

## SCIENTIFIC INVESTIGATIONS

# Effect of Varying Definitions of Hypopnea on the Diagnosis and Clinical Outcomes of Sleep-Disordered Breathing: A Systematic Review and Meta-Analysis

Meghna P. Mansukhani, MD<sup>1</sup>; Bhanu Prakash Kolla, MD<sup>1,2</sup>; Zhen Wang, PhD<sup>3</sup>; Timothy I. Morgenthaler, MD<sup>1,4</sup>

<sup>1</sup>Center for Sleep Medicine, Mayo Clinic, Rochester, Minnesota; <sup>2</sup>Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota; <sup>3</sup>Division of Health Care Policy and Research, Mayo Clinic, Rochester, Minnesota; <sup>4</sup>Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, Minnesota

**Study Objectives:** Various criteria have been used for scoring hypopneas, leading to difficulties when comparing results in clinical and research settings. We conducted a systematic review and meta-analysis to assess the effect of different hypopnea definitions on the diagnosis, severity, and clinical implications of sleep-disordered breathing (SDB).

**Methods:** Ovid MEDLINE, Embase, and Scopus databases were queried for English-language publications from inception through March 7, 2017. Studies that directly compared various hypopnea definitions were eligible. The hierarchical summary receiver operating characteristic model was used to jointly estimate diagnostic performance for comparisons between criteria.

**Results:** The initial search yielded 2,828 abstracts; 28 met inclusion criteria. After reviewing reference lists and expert review, five additional articles were identified. Most of the studies were cross-sectional or retrospective in nature. Eleven studies compared 2007 recommended criteria with 2012 criteria; 6 of these (evaluating 6,628 patients) were suitable for inclusion in the meta-analysis. Using the 2012 definition ( $\geq 3\%$  desaturation or arousal) as the reference standard, the 2007 definition ( $\geq 4\%$  desaturation) showed a sensitivity of 82.7% (95% confidence interval 0.72–0.90) and specificity of 93.2% (95% confidence interval 0.82–0.98). Although 2007 criteria were found to be associated with prevalent cardiovascular (CV) disease and increased risk of CV death, the 2012 criteria appeared to correspond better with intermediate CV risk markers based on two abstracts.

**Conclusions:** As expected, 2012 hypopnea scoring criteria resulted in a greater prevalence and severity of SDB. Data regarding the effect of varying hypopnea definitions on clinical outcomes, quality of life, health care costs, and mortality rates are limited.

**Commentary:** A commentary on this article appears in this issue on page 683.

**Keywords:** health care costs, hypopnea, prevalence, ROC curve, sleep apnea syndromes, sensitivity and specificity

**Citation:** Mansukhani MP, Kolla BP, Wang Z, Morgenthaler TI. Effect of varying definitions of hypopnea on the diagnosis and clinical outcomes of sleep-disordered breathing: a systematic review and meta-analysis. *J Clin Sleep Med*. 2019;15(5):687–696.

### BRIEF SUMMARY

**Current Knowledge/Study Rationale:** Although the American Academy of Sleep Medicine recommended using a new hypopnea definition in 2012, most sleep centers are still using the 2007 “recommended” (4% desaturation) criteria due to insurance reimbursement requirements. To facilitate understanding in clinical and research settings, it would be important to know how the differing definitions affect estimates of disease severity or prevalence.

**Study Impact:** In this systematic review, we examined 33 studies that directly compared different hypopnea criteria and conducted a meta-analysis comparing the diagnostic performance of the 2012 and 2007 “recommended” criteria. Clinical outcomes associated with the different hypopnea definitions are reviewed, but this area needs further study.

## INTRODUCTION

The apnea-hypopnea index (AHI), the number of apneas plus hypopneas divided by time, is ubiquitously used to both diagnose and gauge the severity of sleep-disordered breathing (SDB). There are many causes of imprecision when using this metric, including how airflow is measured, what time is used in the denominator (total sleep time versus recording time), and how apneas and hypopneas are defined. In some patients, hypopneas constitute a relatively large proportion of respiratory events recorded by polysomnography (PSG); thus, varying definitions of hypopnea can influence an individual patient’s diagnosis, treatment, and clinical outcomes of SDB. This article

focuses on how varying definitions of hypopnea may influence classification of the presence and severity of SDB. We also sought to know which definition of hypopneas associate with clinical outcomes of importance.

The scoring criteria for hypopneas during PSG are largely based on expert consensus, and consider the degree of airflow reduction, level of desaturation, and presence of arousals.<sup>1–3</sup> The most widely published definitions are summarized in **Table 1**. In the United States, insurance reimbursement often respects only the 2007 IA (or the 2012 IB) definition of hypopnea, which must have a 4% oxyhemoglobin desaturation (4%OD) and does not respect arousals in the definition. Therefore, despite the more recent 2012 recommendation to define hypopneas by a flow reduction

**Table 1**—Hypopnea scoring criteria.

Source	Year	Criteria	Abbreviation
AASM Task Force	1999	Either $\geq 50\%$ airflow reduction OR a lesser airflow reduction with associated $\geq 3\%$ desaturation or arousal	1999Hyp
AASM Scoring Manual	2007	Recommended or IA: $\geq 30\%$ airflow reduction and $\geq 4\%$ desaturation	4%OD
		Alternate or IB: $\geq 50\%$ airflow reduction and $\geq 3\%$ desaturation	2007IB
AASM Scoring Manual	2012	Recommended or IA: $\geq 30\%$ airflow reduction and $\geq 3\%$ desaturation or arousal	3%ODA
		Acceptable or IB: $\geq 30\%$ airflow reduction and $\geq 4\%$ desaturation	4%OD

AASM = American Academy of Sleep Medicine.

with at least a 3% desaturation or an arousal (3%ODA), in order to ensure patients receive treatment many sleep laboratories report AHIs calculated using both the 4%OD and the 3%ODA definitions, which can cause confusion for patients, payors, durable medical equipment providers, and sometimes clinicians.<sup>4</sup>

Although the recommendation for using the 3%ODA rule considered that such events are associated with physiologic perturbations and can lead to symptoms that respond to treatment, scoring using the 3%ODA rule will identify more hypopneas than using the 4%OD rule.<sup>5-9</sup> The aim of this systematic review and meta-analysis was to identify and describe studies that directly compared different criteria for scoring hypopneas, and to quantify the effect of using varying hypopnea criteria on the diagnosis, severity classification, and clinical implications of obstructive and central SDB in adults.

## METHODS

This systematic review was conducted and reported in accordance with the PRISMA guidelines.<sup>10</sup> This research was exempt from review by the Mayo Clinic Institutional Review Board.

### Search Strategy

We searched the Ovid MEDLINE, Embase, and Scopus databases from inception through March 7, 2017, for all English-language articles of any study design with the following search terms: polysomnography, sleep-wake syndromes, sleep apnea syndromes, sleep-disordered breathing, hypoventilation/hypopnea\* or hypopnoea, oxygen, desaturation, score or scoring/AHI/severity/rules, interrater variation/observer variation, definition or criteria, diagnostic accuracy, sensitivity and specificity, AUC ROC curve, repeatability/reproducibility of results, and quantitation/quantify. The full search strategy is shown in the supplemental material.

### Inclusion and Exclusion Criteria

Studies of humans (age 18 years or older) were eligible for inclusion. We included studies that performed PSG or home sleep apnea testing (HSAT) for the diagnosis of SDB, compared two or more American Academy of Sleep Medicine (AASM) definitions of hypopneas, and reported the AHI using the definitions in **Table 1**. Retrospective, cross-sectional, and prospective studies were included. Book chapters, review articles, and editorials were excluded. The abstracts were reviewed by two investigators (M.P.M. and B.K.). We reviewed the full text of articles that

met inclusion criteria, and searched their reference lists for additional articles. Disagreements were resolved by discussion with another investigator (T.I.M.). We extracted information from individual articles independently and in duplicate (by M.P.M. and B.K.) with reference to participants, intervention, comparisons, outcomes, and study design (PICOS). Discrepancies were resolved through discussion and consensus. The quality of the included studies was appraised based on study design, blinding of investigators, and completeness of data reported.

### Definitions

Originally designated for research purposes, the 1999 definition of hypopnea specified either a  $\geq 50\%$  airflow reduction without desaturation or arousal considered, or a lesser airflow reduction with an associated  $\geq 3\%$  desaturation or an arousal from sleep.<sup>3</sup> We will refer to it as the “1999Hyp” definition (**Table 1**). The 2012 “recommended” criteria per the 2012 AASM scoring manual is referred to in the text as the “3%ODA” criteria.<sup>1,11</sup> The 2007 “Recommended IA” criteria for hypopnea was identical to the 2012 “Acceptable” criteria, and will be referred to by the “4%OD” term to simplify comparisons. The “Alternate or IB” 2007 AASM scoring manual hypopnea criteria included a  $\geq 50\%$  airflow reduction with a  $\geq 3\%$  oxyhemoglobin desaturation (the “2007 IB” criteria).<sup>2</sup>

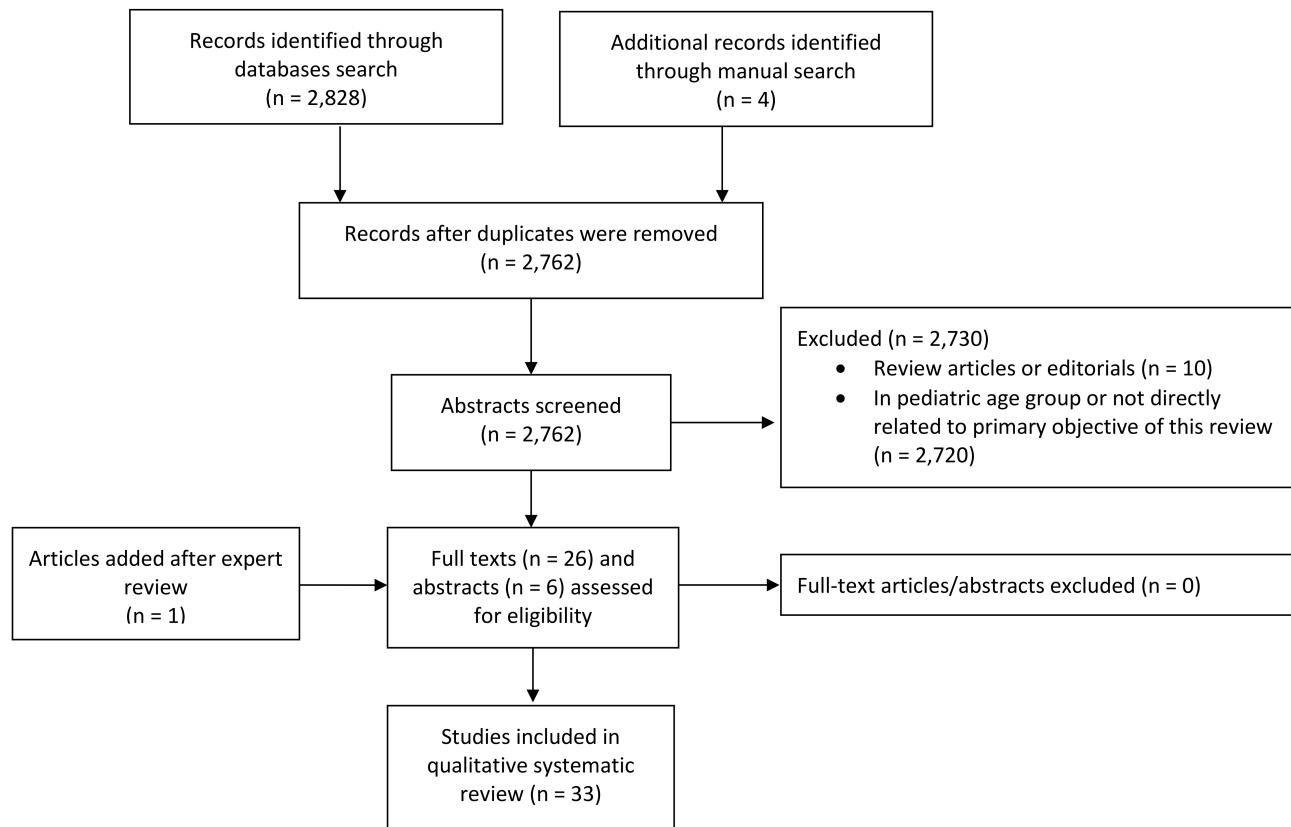
### Statistical Analyses

We compared different criteria for scoring hypopneas on diagnostic performance. We extracted true-positive, true-negative, false-positive, and false-negative outcomes. The symmetric hierarchical summary receiver operating characteristic (HSROC) model was used to jointly estimate sensitivity and specificity, positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio.<sup>12</sup> SDB was defined as AHI  $\geq 5$  events/h, used as a categorical variable. We were unable to evaluate publication bias as the number of studies included in the meta-analysis was relatively small ( $n = 6$ ). All statistical analyses were conducted using Stata software (version 15.1; StataCorp LLC College Station, Texas, United States).

## RESULTS

### Search Results

The initial search identified 2,828 articles. **Figure 1** shows how the final 33 articles were selected for the analysis. We could not obtain full-text articles or further information for five of six

**Figure 1**—PRISMA flow diagram of literature search.

published conference or journal supplement abstracts, despite attempts to contact the corresponding authors; however, sufficient data with reference to PICOS were available to allow their inclusion in the systematic review.

Most of the articles were retrospective ( $n = 19$ ) or cross-sectional studies ( $n = 9$ ). Three prospective cohort studies were included.<sup>13–15</sup> One study involved conduction of PSG in patients with a negative HSAT, and another study evaluated the effect of using a nasal pressure sensor only versus combination thermal and nasal pressure sensors for the diagnosis of OSA.<sup>16,17</sup> No randomized trials or case-control investigations were identified. The studies that were included in this review are summarized in **Table 2**, **Table 3**, **Table 4**, and **Table 5**.<sup>13–45</sup> Further details regarding the studies are shown in the supplemental material (**Table S1**, **Table S2**, **Table S3**, and **Table S4**).

An updated search through April 11, 2018 yielded 367 additional articles; following a review of the abstracts, one additional relevant article was found that is described in the next paragraphs.<sup>46</sup>

### Studies Comparing 3%ODA and 4%OD Criteria

A total of 11 studies compared the 3%ODA and 4%OD criteria (**Table 2**) across various outcomes.<sup>13,18–26,45</sup>

### Effect of Hypopnea Definition on Diagnosis, Severity, and Type of SDB

As expected, the AHI was higher using the 3%ODA definition than when using 4% OD or 2007 IB criteria. The studies were

conducted in various patient populations, including groups with acute transient ischemic attack or stroke and symptomatic heart failure (HF) with reduced ejection fraction (EF).<sup>18,22</sup> Across studies, a considerable proportion of patients designated as having no SDB using older criteria were considered to have SDB with the 3%ODA criteria, with the differences in AHI diminishing with increasing levels of SDB severity.<sup>13,18–26</sup> In the study by Ponsaing et al., most of the patients whose disease classification changed had an initial diagnosis of either no OSA or mild OSA.<sup>18</sup> In the supplement to the article by Ho et al., a gradual convergence in AHI derived using different criteria was observed at higher values (particularly  $> 60$  events/h) where the difference in AHI was less than 5 to 10 events/h, whereas for AHI 5 to 60 events/h, the difference was typically seen to exceed 10 events/h.<sup>21</sup>

The percentage of women with no SDB based on 4%OD criteria decreased when 3%ODA criteria were applied in one study.<sup>20</sup> Subsequent evaluation of the data from the cohort studied by Duce et al. suggested that the prevalence of a certain type of positional obstructive sleep apnea (OSA) that would be most amenable to positional therapy was increased when using the 3%ODA definition, and the proportion of women with this type of positional OSA increased as well.<sup>19,47</sup> Another recent study, published after the initial search for this systematic review was completed, suggested a decrease in rapid eye movement (REM)–predominant or REM-isolated OSA after applying the 3%ODA instead of the 4%OD criteria, because more non-rapid eye movement events were detected with the 3%ODA criteria.<sup>46</sup>

**Table 2**—Studies comparing 2012 and 2007 criteria.

Study	Participants or Records	Hypopnea Definitions and Main Results
Ponsaing et al.	n = 63	4%OD and 2007 IB, 3%ODA No difference in SDB prevalence with 2007 IB and 3%ODA, lower AHI with 4%OD. Effect of rescore with 3%ODA was highest in those with no/mild SDB
Duce et al.	n = 303	4%OD and 3%ODA 3%ODA increased OSA prevalence and proportion of those who could be successfully treated with positional therapy versus 4%OD
Campos-Rodriguez et al.	n = 1,116 women n = 939 elderly individuals	4%OD, 30%AH13%, and 3%ODA Prevalence of severe OSA increased with 3%ODA versus 4%OD and % of women with no SDB decreased. AHI $\geq$ 30 events/h using any definition in women and using 4%OD only in the elderly was associated with increased risk of CV mortality
Ho et al.	n = 6,441	4%OD, 30%AH13%, 3%ODA Median AHI highest with 3%ODA, lowest with 4%OD; relative difference in AHI was reduced in severe SDB
Heinrich et al.	n = 91	4%OD, 2007 IB, 3%ODA Higher AHI with 2007 IB and 3%ODA versus 4%OD; mean duration of CSR was longest with 2007 IB
Duce et al.	n = 112	1999Hyp, 4%OD, 2007 IB, 3%ODA Median AHI with %ODA was higher than 4%OD and 2007 IB, lower than 1999Hyp
Dean et al.	n = 2,040	Multiple, including 4%OD, 30%AH13%, 3%ODA 4%OD was most consistently associated with SBP and DBP
Leow et al.	n = 81	4%OD, 3%ODA Median AHI was higher with 3%ODA and correlated better with CAC scores
BaHamam et al.	n = 100	4%OD, 2007 IB, 3%ODA 4%OD detected fewer events compared with 2007 IB and 3%ODA
Heinzer et al.	n = 2,114	1999Hyp, 4%OD, 3%ODA Mean AHI highest with 1999Hyp, lowest with 4%OD; 3%ODA correlated best with SBP, DBP, and blood glucose
Punjabi et al.	n = 6,106	4%OD versus lesser desaturation thresholds or arousal, i.e., 4%OD versus 3%ODA, 2012 B and 2007 IB, and other comparisons Only 4%OD and higher levels of desaturation were independently associated with prevalent CV disease

30%AH13% =  $\geq$  30% reduction in airflow associated with  $\geq$  3% desaturation, AHI = apnea-hypopnea index, CAC = coronary artery calcification, CSR = Cheyne-Stokes respiration, CV = cardiovascular, DBP = diastolic blood pressure, OSA = obstructive sleep apnea, SBP = systolic blood pressure, SDB = sleep-disordered breathing.

**Table 3**—Studies comparing 4%OD and 2007 IB criteria.

Study	Participants or Records	Hypopnea Definitions and Main Results
Nerfeldt et al.	n = 187	4%OD, 2007 IB Most patients had OSA after rescore with 2007 IB
Ward et al.	n = 181	4%OD, 2007 IB Median AHI was higher with 2007 IB; proportion of obstructive and central SDB events stayed the same
Thornton et al.	n = 164	4%OD, 2007 IB Transducer resulted in higher AHI than thermistor with 4%OD versus 2007 IB; same results with transducer alone versus conventional AHI
Kuna et al.	n = 70	1999Hyp, 4%OD, 2007 IB Mean AHI highest with 1999Hyp, lowest with 4%OD; interscorer agreement best with 4%OD
Guilleminault et al.	n = 35	1999Hyp, 4%OD, 2007 IB Baseline AHI highest with 1999Hyp, lowest with 4%OD; self-reported sleepiness correlated best with 1999Hyp

AHI = apnea-hypopnea index, OSA = obstructive sleep apnea, RDI = respiratory disturbance index, SDB = sleep-disordered breathing.

Last, in a study of patients with HF and reduced EF, the duration of Cheyne-Stokes respiration (CSR) decreased with the 3%ODA criteria compared to 2007 IB and increased with the 2007 IB criteria compared to the 4%OD criteria.<sup>22</sup>

We identified 6 studies comparing 4%OD criteria with the 3%ODA criteria (6,628 patients) with AHI as an outcome suitable for pooling of data.<sup>13,18–20,22,23</sup> We conducted a meta-analysis with information obtained regarding sensitivity and specificity

**Table 4**—Studies comparing 1999Hyp criteria to 2012 or 2007 criteria, not listed previously.

Study	Participants or Records	Hypopnea Definitions and Main Results
Hobson et al.	n = 30	1999Hyp, 4%OD, 2007 IB Median AHI highest with 1999Hyp, lowest with 4%OD; rate of “surgical success” higher with 4%OD
Haba-Rubio et al.	n = 505	1999Hyp, 4%OD Prevalence of mild, moderate and severe OSA in men and women with and without sleepiness was higher with 1999Hyp versus 4%OD
Ruehland et al.	n = 328	1999Hyp, 4%OD, 2007 IB Median AHI highest with 1999Hyp, lowest with 4%OD; large AHI-dependent patient specific differences were noted, with several patients classified as negative for SDB with newer criteria

AHI = apnea-hypopnea index, OSA = obstructive sleep apnea, SDB = sleep-disordered breathing.

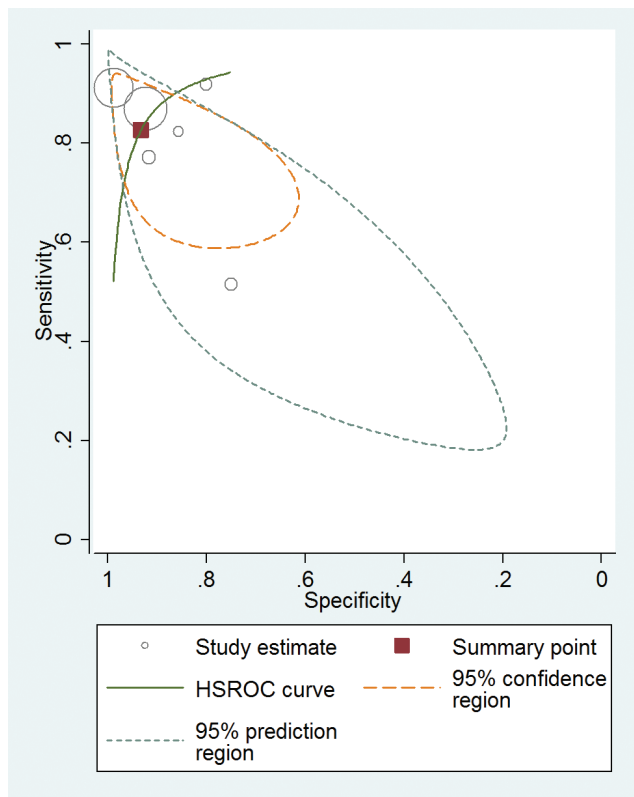
**Table 5**—Studies examining other hypopnea criteria, not listed previously.

Study	Participants or Records	Hypopnea Definitions and Main Results
Myllymaa et al.	n = 68	30%AH12-8% desaturation Higher median AHI with 3% versus 4% desaturation criterion
Aurora et al.	n = 200	3% and 4% desaturation thresholds Automatic scoring underestimated AHI versus manual scoring, differed depending on criteria used
Otero et al.	n = 40	Event duration 5–10 seconds, 2007 IB AHI higher with duration criterion
Ruiz-Lopez et al.	n = 189	30%AH13%, 4%OD Acceptable level of agreement between visual RDI and both criteria
Masa et al.	n = 90	Decrease in thoraco-abdominal motion on inductance plethysmography with $\geq 3\%$ desaturation and/or arousal, Pes amplitude reduction and $\geq 3\%$ desaturation and/or arousal Noninvasive method detected most Pes hypopneas. Inclusion of arousal to criteria markedly increased number of events, association with ESS noted
Ciftci et al.	n = 90	1999Hyp, 50-70% AHlarousal, 50-70% AH13% , 50%effort 1999Hyp had highest sensitivity and specificity for detecting “clinical OSA with sleepiness”
Nigro and Rhodius	n = 20	30% reduction in airflow or 50% in thoraco-abdominal movement with type 1 = 3% desaturation, type 2 = 3% desaturation or arousal, type 3 = 4% desaturation, type 4 = 4% desaturation or arousal AHI higher with types 2 and 4; no difference between types 1 and 2
Quan et al.	n = 91	30%AH13%, 4%OD No bias in RDI between the initial and repeat study with either criteria; significant inter-individual variability (age, sex, BMI not predictive)
Rochford et al.	n = 48	> 10 second plus 50% decrease in $\geq 1$ of 3 (nasal transducer, oral thermistor, thoraco-abdominal plethysmography), 1999Hyp, any reduction in airflow $\geq 10$ seconds plus 3% desaturation Mean AHI highest with 1999Hyp, lowest with last criterion above
Manser et al.	n = 48	Same definitions as Rochford et al. Level of agreement good
Cracowski et al.	n = 15	Hypopneas classified based on (1) cortical arousal and (2) autonomic arousal Autonomic arousal as accurate as cortical in classifying hypopneas
Redline et al.	n = 5,046	Relationships among 10 RDI definitions Median RDI varied by 10-fold
Hibbert et al.	n = 48	Apnea or hypopnea with 4% desaturation, 1999Hyp Mean RDI higher with 1999Hyp; effect of rescoring with 1999Hyp highest in those with mild SDB
Tsai et al.	n = 94	4%OD, 4% desaturation or arousal, arousal only Prevalence of OSA increased with addition of arousal or using arousal only criteria

30%AH12-8% =  $\geq 30\%$  reduction in airflow associated with 2% to 8% desaturation, 30%AH13% =  $\geq 30\%$  reduction in airflow associated with  $\geq 3\%$  desaturation, 50% to 70% AHlarousal = 50% to 70% reduction in airflow associated with arousal, 50% to 70% AH13% = 50% to 70% reduction in airflow associated with  $\geq 3\%$  desaturation, 50%effort =  $\geq 50\%$  decrease in thoraco-abdominal signal amplitude, AHI = apnea-hypopnea index, BMI = body mass index, ESS = Epworth Sleepiness Scale, OSA = obstructive sleep apnea, Pes = esophageal pressure, PSG = polysomnography, RDI = respiratory disturbance index, SDB = sleep-disordered breathing.



**Figure 2**—Meta-analysis of studies comparing 4%OD criteria for hypopnea against 2012 criteria.



Summary	Point Estimate	95% CI
Sensitivity	0.83	0.72–0.90
Specificity	0.93	0.82–0.98
Diagnostic odds ratio	65.74	13.03–331.68
Positive likelihood ratio	12.17	3.92–37.76
Negative likelihood ratio	0.19	0.104–0.33
Inverse of negative likelihood ratio	5.40	3.03–9.63

CI = confidence interval, HSROC = hierarchical summary receiver operating characteristic.

for SDB using  $\text{AHI} \geq 5$  events/h (and not of breathing events themselves) using these criteria. Choosing the more inclusive 3%ODA definition as the reference standard, the 4%OD definition showed a sensitivity of 82.7% (95% confidence interval [CI] 0.72–0.90) and specificity of 93.2% (95% CI 0.82–0.98) for the point estimate in the HSROC curve (Figure 2). Although there were no false-positive outcomes for any of the studies, statistical calculations in the HSROC model applied a small correction in order to calculate sensitivity, specificity, etc., which is a limitation of using this model.

### Effect on Clinical Outcomes

Data were sparse on clinical outcomes associated with the different hypopnea definitions. One study found no differences in sleepiness scores, psychomotor vigilance testing, or quality of life measures when using 4%OD or 3%ODA criteria.<sup>19</sup>

In women,  $\text{AHI} \geq 30$  events/h was independently associated with increased risk of cardiovascular (CV) death regardless

of the hypopnea definition used in one large study. However, in the elderly, this association was only noted in those with a diagnosis of  $\text{AHI} \geq 30$  events/h using the 4%OD and 3% desaturation criteria but not 3%ODA criteria.<sup>20</sup> The Sleep Heart Health Study (SHHS) showed that the 4%OD criteria ( $\geq 4\%$  desaturation) and not those using lesser desaturation thresholds or arousal were associated with prevalent CV disease.<sup>26</sup>

In the multiethnic study of atherosclerosis, Dean et al. found that the  $\text{AHI}$  derived from the 4%OD criterion was most consistently associated with systolic and diastolic blood pressures (BP).<sup>45</sup> In an abstract published by Leow et al., computed tomography-derived coronary artery calcification (CAC) scores appeared to correspond best with the 3%ODA hypopnea definition, but predictive accuracies of both definitions (4%OD and 3%ODA) were similar at higher CAC values.<sup>24</sup> Finally, another abstract noted that systolic and diastolic BP and blood glucose levels corresponded better with the 3%ODA criteria than with 4%OD criteria.<sup>25</sup>

### Studies Comparing 2007 IB and 3%ODA Criteria

Four of the previously described studies compared 2007 IB against 3%ODA criteria as the reference standard.<sup>13,18,22,23</sup> In most of these studies the 3%ODA and 2007 IB definitions (that use a 3% desaturation threshold) appeared to yield similar prevalence rates of SDB, and both differed from prevalence using the 4%OD definition.<sup>13,18,22</sup> In three studies in which diagnostic performance of the criteria could be assessed, the sensitivity against the 3%ODA criteria ranged from 93% to 100% and specificity was 100%.<sup>13,22,23</sup>

### Studies Comparing 4%OD and 2007 IB Criteria

Nine studies compared the 4%OD against 2007 IB criteria, of which four are included in the previous section.<sup>13,18,22,23</sup> The remainder are shown in Table 3.

### Effect on Diagnosis, Severity, and Type of SDB

The use of 2007 IB criteria increased the number of SDB events compared with 4%OD criteria. In one study, 90% of symptomatic patients with a negative HSAT using 4%OD criteria were found to have OSA on PSGs scored using 2007 IB criteria.<sup>16</sup>

A higher discrepancy was noted between oronasal thermistor and nasal pressure transducer signal-derived  $\text{AHI}$ , as well as between nasal pressure-only and conventional (ornasal thermistor plus nasal pressure transducer)  $\text{AHI}$  with 4%OD versus 2007 IB criteria.<sup>17</sup> In this study, the thermal sensor was judged to be of inadequate quality for scoring in more than 20% of epochs in 12 of 164 studies (7%), and the nasal pressure transducer was inadequate in 4 of 164 studies (2.5%).<sup>17</sup> However, in another study, greater disagreement was noted in the  $\text{AHI}$  between scorers when 2007 IB criteria were applied instead of 4%OD criteria.<sup>28</sup> The classification of SDB as OSA or CSA did not appear to be affected by the hypopnea definition used.<sup>27</sup>

The 4%OD criteria showed a sensitivity of 83% to 100% and specificity of 70% to 100% when compared with 2007 IB criteria, in the four studies where this information could be ascertained.<sup>22,23,27,28</sup> A meta-analysis of these studies could not be performed because of marked variability in the populations and outcomes assessed in these investigations.

### Effect on Clinical Outcomes

Only 4%OD criteria corresponded with prevalent CV disease as noted previously.<sup>26</sup> However, symptomatic improvement after treatment of OSA with continuous positive airway pressure devices, oral appliances, and surgery appeared to correspond better with 2007 IB and 1999Hyp criteria ( $r = .6$ ) than with 4%OD criteria ( $r = .4$ ) in one study of patients who were lean.<sup>29</sup>

### Studies Comparing 2012 or 2007 to 1999 Criteria

Four studies that examined the 3%ODA or 2007 criteria and compared it with the 1999Hyp criteria have been described previously.<sup>23,25,28,29</sup> The remaining three studies are shown in **Table 4**.

### Effect on Diagnosis and Severity of SDB

In the studies comparing 1999Hyp criteria to newer hypopnea criteria, the AHI based on the 1999Hyp definition was considerably higher than that based on 4%OD, 2007 IB, and 3%ODA criteria, in both sexes, and across all severity ranges.<sup>23,25,28,29,31,32</sup> As expected, the difference was most marked between the 1999Hyp and 4%OD criteria; the prevalence rate was more than double the former in some investigations.<sup>31,32</sup> One study estimated that a substantial proportion of patients with SDB diagnosed with the 1999Hyp definition would not be classified as having SDB with newer definitions (40% with 4%OD and 25% with 2007 IB).<sup>32</sup> In another study, good agreement was noted between scorers using 4%OD criteria compared with 1999Hyp and 2007 IB criteria.<sup>28</sup>

In the one study in which data regarding the diagnostic performance of 3%ODA versus 1999Hyp criteria could be computed, the sensitivity was 73% (95% CI 0.39–0.93) and specificity was 100% (95% CI 0.95–1.00).<sup>23</sup> When comparing 4%OD versus 1999Hyp criteria, the sensitivity ranged from 73% to 87% and specificity ranged from 85% to 100%,<sup>23,28</sup> whereas with 2007 IB versus 1999Hyp criteria, these values were 89% to 92% and 93% to 100%, respectively, in the studies where this information could be calculated.<sup>23,28,32</sup>

### Effect on Clinical Outcomes

The effect on posttreatment outcomes in patients who were lean and who had OSA has been described previously, and one other study showed that the success rate of surgery performed to treat OSA was higher with 4% OD versus 1999Hyp criteria.<sup>29,30</sup> In one study, as noted previously, BP and blood glucose levels were most closely related to the 3%ODA definition and not to the 1999Hyp or 4%OD definitions.<sup>25</sup>

### Studies Using Other Hypopnea Criteria

Fourteen studies compared various miscellaneous definitions of hypopnea (**Table 5**).

### Effect on Diagnosis and Severity of SDB

In the studies that used miscellaneous definitions for hypopnea, there was greater concordance in AHI obtained by using hypopnea criteria that incorporated a similar degree of desaturation compared with those based on amplitude reduction or arousal.<sup>42,44</sup> One abstract suggested that the effect of rescoring hypopneas (that were initially scored using a 4% desaturation

threshold) using 1999Hyp criteria was greater in patients classified as having “mild” SDB and had the effect of combining previously diagnosed mild and moderate disease.<sup>43</sup> There was limited evidence assessing the accuracy of obstructive-appearing hypopneas scored on PSG against those recorded via esophageal manometry.<sup>14,15</sup> The wide range of hypopnea definitions used precluded a meta-analysis of this subgroup of studies.

### Effect on Clinical Outcomes

The 1999Hyp criteria appeared to have good sensitivity and specificity for detecting patients with OSA and sleepiness compared to other criteria.<sup>37</sup> One study showed an association between arousal-based (and not desaturation-based) criteria with Epworth Sleepiness Scale scores whereas another study did not show a difference.<sup>14,37</sup> These studies did not report other clinical outcomes.

## DISCUSSION

In the practice and study of sleep medicine, the AHI is currently used as both a diagnostic criterion and as a means for stratifying severity of sleep apnea. In this review, we have examined the effect of changing recommendations for how to define a hypopnea. In our meta-analysis comparing the 2007 4%OD criteria to the latest 3%ODA criteria, using a diagnostic AHI threshold  $\geq 5$  events/h, the sensitivity was 83% and specificity was 93%. These findings suggest that these changes in criteria might lead to a 20% or greater difference in the number of diagnoses of OSA. Although the direction of this finding is not surprising, the magnitude of the difference is significant. If one assumes that approximately 5% to 15% of the adult population in the United States has at least mild OSA using the 4%OD rule (approximately 32.7 million adults), an increase of 20% using the 3%ODA rule may lead to an additional 6.5 million adults receiving a diagnosis of OSA that may benefit from treatment.<sup>48,49</sup> Conversely, working to diagnose these milder cases may divert attention and resources from treating more severe ones that may have greater association with important cardiovascular and other outcomes.<sup>50</sup> The change in classification between disease and health has significant implications for resource utilization as well as health outcomes.

The differing definitions also may lead to alterations in disease subtype classification, such as REM-related or positional SDB.<sup>46,47</sup> These may have both prognostic and therapeutic importance. There appeared to be no effect on the classification of SDB into OSA and CSA.<sup>27</sup> However, statistically significant differences in the scoring of CSR were noted on PSG.<sup>22</sup> Therefore, it is not clear whether the different hypopnea definitions affect the categorization of SDB into OSA or CSA and its downstream treatment implications.

Do the changes in hypopnea definition lead to better correlation between AHI and other symptoms of the syndrome such as sleepiness? Data regarding the effect of adding arousals to  $\geq 3\%$  desaturation hypopnea criteria on self-reported measures of sleepiness were conflicting.<sup>14,37</sup> In one study, the more liberal arousal-based criteria appeared to be better at gauging

improvement in sleepiness posttreatment in patients with OSA who were lean and in another study a dose-response relationship between increasing respiratory disturbance index (RDI) (based on arousal criteria) and Epworth Sleepiness Scale score was noted.<sup>29,14</sup> In other studies there was no difference in measures of sleepiness, vigilance, or quality of life with the different hypopnea definitions.<sup>19,37</sup> These findings indicate that the inclusion of arousals in the hypopnea definition might differentially associate with daytime alertness and the prevalence and course of various neurocognitive sequelae of SDB, perhaps more so in those with lower BMI, but this will need to be confirmed in future studies.

The effect of changing hypopnea definitions also depends on sex and age. In one large study, severe SDB defined using the 4%OD (but not the 3%ODA) criterion was associated with increased CV mortality risk in the elderly regardless of sex; in women, severe SDB, regardless of hypopnea definition, was associated with this outcome.<sup>20</sup> The percentage of women with a diagnosis of SDB was higher with 3%ODA versus 4%OD criteria in this study. In another study the proportion of women increased in the category of individuals with positional OSA who might be successfully treated with positional therapy techniques with 3%ODA versus 4%OD criteria.<sup>19</sup> Yet another study showed that age, sex, and BMI were not predictive factors for the difference in RDIs determined by 3% and 4% desaturation criteria (not including arousal).<sup>39</sup> These results suggest that there may be sex differences in the prevalence, type, and effect of SDB based on varying definitions of hypopneas, particularly with regard to criteria that include arousals in women compared to those that do not. Further studies are required to clearly determine if there are age, sex and/or BMI-related differences in the diagnosis and outcomes of SDB with different hypopnea criteria.

Among the studies reviewed, the SHHS and the study of Heinzer et al. were unique in that they were not performed in a referral population.<sup>20,25,26,31,39,42</sup> The SHHS demonstrated that the 4%OD criteria best associated with prevalent CV disease.<sup>26</sup> The multiethnic study of atherosclerosis also showed that the 4%OD criteria corresponded best with systolic and diastolic BP.<sup>45</sup> However, based on two published abstracts, the 3%ODA criteria appeared to correspond best with intermediate markers of CV risk such as CAC scores, systolic and diastolic BP, and blood glucose levels.<sup>24,25</sup> Thus, the question of which criteria are best associated with adverse CV outcomes remains unresolved. This remains an extremely important question, particularly in nonsleepy individuals, where the decision to treat is often based on the degree of severity of SDB as judged by the AHI.

Sleep medicine is not alone in adjusting important diagnostic criteria for chronic diseases. Over the past 25 years there have been significant changes in the diagnostic criteria for sepsis, diabetes mellitus, hypertension, obesity, and hypercholesterolemia, to name only a few.<sup>51</sup> Making changes to diagnostic criteria in medicine requires an investment of significant resources—an investment that is hoped to have a good return. For those practicing sleep medicine, the changes in hypopnea definition over the years has meant acquiring new sensors, training technologists and physicians to score against different criteria,

changing polysomnographic computer algorithms, design and implementation of new reports, and managing the downstream effects of those reports such as educating community physicians, durable medical equipment providers, and insurance regulators. The costs of these investments have not been well characterized. What is the return on these investments? The findings of this study are clear regarding how hypopnea definitions lead to significant changes in disease classification, but the findings are not clear regarding whether these changes improve the ability to prognosticate, and there were relatively few data regarding the effect on correlation with other important complications of OSA.

## Limitations

This meta-analysis and systematic review should be viewed in light of some limitations. Methodologically, few prospective studies have been published that explicitly examine the effect of varying definitions on the prevalence and consequences of sleep apnea. Most of the studies were cross-sectional or retrospective, were conducted in differing populations and settings (community-based to tertiary academic institutions), and most investigations assessed patients suspected of having OSA. Information that could be abstracted and included in a meta-analysis was available only for studies comparing the 4%OD and 3%ODA criteria. It seems unlikely that the inherently increased risk of bias or confounding encountered with limited populations and retrospective study designs would alter the conclusion that the 3%ODA criteria would be more sensitive in detecting breathing abnormalities, but the observed 20% difference in sensitivity might be lower in more generalized populations, and the important issue of clinical relevance probably requires more and larger prospective studies to reduce bias and improve the accuracy of the prevalence implications of differing hypopnea definitions.

## Future Directions

The AHI is an important metric that conveys information regarding the severity of sleep apnea, influences treatment decisions, and has prognostic implications. Currently there is no consensus regarding the appropriate definition of a hypopnea. Future studies are needed to examine the effect of various definitions used to calculate AHI in order to determine which definition correlates best with long-term outcomes (eg, cardiovascular, neurocognitive, psychiatric, metabolic) in different age, sex, and BMI subgroups. Studies comparing various definitions should also examine the correlation between AHI calculated using these definitions and symptoms of SDB such as sleepiness. In addition, studies examining the effect of hypopnea definitions on the classification of SDB into various types (obstructive versus central, positional or REM-related), and on the methods of scoring (automated versus manual, sensors utilized) and interscorer reliability are required. Finally, there is a need to examine the cost-effectiveness of the differential definitions of hypopneas, perhaps in a population estimate of quality adjusted life-years assuming certain treatment effects. The answers to these questions will help determine what would constitute an ideal definition of a hypopnea.



## ABBREVIATIONS

AASM, American Academy of Sleep Medicine  
 AHI, apnea-hypopnea index  
 BP, blood pressure  
 CAC, coronary artery calcification  
 CSR, Cheyne-Stokes respiration  
 CV, cardiovascular  
 EF, ejection fraction  
 HF, heart failure  
 HSAT, home sleep apnea testing  
 OSA, obstructive sleep apnea  
 PICOS, participants, intervention, comparisons, outcomes, and study design  
 PSG, polysomnography  
 REM, rapid eye movement  
 SDB, sleep-disordered breathing  
 SHHS, Sleep Heart Health Study

## REFERENCES

- 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 2.3. Darien, IL: American Academy of Sleep Medicine; 2016.
- Iber C, Ancoli-Israel S, Chesson AL Jr, Quan SF; 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.
- 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):667–689.
- Thomas RJ, Guilleminault C, Ayappa I, Rapoport DM. Scoring respiratory events in sleep medicine: who is the driver--biology or medical insurance? *J Clin Sleep Med*. 2014;10(11):1245–1247.
- American Academy of Sleep M. *International Classification of Sleep Disorders*. Darien, IL: American Academy of Sleep Medicine; 2014.
- Krakow B, Krakow J, Ulibarri VA, McIver ND. Frequency and accuracy of "RERA" and "RDI" terms in the Journal of Clinical Sleep Medicine from 2006 through 2012. *J Clin Sleep Med*. 2014;10(2):121–124.
- Collop N. Breathing related arousals: call them what you want, but please count them. *J Clin Sleep Med*. 2014;10(2):125–126.
- Meoli AL, Casey KR, Clark RW, et al. Hypopnea in sleep-disordered breathing in adults. *Sleep*. 2001;24(4):469–470.
- Sankari A, Pranathigeswaran S, Maresh S, Hosni AM, Badr MS. Characteristics and consequences of non-apneic respiratory events during sleep. *Sleep*. 2017;40(1).
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
- Berry RB, Brooks R, Gamaldo C, et al. AASM Scoring Manual updates for 2017 (Version 2.4). *J Clin Sleep Med*. 2017;13(5):665–666.
- Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20(19):2865–2884.
- BaHammam AS, Obeidat A, Barataman K, Bahammam SA, Olaish AH, Sharif MM. A comparison between the AASM 2012 and 2007 definitions for detecting hypopnea. *Sleep Breath*. 2014;18(4):767–773.
- Masa JF, Corral J, Teran J, et al. Apnoeic and obstructive nonapnoeic sleep respiratory events. *Eur Respir J*. 2009;34(1):156–161.
- Cracowski C, Pepin JL, Wuyam B, Levy P. Characterization of obstructive nonapneic respiratory events in moderate sleep apnea syndrome. *Am J Respir Crit Care Med*. 2001;164(6):944–948.
- Nerfeldt P, Aoki F, Friberg D. Polygraphy vs. polysomnography: missing osas in symptomatic snorers--a reminder for clinicians. *Sleep Breath*. 2014;18(2):297–303.
- Thornton AT, Singh P, Ruehland WR, Rochford PD. AASM criteria for scoring respiratory events: interaction between apnea sensor and hypopnea definition. *Sleep*. 2012;35(3):425–432.
- Ponsaing LB, Iversen HK, Jennum P. Sleep apnea diagnosis varies with the hypopnea criteria applied. *Sleep Breath*. 2016;20(1):219–226.
- Duce B, Kulkas A, Langton C, Toyras J, Hukins C. The AASM 2012 recommended hypopnea criteria increase the incidence of obstructive sleep apnea but not the proportion of positional obstructive sleep apnea. *Sleep Med*. 2016;26:23–29.
- Campos-Rodriguez F, Martinez-Garcia MA, Reyes-Nunez N, Selma-Ferrer MJ, Punjabi NM, Farre R. Impact of different hypopnea definitions on obstructive sleep apnea severity and cardiovascular mortality risk in women and elderly individuals. *Sleep Med*. 2016;27–28:54–58.
- Ho V, Crainiceanu CM, Punjabi NM, Redline S, Gottlieb DJ. Calibration model for apnea-hypopnea indices: impact of alternative criteria for hypopneas. *Sleep*. 2015;38(12):1887–1892.
- Heinrich J, Spiesshofer J, Bitter T, Horstkotte D, Oldenburg O. Implications of revised AASM rules on scoring apneic and hypopneic respiratory events in patients with heart failure with nocturnal Cheyne-Stokes respiration. *Sleep Breath*. 2015;19(2):489–494.
- 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.
- Leow LC, Baskaran L, Tan SY, Ong TH. Impact of different hypopnea definitions on predictive accuracy of apnea hypopnea index for coronary artery calcification. *Eur Respir J*. 2014;44(Suppl 58):P2010.
- Heinzer R, Haba-Rubio J, Vat S, Andries D, Tobback N, Tafti M. Which numbers do you want? Scoring and OSA prevalence [abstract]. *Sleep Biol Rhythms*. 2013;11(Suppl 2):1.
- 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):1150–1155.
- 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(9):1341–1348.
- Kuna ST, Benca R, Kushida CA, et al. Agreement in computer-assisted manual scoring of polysomnograms across sleep centers. *Sleep*. 2013;36(4):583–589.
- 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(4):341–347.
- Hobson JC, Robinson S, Antic NA, et al. What is "success" following surgery for obstructive sleep apnea? The effect of different polysomnographic scoring systems. *Laryngoscope*. 2012;122(8):1878–1881.
- Haba-Rubio J, Andries D, Bastardot F, et al. Prevalence of sleep disordered breathing in middle-aged general population: The hypnolaus study. *Sleep*. 2011;34(Abstr Suppl):A123.
- 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.
- Myllymaa S, Myllymaa K, Kupari S, et al. Effect of different oxygen desaturation threshold levels on hypopnea scoring and classification of severity of sleep apnea. *Sleep Breath*. 2015;19(3):947–954.
- Aurora RN, Swartz R, Punjabi NM. Misclassification of OSA severity with automated scoring of home sleep recordings. *Chest*. 2015;147(3):719–727.
- Otero A, Felix P, Presedo J, Zamarron C. Evaluation of an alternative definition for the apnea-hypopnea index. *Conf Proc IEEE Eng Med Biol Soc*. 2010;2010:4654–4657.
- Ruiz-Lopez FJ, Fernandez-Suarez B, Guardiola-Martinez J, Vergara-LaHuerta I, Latour-Perez J, Lorenzo-Cruz M. Quality control of the ambulatory polygraphy using automatic analysis. *Chest*. 2009;135(1):194–200.
- Ciftci TU, Kokturk O, Ozkan S. Apnea-hypopnea indexes calculated using different hypopnea definitions and their relation to major symptoms. *Sleep Breath*. 2004;8(3):141–146.

38. Nigro CA, Rhodius EE. Effect of the definition of hypopnea on apnea/hypopnea index. *Medicina (B Aires)*. 2003;63(2):119–124.
39. Quan SF, Griswold ME, Iber C, et al. Short-term variability of respiration and sleep during unattended nonlaboratory polysomnography--the Sleep Heart Health Study. [corrected]. *Sleep*. 2002;25(8):843–849.
40. Rochford P, Manser R, Pierce R, Campbell D. Impact of different criteria for defining hypopnea on the apnoea-hypopnoea index. *Respirology*. 2001;6(Suppl 1):A56.
41. 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):909–914.
42. 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 Pt 1):369–374.
43. Hibbert M, Joffe D, Phillips C, Berend N. Impact of Chicago draft criteria on assessment of respiratory disturbance index and severity grading. *Respirology*. 1999;4(Suppl 1):A13.
44. Tsai WH, Flemons WW, Whitelaw WA, Remmers JE. A comparison of apnea-hypopnea indices derived from different definitions of hypopnea. *Am J Respir Crit Care Med*. 1999;159(1):43–48.
45. Dean DA, Wang R, Jacobs DR, et al. A systematic assessment of the association of polysomnographic indices with blood pressure: the Multi-Ethnic Study of Atherosclerosis (MESA). *Sleep*. 2015;38(4):587–596.
46. Duce B, Kulkas A, Langton C, Toyras J, Hukins C. The prevalence of REM-related obstructive sleep apnoea is reduced by the AASM 2012 hypopnoea criteria. *Sleep Breath*. 2018;22(1):57–64.
47. Duce B, Kulkas A, Langton C, Toyras J, Hukins C. Amsterdam positional OSA classification: the AASM 2012 recommended hypopnoea criteria increases the number of positional therapy candidates. *Sleep Breath*. 2017;21(2):411–417.
48. Punjabi NM. The epidemiology of adult obstructive sleep apnea. *Proc Am Thorac Soc*. 2008;5(2):136–143.
49. Garvey JF, Pengo MF, Drakatos P, Kent BD. Epidemiological aspects of obstructive sleep apnea. *J Thorac Dis*. 2015;7(5):920–929.
50. Young T, Finn L, Peppard PE, et al. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. *Sleep*. 2008;31(8):1071–1078.
51. Schwartz LM, Woloshin S. Changing disease definitions: implications for disease prevalence. Analysis of the Third National Health and Nutrition Examination Survey, 1988-1994. *Eff Clin Pract*. 1999;2(2):76–85.

## ACKNOWLEDGMENTS

The authors thank Patricia J. Erwin, MLS, Mayo Clinic Libraries, Mayo Clinic, Rochester, Minnesota, for her assistance with the literature search. Ms. Erwin has provided written consent for this acknowledgment.

## SUBMISSION & CORRESPONDENCE INFORMATION

**Submitted for publication July 3, 2018**

**Submitted in final revised form November 14, 2018**

**Accepted for publication January 10, 2019**

Address correspondence to: Meghna P. Mansukhani, MD, Center for Sleep Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905; Email: mansukhani.meghna@mayo.edu

## DISCLOSURE STATEMENT

Work for this study was performed at Mayo Clinic, Rochester, Minnesota. All authors have seen and approved the manuscript. Results of this study were previously presented at SLEEP 2018, June 2018, Baltimore, Maryland. Dr. Mansukhani is the principal investigator for a study funded by ResMed Corporation to evaluate the effects of adaptive servoventilation therapy of central sleep apnea syndromes on health care utilization and has received a benefactor-sponsored career development award at Mayo Clinic (Rochester, Minnesota); these projects are unrelated to the current study. The authors report no conflicts of interest.