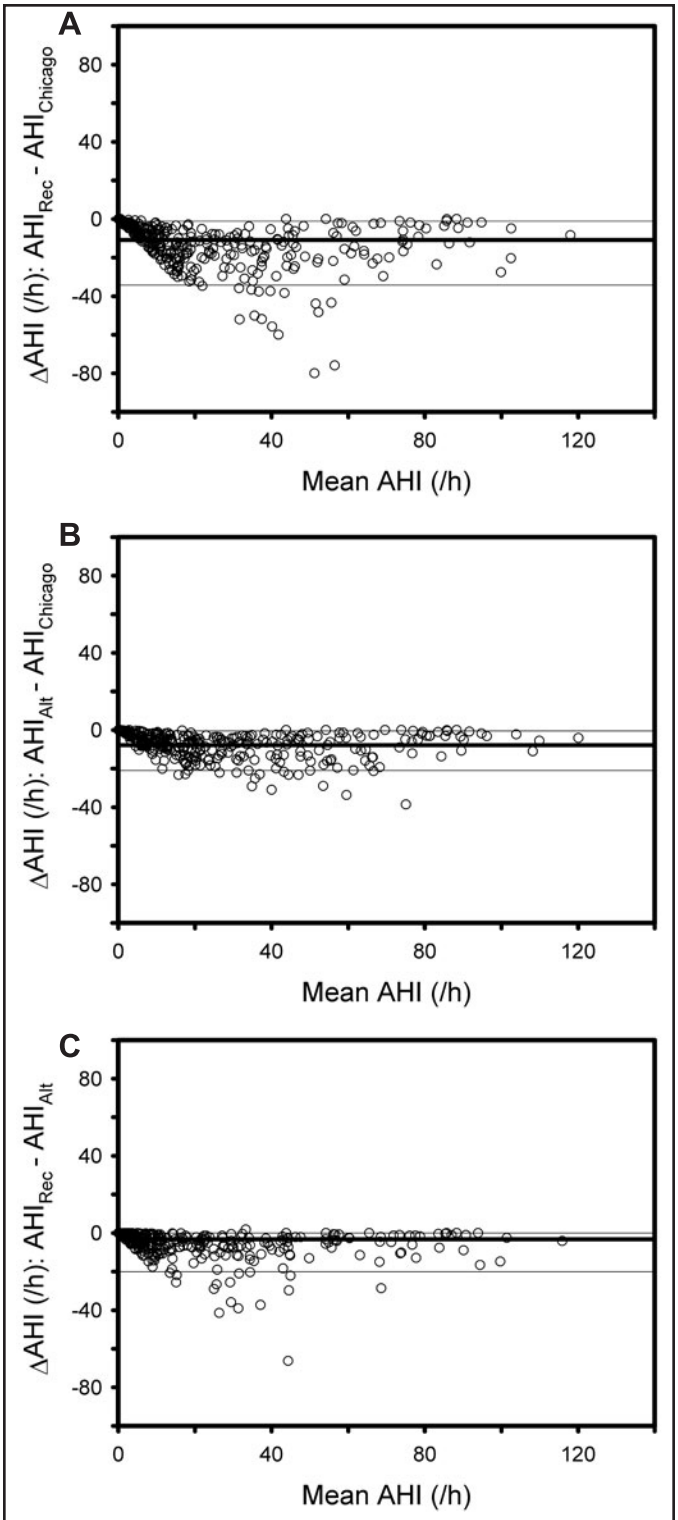

Figure 2—Bland-Altman plots illustrating agreement between AHIs: (A) AHI_{Rec} and $AHI_{Chicago}$, (B) AHI_{Alt} and $AHI_{Chicago}$, and (C) AHI_{Alt} and AHI_{Rec}. Thick solid line represents median difference. Thin solid lines represent 5th and 95th percentiles.

Table 4—Equivalent AHI by Method for Various AHI Cut-Points

	$AHI_{Chicago}$ (events/h)		
Parameter	**5**	**15**	**30**
Equivalent AHI_{Rec}	1.4	4.4	10.8
Sensitivity (%)	87.2	82.5	91.0
Specificity (%)	85.2	82.1	89.9
Equivalent AHI_{Alt}	2.8	8.9	18.4
Sensitivity (%)	91.6	90.3	95.5
Specificity (%)	92.6	90.6	95.2

gies and highlight the importance of consideration and understanding of the method used to determine AHI when comparing clinical and research studies. However, even when the method is taken into consideration, the results demonstrate that caution should be exercised in applying a simple correction factor. Sensitivity and specificity values from ROC analysis show that even if the best equivalent cut-points are utilized, variations in diagnosis will still occur.

These results also highlight the importance of standardizing methodology. Lack of standardization has numerous implications including the impact on disease identification, severity grading, and comparability of results between different laboratories and research studies. Lack of standardization may also impact treatment decisions, treatment funding by third parties, OSA prevalence estimates, estimates of the public health impact of OSA, and establishment of links between OSA and comorbidities. In research, if scoring is performed in multiple laboratories, lack of standardization and increased AHI variation can result in less robust correlation coefficients, increased sample size requirements, reduction in statistical power, and increased trial costs.[8]

The *AASM Manual for Scoring of Sleep and Associated Events*[14] suggests that either the recommended or alternative hypopnea definition can be used at the discretion of the clinician or investigator. Providing options for scoring criteria may lead to different scoring approaches between laboratories and research studies, and we have shown that this can have a substantial impact on AHI. We propose, therefore, that consideration should be given to adopting a single standardized hypopnea definition.

Determining the most appropriate hypopnea definition requires studies comparing alternative hypopnea definitions in terms of their association with physical and/or clinical outcomes such as hypertension, cardiovascular disease, sleepiness, impaired quality of life, or accidents. Although our study was not designed to investigate associations between definitions and clinical outcomes recent data from the Sleep Heart Health Study (SHHS) provides some clues in this regard. It has shown that hypopneas with 4% desaturation, but not those with lesser desaturation, are associated with increased prevalence of self-reported cardiovascular disease,[18] that hypopneas with as low as 2% desaturation are associated with increased prevalence of metabolic impairment,[19] and that association between hypopnea indices and these clinical outcomes is not improved by inclusion of arousal association criteria.[18,19] In general, these findings support the requirement of associated desaturation of ≥ 2% in hypopnea definition criteria which, in terms of desaturation, the AASM alternative definition comes closest to matching. However, a number of considerations should be kept in mind when interpreting these results including: (i) At present it is uncertain whether similar findings will be revealed for other outcomes such as sleepiness, impaired quality of life or accidents. (ii) The SHHS study used thermal sensors rather than the current

Table 5—The Contribution of Specific Hypopnea Definition Criteria to Differences in Hypopnea Scoring Between Methods (for a Subset of 20 Patients). The Contribution is Quantified as: (a) the Percentage Relative to All Hypopnea Scoring Differences Between Methods, and (b) the Resultant Median (Interquartile Range) Change in AHI (/h).

Hypopnea Definition Pair	Hypopnea scored			Criteria	%	ΔAHI(/h)
	Chic	Rec	Alt			
Alt / Rec	—	No	Yes	≥ 50% airflow reduction plus arousal	54.8	5.0 (1.3, 8.7)
	—	No	Yes	≥ 50% airflow reduction plus 3% desaturation (but not ≥ 4%)	24.1	2.2 (1.0, 4.2)
	—	No	Yes	≥ 50% airflow reduction plus arousal and 3% desaturation (but not ≥ 4%)	18.0	1.4 (0.5, 3.5)
	—	Yes	No	≥ 4% desaturation, ≥ 30% but < 50% airflow reduction	3.1	0.0 (0.0, 0.0)
Chic / Rec	Yes	No	—	< 4% desaturation	99.7	19.6 (10.5, 31.8)
	Yes	No	—	≥ 4% desaturation but < 30% airflow reduction	0.3	0.0 (0.0, 0.0)
Chic / Alt	Yes	—	No	< 3% desaturation and no arousal	83.3	6.4 (4.0, 13.9)
	Yes	—	No	Arousal or ≥ 3% desaturation but < 50% airflow reduction	16.7	0.6 (0.1, 2.0)

Chic, Alt, and Rec refer to "Chicago," "Recommended," and "Alternative" hypopnea definitions respectively.

AASM recommendation of nasal pressure to measure airflow[14] and to date, there are few published outcome studies that use nasal pressure.[20] (iii) It is also possible that the relatively lower scoring reliability demonstrated for arousal scoring and respiratory events without associated desaturation[21] compared to events with desaturation may impair the capacity to detect associations with clinical outcomes.[8]

Although other studies have examined the impact of hypopnea definition on AHI, comparisons with the current study are difficult due to methodology and hypopnea definition differences. One key point of difference is that no other published studies have used nasal pressure as the primary method for detecting airflow reduction in their comparison. Instead previous studies have detected hypopneas using thoracoabdominal movement alone[6] or a combination of thoracoabdominal movement and thermal sensors.[3-5] In addition they have either not examined different desaturation[6] or airflow reduction requirements[4,5] or, if they have examined these,[3] it has not been while utilizing hypopnea definitions similar to the new AASM criteria used in this study.

Bearing in mind the relatively small sample size, our analysis examining which aspects of the scoring criteria contribute to differences between AHIs produced a number of interesting findings. Differences in arousal association criteria played a larger role in the differences in AHI between the recommended and alternative hypopnea definitions than the difference in desaturation association criteria. This finding has implications for using the alternative hypopnea definition with portable monitoring devices that do not record EEG and therefore do not allow for the scoring of arousals. Our results indicate that using the alternative criteria with such devices is likely to result in underestimation of AHI.

Our analysis also demonstrated that there were very few events that met the recommended hypopnea definition criteria of ≥ 4% desaturation and ≥ 30% airflow reduction but did not meet the alternative definition criterion of ≥ 50% airflow reduction. This possibly suggests that hypopnea definitions with ≥ 4% desaturation requirement can have a ≥ 30% or ≥ 50% airflow reduction requirement with minimal impact on AHI.

It could be argued that scoring RERAs, which can be optionally scored using nasal pressure according to the 2007 AASM manual,[14] will capture events previously scored as hypopneas using Chicago criteria but not the newer criteria. Our results suggest that this may not be the case. We found that when using the alternative hypopnea definition as opposed to the Chicago definition, the majority of differences were due to events that had minimal (0% to 2%) desaturation and no arousal, despite having > 50% nasal pressure reduction. As these events do not have associated arousal, they will also fall outside the current criteria for RERA scoring.

With regard to methodology, in contrast to similar studies where PSGs were prospectively rescored,[3,6] in the current study a single experienced scorer in each laboratory reviewed PSGs previously scored using Chicago criteria and removed events that did not meet new criteria. This approach was feasible, as both sets of hypopnea scoring criteria published in the new AASM *Manual for scoring of Sleep and Associated Events* have stricter airflow reduction and desaturation/arousal association requirements compared to Chicago criteria and therefore represent subsets of Chicago-scored events. A limitation of this methodology is that in deriving the new AHIs the investigators were not blinded to the original scoring, leaving open the possibility of bias in determining whether to exclude scored hypopneas. In excluding hypopneas, as many as 3 decisions needed to be made: (i) degree of desaturation, (ii) presence or absence of a previous independently scored arousal, (iii) degree of airflow reduction. We contend that the first 2 of these are relatively objective decisions and are unlikely to be influenced by bias; the third decision, while less objective, was in our experience the least likely to be required.

In deciding upon preferred event definitions, as well as the validity of the scoring approach determined by the strength of associations with major morbidities, scoring reliability is an important consideration in examining the utility of scoring methods. Inter and intra-scorer reliability of the 3 scoring methods has not been addressed in this study and is therefore an area that warrants further investigation.

In scoring apneas, in the current study nasal pressure measurements were used, whereas published standards have recommended use of thermal sensors.[9,14] These recommendations mainly relate to the possibility that nasal pressure may be affected by mouth breathing and therefore may indicate an apneic event even if ventilation is apparent using thermistor.[22] In the present study, it is possible that the use of nasal pressure may have resulted in a degree of underestimation of the number of hypopneas originally scored using Chicago criteria. If anything, this would only result in an underestimation of the large differences observed between AHIs derived using different methodologies.

With regard to OSA diagnosis, the current study examined the impact of differing hypopnea definitions on the measured prevalence of OSA at various AHI cut-points. In practice, however, clinicians generally do not rely solely on AHI for OSA diagnosis and for determining treatment plans, but also base decisions on symptoms, sleep architecture, arousal indices, degree of desaturation, and examination of raw PSG data. Further investigation would be useful to examine how differences in AHI impact on clinical diagnosis and treatment by clinicians given a full clinical picture. Nevertheless, as previously mentioned, the AHI is widely used as a metric for aspects other than OSA diagnosis, such as for case finding, to classify OSA severity, and to determine eligibility for third party treatment funding, highlighting the importance of this analysis.

One other specific benefit of our data is that it may provide insight when comparing prevalence estimates in published studies of normal and clinical populations where different hypopnea definitions have been used. For example, when a historical comparison to a normal population is required, that most commonly referred to is the Wisconsin study of middle-aged adults from the general population.[23] The hypopnea scoring rule employed in the Wisconsin study—any discernable decrement in sum of thoracic and abdominal movement accompanied by ≥ 4% desaturation[9]—is most closely approximated by the current AASM recommended definition. Although thermal sensors[22,24-28] and, to a lesser extent, respiratory movement measurements[22,24] have been shown to detect a lower number of respiratory events than nasal pressure recordings, no studies have made this comparison using definitions requiring 4% oxygen desaturation; in our experience, more severe hypopneas such as these are less likely to be underestimated by thermal sensors and respiratory movement than less severe hypopneas. Given this, our data suggests that the AHI cut-off of 5/h used to define sleep disordered breathing in the Wisconsin study is approximately equivalent to an AHI of 15/h using the Chicago hypopnea definition and 10/h using the alternative AASM definition.

CONCLUSION

The current study has demonstrated that using different published standard hypopnea definitions leads to marked differences in AHI, with implications for disease identification, severity grading, comparability of research and clinical results, treatment decisions, treatment funding by third parties, OSA prevalence estimates, estimates of the public health impact of OSA, and establishment of links between OSA and comorbidities. This study provides insight to clinicians and researchers when comparing results obtained using different standard hypopnea definitions and suggests that consideration should be given to revising the current scoring recommendations to include a single standardized hypopnea definition.

ACKNOWLEDGMENTS

We thank Bronwyn Stevens for assistance with the detailed sub-analysis examining the scoring criteria leading to differences in AHI.

Institutions at which work was performed: Institute for Breathing and Sleep, Austin Health, Heidelberg, Victoria, Australia; Royal Adelaide Hospital, Adelaide, South Australia, Australia.

DISCLOSURE STATEMENT

This was not an industry supported study. Mr. Rochford has received research support from Compumedics. Dr. O'Donoghue has participated in research studies sponsored by Sanofi-Aventis. Dr. Pierce is on the advisory board of Compumedics and has received research support from Compumedics, Resmed, and Boehringer Ingelheim. Ms. Singh and Dr. Thornton have participated in research studies sponsored by Takeda and Sanofi-Aventis. M. Ruehland has indicated no financial conflicts of interest.

REFERENCES

1. Manser RL, Rochford P, Naughton MT, et al. Measurement variability in sleep disorders medicine: the Victorian experience. Intern Med J 2002;32(8):386-93.
2. Moser NJ, Phillips BA, Berry DT, Harbison L. What is hypopnea, anyway? Chest 1994;105(2):426-8.
3. Manser RL, Rochford P, Pierce RJ, Byrnes GB, Campbell DA. Impact of different criteria for defining hypopneas in the apnea-hypopnea index. Chest 2001;120(3):909-14.
4. Nigro CA, Rhodius EE. Effect of the definition of hypopnea on apnea/hypopnea index. Medicina (B Aires) 2003;63(2):119-24.
5. Redline S, Kapur VK, Sanders MH, et al. Effects of varying approaches for identifying respiratory disturbances on sleep apnea assessment. Am J Respir Crit Care Med 2000;161(2):369-74.
6. Tsai Willis H, Flemons WW, Whitelaw William A, Remmers John E. A comparison of apnea-hypopnea indices derived from different definitions of hypopnea. Am.J Respir Crit Care Med 1999;159(1):43-8.
7. Redline S, Sanders M. Hypopnea, a floating metric: implications for prevalence, morbidity estimates, and case finding. Sleep 1997;20(12):1209-17.
8. Stepnowsky CJ Jr, Berry C, Dimsdale JE. The effect of measurement unreliability on sleep and respiratory variables. Sleep 2004;27(5):990-5.
9. 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(5):667-89.
10. Quan SF, Gillin JC. New definitions of sleep disordered breathing--not yet a mandate in clinical practice. Sleep 1999;22(5):662.
11. Meoli AL, Casey KR, Clark RW, et al. Hypopnea in sleep-disordered breathing in adults. Sleep 2001;24(4):469-70.
12. Centers for Medicare and Medicaid Services. National Coverage Determination for Continuous Positive Airway Pressure (CPAP) Therapy for Obstructive Sleep Apnea (OSA). NCD #240.4. 2005 [cited 2008 22 September.]; Available from: http://www.cms.hhs.gov.

13. 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(4):499-521.
14. Iber C, Ancoli-Israel S, Chesson A, Quan S; for the American Academy of Sleep Medicine. 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.
15. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Los Angeles: Brain Information Service, Brain Research Institute, UCLA, 1968.
16. EEG arousals: scoring rules and examples: a preliminary report from the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association. Sleep 1992;15(2):173-84.
17. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8(2):135-60.
18. Punjabi NM, Newman AB, Young TB, Resnick HE, Sanders MH. Sleep-disordered breathing and cardiovascular disease: an outcome-based definition of hypopneas. Am J Respir Crit Care Med 2008;177(10):1150-5.
19. Stamatakis K, Sanders MH, Caffo B, et al. Fasting glycemia in sleep disordered breathing: lowering the threshold on oxyhemoglobin desaturation. Sleep 2008;31(7):1018-24.
20. Redline S, Budhiraja R, Kapur V, et al. The scoring of respiratory events in sleep: reliability and validity. J Clin Sleep Med 2007;3(2):169-200.
21. Whitney CW, Gottlieb DJ, Redline S, et al. Reliability of scoring respiratory disturbance indices and sleep staging. Sleep 1998;21(7):749-57.
22. Series F, Marc I. Nasal pressure recording in the diagnosis of sleep apnoea hypopnoea syndrome. Thorax 1999;54(6):506-10.
23. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 1993;328(17):1230-5.
24. Ballester E, Badia JR, Hernandez L, Farre R, Navajas D, Montserrat JM. Nasal prongs in the detection of sleep-related disordered breathing in the sleep apnoea/hypopnoea syndrome. Eur Respir J 1998;11(4):880-3.
25. Epstein MD, Chicoine SA, Hanumara RC. Detection of upper airway resistance syndrome using a nasal cannula/pressure transducer. Chest 2000;117(4):1073-7.
26. Hernandez L, Ballester E, Farre R, et al. Performance of nasal prongs in sleep studies: spectrum of flow-related events. Chest 2001;119(2):442-50.
27. Norman RG, Ahmed MM, Walsleben JA, Rapoport DM. Detection of respiratory events during NPSG: nasal cannula/pressure sensor versus thermistor. Sleep 1997;20(12):1175-84.
28. Teichtahl H, Cunnington D, Cherry G, Wang D. Scoring polysomnography respiratory events: the utility of nasal pressure and oronasal thermal sensor recordings. Sleep Med 2003;4(5):419-25.