Misclassification of OSA Severity With Automated Scoring of Home Sleep Recordings

R. Nisha Aurora, MD; Rachel Swartz, BA; and Naresh M. Punjabi, MD, PhD, FCCP

BACKGROUND: The advent of home sleep testing has allowed for the development of an ambulatory care model for OSA that most health-care providers can easily deploy. Although automated algorithms that accompany home sleep monitors can identify and classify disordered breathing events, it is unclear whether manual scoring followed by expert review of home sleep recordings is of any value. Thus, this study examined the agreement between automated and manual scoring of home sleep recordings.

METHODS: Two type 3 monitors (ApneaLink Plus [ResMed] and Embletta [Embla Systems]) were examined in distinct study samples. Data from manual and automated scoring were available for 200 subjects. Two thresholds for oxygen desaturation ($\geq 3\%$ and $\geq 4\%$) were used to define disordered breathing events. Agreement between manual and automated scoring was examined using Pearson correlation coefficients and Bland-Altman analyses.

RESULTS: Automated scoring consistently underscored disordered breathing events compared with manual scoring for both sleep monitors irrespective of whether a $\geq 3\%$ or $\geq 4\%$ oxygen desaturation threshold was used to define the apnea-hypopnea index (AHI). For the ApneaLink Plus monitor, Bland-Altman analyses revealed an average AHI difference between manual and automated scoring of 6.1 (95% CI, 4.9-7.3) and 4.6 (95% CI, 3.5-5.6) events/h for the $\geq 3\%$ and $\geq 4\%$ oxygen desaturation thresholds, respectively. Similarly for the Embletta monitor, the average difference between manual and automated scoring was 5.3 (95% CI, 3.2-7.3) and 8.4 (95% CI, 7.2-9.6) events/h, respectively.

CONCLUSIONS: Although agreement between automated and manual scoring of home sleep recordings varies based on the device used, modest agreement was observed between the two approaches. However, manual review of home sleep test recordings can decrease the misclassification of OSA severity, particularly for those with mild disease.

TRIAL REGISTRY: ClinicalTrials.gov; No.: NCT01503164; www.clinicaltrials.gov

CHEST 2015; 147(3):719-727

Manuscript received April 17, 2014; revision accepted October 1, 2014; originally published Online First November 20, 2014.

ABBREVIATIONS: AHI = apnea-hypopnea index; ODI = oxygen desaturation index

AFFILIATIONS: From the Department of Medicine (Drs Aurora and Punjabi and Ms Swartz) and Department of Epidemiology (Dr Punjabi), The Johns Hopkins University School of Medicine, Baltimore, MD.

FUNDING/SUPPORT: This study was supported by grants from the National Institutes of Health [R01-HL075078, R01-HL117167, and K23-HL118414].

CORRESPONDENCE TO: Naresh M. Punjabi, MD, PhD, FCCP, Division of Pulmonary and Critical Care Medicine, The Johns Hopkins University School of Medicine, 5501 Hopkins Bayview Cir, Baltimore, MD 21224; e-mail: npunjabi@jh.u.edu

© 2015 AMERICAN COLLEGE OF CHEST PHYSICIANS. Reproduction of this article is prohibited without written permission from the American College of Chest Physicians. See online for more details.

DOI: 10.1378/chest.14-0929

OSA is a common sleep disorder that affects 5% to 15% of the general population.¹⁻³ Despite the increasing awareness of OSA and the subsequent proliferation of clinical sleep laboratories, OSA remains significantly underdiagnosed.^{4,5} The low level of case identification is partly due to the patient inconvenience, costs, and inaccessibility often associated with in-laboratory testing. The public health consequences of underdiagnosed OSA are significant, particularly if the condition is left untreated. OSA is associated with a multitude of adverse health outcomes, including incident hypertension, abnormalities in glucose metabolism, cardiovascular disease, and stroke.⁶ Accordingly, the impetus for health-care delivery systems and payers to create cost-effective and convenient ambulatory models of OSA diagnosis and treatment has led to the development of home sleep testing.

Tremendous advancements have been made in home sleep testing over the past several years. Portable monitoring devices have become easier and more convenient to use for both patients and health-care providers. Equipment that is simple to apply and less cumbersome to use has made home testing more comfortable for the patient. Correspondingly for the health-care provider, automated scoring algorithms embedded within the software accompanying the device have allowed for

expeditious diagnosis of OSA by practitioners not necessarily trained in sleep medicine. In 2008, the Centers for Medicare & Medicaid Services approved the widespread use of home sleep testing, stating that a home “sleep test must have been previously ordered by the beneficiary’s treating physician.”⁷ However, supervision by a sleep physician in the ordering or interpretation of acquired data was not specified. In contrast, the clinical guidelines on the use of unattended portable monitors for the diagnosis of OSA recommend that “evaluation of portable monitoring data must include review of the raw data by a board certified sleep specialist or an individual who fulfills the eligibility criteria for the sleep medicine certification examination.”⁸ Although the recommended guidelines are reasonable and appropriate, limited evidence supports or refutes the value of manual scoring of home sleep recordings and subsequent expert review by a sleep specialist. Specifically, it remains to be determined whether the severity of OSA as determined by automated scoring will be sufficiently different from manual scoring and expert review, thus justifying the need for such expertise. The objective of the current study was to assess the level of agreement in OSA severity between automated scoring without expert review and manual scoring using two distinct type 3 devices for the diagnosis of OSA.

Materials and Methods

Study Sample

To accomplish the objectives of the current study, two type 3 devices (ApneaLink Plus [ResMed] and Embletta [Embla Systems]) were used for home sleep testing. A distinct study sample was recruited for each device. The first sample consisted of a community-based cohort of subjects ($n = 100$) between the ages of 21 and 80 years who did not have an established diagnosis of OSA at the time of enrollment and were not receiving OSA therapy. The second sample consisted of patients recruited from a cardiology clinic ($n = 100$) with either known cardiovascular disease or at least two of the following risk factors for cardiovascular disease: hypertension, high cholesterol, or elevated BMI (≥ 30 kg/m²). Exclusion criteria for the second sample were known diagnosis of or any treatment for OSA. The study protocol was approved by The Johns Hopkins University institutional review board on human research (IRB approval number NA_00036672).

Home Sleep Testing

The first sample underwent home sleep testing with the ApneaLink Plus device. Nasal airflow was recorded with a nasal cannula connected to a pressure transducer. Pulse oximetry was used to assess oxyhemoglobin saturation, and respiratory effort was measured with a pneumatic sensor attached to an effort belt. The second sample had home sleep testing using the Embletta device. Signals recorded included nasal pressure through a nasal cannula attached to a pressure transducer, oronasal thermistor, pulse oximetry for oxyhemoglobin saturation, respiratory effort through piezoelectric belts, and body position. For both devices, the overnight sleep recordings were subjected to automated scoring without expert review, and at least 4 h of interpretable recording time were required for inclusion in the study. Each sleep recording from both

devices was also manually scored and reviewed by a physician (R. N. A.) board certified in sleep medicine who was blinded to the results of the automated scoring. Oxyhemoglobin desaturations of at least 3% were marked separately on the oxygen saturation channel, and an oxygen desaturation index (ODI) of $\geq 3\%$ and $\geq 4\%$ was calculated as the number of oxygen desaturation events per hour of recording time. Apneas and hypopneas were defined using the following criteria: Apneas were identified if there was a $\geq 90\%$ reduction in airflow for at least 10 s, and hypopneas were identified if there was a $\geq 30\%$ reduction in airflow for at least 10 s associated with either an oxyhemoglobin desaturation of $\geq 3\%$ (American Academy of Sleep Medicine definition) or $\geq 4\%$ (Medicare definition). For manual scoring, if a loss of airflow signal was observed in the primary channel used to detect apneas and hypopneas (ie, nasal pressure), then alternate channels (eg, effort channels) were used for scoring apneas and hypopneas. The scoring software for both devices was configured using these criteria. Similarly, manual scoring was conducted using these definitions. The apnea-hypopnea index (AHI), the disease-defining metric for OSA, was calculated as the number of apneas and hypopneas per hour of recording time. Two distinct AHI values were determined corresponding to the $\geq 3\%$ (AHI_{3%}) and the $\geq 4\%$ (AHI_{4%}) thresholds in oxyhemoglobin desaturation. The following AHI thresholds were used to define OSA severity: < 5.0 events/h (normal), 5.0 to 14.9 events/h (mild), 15.0 to 29.9 events/h (moderate), and ≥ 30 events/h (severe). Finally, the ODI was calculated as the number of oxygen desaturations per hour of recording based on the $\geq 3\%$ (ODI_{3%}) and $\geq 4\%$ (ODI_{4%}) thresholds.

Statistical Analysis

Complete data from manual (expert reviewed) and automated scoring was available for the 200 subjects. Using the $\geq 3\%$ and $\geq 4\%$ thresholds to define disordered breathing events, Pearson correlation coefficients

were derived between the manual and automated scoring for AHI and ODI. Agreement between manual and automated scoring was also examined using the method proposed by Bland and Altman.⁹ The average difference, or bias, in AHI and ODI was determined between manual and automated scoring results for the two devices. SAS 9.1 statistical software (SAS Institute Inc) was used for the analyses.

Results

The demographic and anthropometric variables along with the prevalence of various medical conditions assessed in the two study samples are shown in Table 1. Not surprisingly, the community-based sample was younger and less obese than the cardiology clinic sample. No differences were noted in the distribution of sex, race, or self-reported sleepiness (ie, Epworth Sleepiness Scale) between the two samples. Medical conditions such as hypertension, high cholesterol, myocardial infarction, and type 2 diabetes were more prevalent in the cardiology clinic sample. The overall failure rates for the ApneaLink Plus and the Embletta devices were 2.2% and 8.0%, respectively, with poor oximetry signal being the most common reason for failure.

Pearson correlation coefficients between automated and manually scored AHI values varied as a function of monitor type. With the ApneaLink Plus monitor, the correlation coefficients between manual and automated scoring were 0.968 (95% CI, 0.952-0.978) and 0.977 (95% CI, 0.965-0.984) for AHI_{3%} and AHI_{4%}, respectively (Fig 1). In contrast, with the Embletta monitor, the correlation coefficients between manual scoring and automated scoring were 0.639 (95% CI, 0.504-0.741) and 0.758 (95% CI, 0.658-0.829) for AHI_{3%} and AHI_{4%},

respectively (Fig 1). Figure 2 shows the Bland-Altman plots comparing the automated and manually scored AHI_{3%} and AHI_{4%} for both devices. For the ApneaLink Plus monitor, the average bias (manual – automated scoring) was 6.1 (95% CI, 4.9-7.3) and 4.6 (95% CI, 3.5-5.6) events/h for AHI_{3%} and AHI_{4%}, respectively. Thus, automated scoring underestimated the AHI compared with manual scoring. For the Embletta monitor (Fig 2), the average bias was 5.3 (95% CI, 3.2-7.3) and 8.4 (95% CI, 7.2-9.6) events/h for AHI_{3%} and AHI_{4%}, respectively, indicating an underestimation of the AHI between the two scoring approaches.

Differences between automated and manual scoring were less for ODI than for AHI for both devices. Correlation coefficients between automated and manually scored ODI values for the ApneaLink Plus monitor were 0.994 (95% CI, 0.991-0.996) and 0.996 (95% CI, 0.993-0.997) for ODI_{3%} and ODI_{4%}, respectively (Fig 3). Similarly for the Embletta monitor, correlation coefficients of 0.992 (95% CI, 0.987-0.994) and 0.997 (95% CI, 0.995-0.998) were noted for ODI_{3%} and ODI_{4%}, respectively (Fig 3). Bland-Altman plots comparing automated and manual scoring for ODI_{3%} and ODI_{4%} for both devices are shown in Figure 4. For the ApneaLink Plus monitor, the average bias (manual – automated

TABLE 1] Characteristics of the Study Samples

Characteristic	ApneaLink Plus Sample	Embletta Sample
Age, y	55.8 (21-79)	60.8 (45-74)
Male sex	65.0	67.1
Race		
White	65.7	71.3
Black	30.3	19.5
Other	4.0	9.2
BMI, kg/m ²	31.8 (21.0-40.3)	35.0 (20.7-64.5)
Waist, cm	109.3 (80.5-140.2)	117.9 (79.0-157.5)
Neck, cm	41.7 (33.6-50.2)	42.7 (32.8-52.3)
Epworth Sleepiness Scale score	10.0 (2.0-21.0)	9.5 (1.0-15.0)
Prevalent medical conditions		
Hypertension	61.3	86.2
High cholesterol	53.0	92.0
Myocardial infarction	3.8	22.0
Type 2 diabetes	30.9	50.6

Data are presented as mean (range) or %.

Figure 1 – A-D, Scatter plots of automated vs manual scoring of the AHI for the ApneaLink Plus (A, B) and Embletta (C, D) monitors using $a \geq 3\%$ (A, C) and $\geq 4\%$ (B, D) oxygen desaturation threshold. Pearson correlation coefficients and the line of identity (dashed line) are shown for each plot. AHI = apnea-hypopnea index.

scoring) was -0.1 (95% CI, -0.66 to 0.50) and 0.3 (95% CI, -0.2 to $0.0.8$) events/h for $ODI_{3\%}$ and $ODI_{4\%}$, respectively. With the Embletta monitor, the average bias was 2.2 (95% CI, 1.9 - 2.5) and 0.8 (95% CI, 0.6 - 0.9) events/h for $ODI_{3\%}$ and $ODI_{4\%}$, respectively. Overall, a high level of agreement was seen between automated and manual scoring for both devices when ODI was used as the disease-defining metric.

OSA severity as determined by ApneaLink Plus- and Embletta-derived AHI values is shown in Figures 5 and 6, respectively. With the ApneaLink Plus monitor, overall misclassification was 22% and 18% for $AHI_{3\%}$ and $AHI_{4\%}$, respectively. However, clinically significant misclassification was $< 10\%$ for the sample irrespective of the AHI definition. For the Embletta monitor, overall misclassification was significantly higher at 55% and 57% for $AHI_{3\%}$ and $AHI_{4\%}$, respectively, whereas clinically relevant misclassification with the Embletta monitor was 29% and 41%. As expected, disease misclassification was most notable

for subjects with mild disease irrespective of the device used.

Discussion

The results of the current study demonstrate that agreement between automated and expert-reviewed manual scoring for type 3 home portable monitoring devices varies with the type of device and the specific metric (ie, AHI, ODI) used for the diagnosis of OSA. For the two devices examined, automated scoring consistently underestimated the AHI compared with manual scoring. This underestimation resulted in misclassification of OSA severity, specifically in those with mild to moderate disease. Although some differences were noted for the ODI, automated and manually scored ODI values were highly correlated.

The past two decades have seen a proliferation of portable sleep monitoring devices, which have typically included automated scoring platforms designed to ease the diagnostic burden of OSA. The current investigation

Figure 2 – A-D, Bland-Altman plots for automated vs manual scoring of the AHI for the ApneaLink Plus (A, B) and Embletta (C, D) monitors using a $\geq 3\%$ (A, C) and $\geq 4\%$ (B, D) oxygen desaturation threshold. Average bias (solid line) and limits of agreement (dashed line) are shown for each device and oxygen desaturation threshold. See Figure 1 legend for expansion of abbreviation.

confirms and extends prior research on the topic of automated vs manual scoring for two portable sleep monitoring devices used for home sleep testing. Available evidence to date suggests that the accuracy of type 3 monitors in identifying and classifying OSA varies with the location of testing (in-laboratory vs home) and the method used for scoring (automated vs manual). OSA severity can also affect automated scoring accuracy, which improves with increasing disease severity.¹⁰ Not surprisingly, visual (manual) scoring of portable monitoring data is superior to automated scoring and improves the reliability of automated scoring for both type 3¹¹⁻¹⁶ and type 4¹⁷⁻²³ portable monitors when AHI is used to identify and classify OSA. Similar to the current findings, previously examined automated algorithms tend to underestimate the AHI compared with manual scoring.^{11,12,14} Underestimation of the AHI could lead to misclassification of patients with mild OSA as being normal and thus limit the utility of automated scoring algorithms in screening for OSA. In contrast to the AHI, higher concordance between manual and automated

scoring was observed with the ODI. The differences in bias comparing ODI and AHI is not surprising given that estimation of the ODI is based on assessing changes in a single channel (ie, oximetry) that is characterized by slower oscillations (ie, desaturations) compared with airflow.

The present study has several strengths that merit discussion. First, the large sample size adds to the generalizability of the findings. Second, the use of two portable monitoring devices with distinct automated scoring software allowed for the assessment of heterogeneity across different devices. Ideally, any portable sleep monitoring device deployed for clinical care should be evaluated for concordance between automated and manual scoring given the proprietary content and nature of these devices. Although the current findings cannot be extrapolated to other monitors, the consistency between the two devices studied suggests potential underestimation of disease severity, which, as previously noted, could limit the use of these devices in population

Figure 3 – A-D, Scatter plots of automated vs manual scoring of the ODI for the ApneaLink Plus (A, B) and Embletta (C, D) monitors using $a \geq 3\%$ (A, C) and $\geq 4\%$ (B, D) oxygen desaturation threshold. Pearson correlation coefficients and the line of identity (dashed line) are shown for each plot. ODI = oxygen desaturation index.

screening efforts for OSA. Third, and perhaps most importantly, inclusion of community-based and clinic-based samples across the full spectrum of OSA severity that were not recruited based on sleep-related complaints increases the generalizability of the results given that these populations are most likely those targeted for home sleep testing.

The current study also has several important limitations. Expert review was performed by a single scorer, and variability across scorers could have undoubtedly influenced the level of misclassification of OSA severity between automated and manual scoring. However, the intraclass correlation coefficient for the scoring of respiratory events on a sample of recordings between the physician scorer and another scorer was 0.98. Data reported in several studies, which have been summarized by the American Academy of Sleep Medicine Task Force, show that interscorer and intrascorer reliability are indeed good when hypopneas are defined by an oxy-hemoglobin desaturation as done in the current study.²⁴

Thus, although problems of reliability with manual scoring can influence the difference between automated and manual scoring, any bias is likely to be small. Moreover, although two non-sleep clinic samples were recruited to increase generalizability, the prevalence of OSA in these samples was high. The relatively small number of subjects without OSA can alter the level of agreement between automated and manual scoring. Thus, comparisons of manual and automated scoring in samples with a prevalence of OSA equal to that in the general population would represent a natural extension of the current work because they would help to determine the utility of such algorithms in population-wide screening for OSA.

Given that many non-sleep-trained physicians are the first-line providers for patients with such chronic diseases as COPD, type 2 diabetes, hyperlipidemia, and hypertension, it is not unreasonable to suggest that they also manage OSA. However, for most health-care providers, adding OSA with its associated portable

Figure 4 – A-D, Bland-Altman plots for automated vs manual scoring of the ODI for the ApneaLink Plus (A, B) and Embletta (C, D) monitors using a $\geq 3\%$ (A, C) and $\geq 4\%$ (B, D) oxygen desaturation threshold. Average bias (solid line) and limits of agreement (dashed line) are shown for each device and oxygen desaturation threshold. See Figure 3 legend for expansion of abbreviation.

monitoring set-up, diagnosis, patient education, and treatment to an already long list of diseases to diagnose and manage may be overwhelming. Thus far, the evidence suggests that identification of adult patients with OSA in the primary care setting has been low, with a significant proportion of patients not being identified

because no systematic recommendations or processes are in place to determine when and how to screen for OSA and other sleep disorders.²⁵ The solution may be collaboration between sleep specialists and other health-care providers so that home sleep monitoring can be deployed in the most diagnostically accurate,

		Apnealink (Automated Scoring)							
		(3% desaturation threshold)				(4% desaturation threshold)			
		Normal	Mild	Moderate	Severe	Normal	Mild	Moderate	Severe
Apnealink (Manual Scoring)	Normal	6.0%	0.0%	0.0%	0.0%	13.0%	0.0%	0.0%	0.0%
	Mild	3.0%	11.0%	0.0%	0.0%	1.0%	17.0%	1.0%	0.0%
	Moderate	0.0%	6.0%	29.0%	1.0%	0.0%	7.0%	27.0%	0.0%
	Severe	0.0%	0.0%	12.0%	32.0%	0.0%	0.0%	9.0%	25.0%
		N=100				N=100			

Figure 5 – Classification of OSA with the ApneaLink Plus monitor using the $\geq 3\%$ and $\geq 4\%$ oxygen desaturation threshold to define the AHI. Gray shading represents complete agreement between automated and manual scoring. Black shading represents misclassification likely to have an impact on clinical decision-making regarding OSA therapy. See Figure 1 legend for expansion of abbreviation.

		Embletta (Automated Scoring)							
		(3% desaturation threshold)				(4% desaturation threshold)			
		Normal	Mild	Moderate	Severe	Normal	Mild	Moderate	Severe
Embletta (Manual Scoring)	Normal	2.0%	0.0%	0.0%	0.0%	6.0%	0.0%	0.0%	0.0%
	Mild	5.0%	2.0%	0.0%	0.0%	13.0%	22.0%	0.0%	0.0%
	Moderate	4.0%	24.0%	29.0%	6.0%	7.0%	28.0%	13.0%	0.0%
	Severe	0.0%	3.0%	13.0%	12.0%	0.0%	2.0%	7.0%	2.0%
N=100					N=100				

Figure 6 – Classification of OSA with the Embletta monitor using the $\geq 3\%$ and $\geq 4\%$ oxygen desaturation threshold to define the AHI. Gray shading represents complete agreement between automated and manual scoring. Black shading represents misclassification likely to have an impact on clinical decision-making regarding OSA therapy. See Figure 1 legend for expansion of abbreviation.

expeditious, cost-effective, and least burdensome manner. The application of computer-aided techniques to help to improve diagnostic efficiency is common practice in the clinical arena. Automated analysis and interpretation of pulmonary function tests, ECGs, and radiographic imaging are examples of innovative uses of technology that assist in disease diagnosis. The information presented herein supports that expert review enhances the accuracy of diagnosis and interpretation²⁶⁻²⁹ and that portable sleep monitoring with automated scoring may be an acceptable approach to diagnosing OSA with sufficient accuracy in the majority of cases. As is the case with most diagnostic tests, however, the potential for misclassification of disease severity with portable sleep monitoring is greatest in those with mild disease, and thus, decisions regarding additional testing or clinical management should integrate the larger clinical context. An alliance between sleep-trained and other health-care providers with development of systematic recommendations and some oversight and comprehension of the automated scoring systems could potentially improve the diagnostic accuracy of these devices, decrease misclassification, expedite care, and reduce costs while minimizing or avoiding additional time burden for clinical practitioners. The burden

imposed by the manual scoring of respiratory events is not trivial. In the current study, the time required for manual scoring varied as a function of OSA severity. Studies in subjects with mild or no OSA required approximately 10 min, whereas studies in subjects with moderate or severe OSA required 60 to 90 min. Manual editing of automated scoring would be an acceptable alternative and one that seems to be commonly implemented in clinical practice. Moving forward, it is important to learn from the use of available automated diagnostic technologies, to achieve the optimum balance between cost and quality of OSA management, and to allocate health-care resources in the most effective manner.

In conclusion, the current investigation suggests that agreement between automated and manual scoring of home sleep tests varies as a function of the portable device and definition of disordered breathing event used. Although modest agreement exists between automated and manual scoring, input by a sleep specialist or a certified polysomnologist in reviewing home sleep study results may help to improve diagnostic accuracy and classification of OSA, particularly if there is mild disease.

Acknowledgments

Author contributions: N. M. P. had full access to all of the data in the study and responsibility for the integrity of the data and the accuracy of the data analysis. R. N. A. contributed to the data collection, statistical analysis and interpretation, initial drafting of the manuscript, critical review of the manuscript for important intellectual content, and final approval of the manuscript; R. S. contributed to the data collection, critical review of the manuscript for important intellectual content, and final approval of the manuscript; and N. M. P. contributed to the study concept and design, statistical analysis and interpretation, initial drafting of the manuscript, critical review of the manuscript for important intellectual content, and final approval of the manuscript

Financial/nonfinancial disclosures: The authors have reported to *CHEST* the following conflicts of interest: Dr Punjabi has received grant support from ResMed and Koninklijke Philips NV (Respironics) unrelated to the subject of this article. Dr Aurora and Ms Swartz have reported that no potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

Role of sponsors: The sponsor had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript.

References

- Punjabi NM. The epidemiology of adult obstructive sleep apnea. *Proc Am Thorac Soc*. 2008;5(2):136-143.
- Stradling JR, Davies RJ. Sleep. 1: obstructive sleep apnoea/hypopnoea syndrome: definitions, epidemiology, and natural history. *Thorax*. 2004;59(1):73-78.
- Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med*. 2002;165(9):1217-1239.
- Gibson GJ. Obstructive sleep apnoea syndrome: underestimated and undertreated. *Br Med Bull*. 2004;72(1):49-65.
- Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep*. 1997;20(9):705-706.
- Gharibeh T, Mehra R. Obstructive sleep apnea syndrome: natural history, diagnosis, and emerging treatment options. *Nat Sci Sleep*. 2010;2010(2):233-255.
- Decision memo for sleep testing for obstructive sleep apnea (OSA) (CAG-00405N). Centers for Medicare & Medicaid Services website. <https://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/downloads/R86NCD.pdf>. Accessed December 15, 2014.
- Collop NA, Anderson WM, Boehlecke B, et al; Portable Monitoring Task Force of the American Academy of Sleep Medicine. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. *J Clin Sleep Med*. 2007;3(7):737-747.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307-310.
- El Shayeb M, Topfer LA, Stafinski T, Pawluk L, Menon D. Diagnostic accuracy of level 3 portable sleep tests versus level 1 polysomnography for sleep-disordered breathing: a systematic review and meta-analysis. *CMAJ*. 2014;186(1):E25-E51.
- Calleja JM, Esnaola S, Rubio R, Durán J. Comparison of a cardiorespiratory device versus polysomnography for diagnosis of sleep apnoea. *Eur Respir J*. 2002;20(6):1505-1510.
- Carrasco O, Montserrat JM, Lloberes P, et al. Visual and different automatic scoring profiles of respiratory variables in the diagnosis of sleep apnoea-hypopnoea syndrome. *Eur Respir J*. 1996;9(1):125-130.
- Dingli K, Coleman EL, Vennelle M, et al. Evaluation of a portable device for diagnosing the sleep apnoea/hypopnoea syndrome. *Eur Respir J*. 2003;21(2):253-259.
- Fietze I, Glos M, Röttig J, Witt C. Automated analysis of data is inferior to visual analysis of ambulatory sleep apnea monitoring. *Respiration*. 2002;69(3):235-241.
- Masa JF, Corral J, Pereira R, et al; Spanish Sleep Group. Effectiveness of sequential automatic-manual home respiratory polygraphy scoring. *Eur Respir J*. 2013;41(4):879-887.
- Zucconi M, Ferini-Strambi L, Castronovo V, Oldani A, Smirne S. An unattended device for sleep-related breathing disorders: validation study in suspected obstructive sleep apnoea syndrome. *Eur Respir J*. 1996;9(6):1251-1256.
- BaHammam AS, Sharif M, Gacuan DE, George S. Evaluation of the accuracy of manual and automatic scoring of a single airflow channel in patients with a high probability of obstructive sleep apnea. *Med Sci Monit*. 2011;17(2):MT13-MT19.
- Chen H, Lowe AA, Bai Y, Hamilton P, Fleetham JA, Almeida FR. Evaluation of a portable recording device (ApneaLink) for case selection of obstructive sleep apnea. *Sleep Breath*. 2009;13(3):213-219.
- Crowley KE, Rajaratnam SM, Shea SA, Epstein LJ, Czeisler CA, Lockley SW; Harvard Work Hours, Health and Safety Group. Evaluation of a single-channel nasal pressure device to assess obstructive sleep apnea risk in laboratory and home environments. *J Clin Sleep Med*. 2013;9(2):109-116.
- Esnaola S, Durán J, Infante-Rivard C, Rubio R, Fernández A. Diagnostic accuracy of a portable recording device (MESAM IV) in suspected obstructive sleep apnoea. *Eur Respir J*. 1996;9(12):2597-2605.
- Koziej M, Cieśllicki JK, Gorzelak K, Sliwiński P, Zieliński J. Hand-scoring of MESAM 4 recordings is more accurate than automatic analysis in screening for obstructive sleep apnoea. *Eur Respir J*. 1994;7(10):1771-1775.
- Nigro CA, Dibur E, Aimaretti S, González S, Rhodius E. Comparison of the automatic analysis versus the manual scoring from ApneaLink™ device for the diagnosis of obstructive sleep apnoea syndrome. *Sleep Breath*. 2011;15(4):679-686.
- Oktay B, Rice TB, Atwood CW Jr, et al. Evaluation of a single-channel portable monitor for the diagnosis of obstructive sleep apnea. *J Clin Sleep Med*. 2011;7(4):384-390.
- 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.
- Mold JW, Quattlebaum C, Schinnerer E, Boeckman L, Orr W, Hollabaugh K. Identification by primary care clinicians of patients with obstructive sleep apnea: a practice-based research network (PBRN) study. *J Am Board Fam Med*. 2011;24(2):138-145.
- Estes NA III. Computerized interpretation of ECGs: supplement not a substitute. *Circ Arrhythm Electrophysiol*. 2013;6(1):2-4.
- Garg A, Lehmann MH. Prolonged QT interval diagnosis suppression by a widely used computerized ECG analysis system. *Circ Arrhythm Electrophysiol*. 2013;6(1):76-83.
- Geddes DM, Green M, Emerson PA. Comparison of reports on lung function tests made by chest physicians with those made by a simple computer program. *Thorax*. 1978;33(2):257-260.
- Shiraishi J, Li Q, Appelbaum D, Doi K. Computer-aided diagnosis and artificial intelligence in clinical imaging. *Semin Nucl Med*. 2011;41(6):449-462.