

Research Article

Accuracy of Positive Airway Pressure Device—Measured Apneas and Hypopneas: Role in Treatment Followup

Carl Stepnowsky,^{1,2} Tania Zamora,¹ Robert Barker,³ Lin Liu,⁴ and Kathleen Sarmiento^{2,3}

¹ Health Services Research & Development Unit, Veterans Affairs San Diego Healthcare System, San Diego, CA 92161, USA

² Department of Medicine, University of California, San Diego, CA 92037, USA

³ Pulmonary Service, Veterans Affairs San Diego Healthcare System, San Diego, CA 92161, USA

⁴ Department of Family and Preventive Medicine, University of California, San Diego, CA 92037, USA

Correspondence should be addressed to Carl Stepnowsky; cstepnowsky@ucsd.edu

Received 30 April 2013; Revised 16 July 2013; Accepted 19 July 2013

Academic Editor: Giora Pillar

Copyright © 2013 Carl Stepnowsky et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Improved data transmission technologies have facilitated data collected from positive airway pressure (PAP) devices in the home environment. Although clinicians' treatment decisions increasingly rely on autoscorer of respiratory events by the PAP device, few studies have specifically examined the accuracy of autoscored respiratory events in the home environment in ongoing PAP use. "PAP efficacy" studies were conducted in which participants wore PAP simultaneously with an Emblette sleep system (Embla, Inc., Broomfield, CO), which was directly connected to the ResMed AutoSet S8 (ResMed, Inc., San Diego, CA) via a specialized cable. Mean PAP-scored Apnea-Hypopnea Index (AHI) was 14.2 ± 11.8 (median: 11.7; range: 3.9–46.3) and mean manual-scored AHI was 9.4 ± 10.2 (median: 7.7; range: 1.2–39.3). Ratios between the mean indices were calculated. PAP-scored HI was 2.0 times higher than the manual-scored HI. PAP-scored AHI was 1.5 times higher than the manual-scored AHI, and PAP-scored AI was 1.04 of manual-scored AI. In this sample, PAP-scored HI was on average double the manual-scored HI. Given the importance of PAP efficacy data in tracking treatment progress, it is important to recognize the possible bias of PAP algorithms in overreporting hypopneas. The most likely cause of this discrepancy is the use of desaturations in manual hypopnea scoring.

1. Introduction

Obstructive sleep apnea (OSA) is a chronic medical condition requiring nightly application of therapy to effectively limit the number of apneas and hypopneas that would occur without intervention. The gold-standard treatment for OSA is continuous positive airway pressure therapy (PAP), which provides a pneumatic splint of the soft tissue in the upper airway [1]. PAP devices can measure and record airflow and pressure levels whenever the device is worn. They contain internal, proprietary (i.e., differing by manufacturer) algorithms that identify breathing disturbances and whether these disturbances are due to persistent obstructive or nonobstructive events. Thus, PAP devices can provide a measure of "residual" Apnea-Hypopnea Index (AHI) and its components, the Hypopnea Index (HI) and Apnea Index (AI). Although not equivalent to the indices measured by polysomnography or

home sleep testing via Type III devices, the PAP terminology is nonetheless the same.

American Academy of Sleep Medicine practice parameters and clinical guidelines recommend routine monitoring of adherence and efficacy data provided by PAP devices as an indication of treatment progress [2, 3]. Because residual AHI is primarily used to inform pressure changes and because its measurement by the PAP device is different relative to polysomnography (PSG) or Type III devices, it requires further study. PAP-scored AHI is different from that scored by PSG for two main reasons: (1) PAP measures are based solely on an airflow signal, and (2) they are based on an automated, proprietary algorithm. Several studies have examined PAP-scored AHI but have primarily attempted to evaluate the ability of the PAP device (autoadjusting PAP, in particular) to provide an initial baseline AHI value. Most have reported a strong correlation between PAP-scored AHI and

manual-scored AHI [4, 5]. However, a certain percentage of AHI values would have resulted in different classifications, which can affect clinical management decisions.

A related but different issue concerns the accuracy of the Apnea-Hypopnea Index (AHI), as measured by the PAP unit in the home environment for the purposes of treatment efficacy (i.e., after a period of use). AHI accuracy is particularly important, given the increasing use of and reliance upon PAP data by providers, patients, and intermediaries (i.e., durable medical equipment staff). Ambulatory models of OSA care are gaining popularity, particularly the use of autotitrating PAP devices in lieu of in-laboratory CPAP titrations. In contrast to fixed pressure devices, which simply count the number of apneas occurring while PAP is applied, autoadjusting devices can make pressure changes based on the identification of these disturbances. With an ever-increasing demand for sleep apnea care, the ability to identify patients who may not be therapeutic on their PAP devices is critical. Efficacy of therapy is also an important factor in patient adherence. New technologies allow for data transmission directly from the PAP device to software accessible to the provider and, more recently, to the patients themselves (e.g., SleepMapper, Philips Respironics, Murrysville, PA). A variety of data transmission methods are possible, including the use of a smartcard, wired modem (via telephone line), wireless modem (via cellular network), and, more recently, Bluetooth modems to connect directly into home computers, tablets, or Smartphones. Remote monitoring is a trend within healthcare that is clearly accelerating, and in the sleep field, it facilitates the evaluation of compliance and efficacy of PAP therapy [6].

Given the improved PAP data transmission technologies and resultant increased use of these data, we sought to investigate the accuracy of the PAP-measured AHI. We had the opportunity to conduct “PAP efficacy” studies in which participants wore PAP devices simultaneously with Type III cardiopulmonary recording equipment. Therefore, the goal of the present study was to specifically examine the accuracy of the identification of apneas and hypopneas by the PAP device.

2. Methods

2.1. Procedures. Twelve research participants from a larger trial evaluating a PAP adherence intervention were included in this study. The PAP adherence intervention study compared a usual care group to a group that was provided with extra education and clinical support via interactive website, phone calls, and in-person clinic visits [7]. They were also provided with daily access to their PAP data. Inclusion criteria for the study were purposefully broad and included those diagnosed with OSA (as defined by AHI >15 with predominately obstructive events) and prescribed PAP therapy. Participants who had a clinical indication for performing an efficacy study (e.g., either high residual PAP-measured AHI or subjective report that was inconsistent with PAP data) [3] were included. These participants underwent a home efficacy study, in which autoadjusting positive airway pressure therapy (APAP) devices was worn simultaneously with Embletta, a Type III cardiopulmonary recording device.

TABLE 1: Baseline characteristics.

| | Mean \pm SD | Range |
|--------------------------------|-----------------|-----------|
| Age | 62.0 \pm 12.3 | 46–78 |
| Body Mass Index (BMI) | 30.4 \pm 5.9 | 21.5–41.7 |
| Apnea-Hypopnea Index (AHI) | 42.9 \pm 21.5 | 11.3–76.9 |
| Epworth Sleepiness Scale (ESS) | 14.2 \pm 3.9 | 8.0–19.0 |

2.2. Equipment Used. The Embletta (Embla, Inc., Broomfield, CO) was directly connected to the ResMed AutoSet S8 (ResMed, Inc., San Diego, CA) via a specialized cable that allowed for the direct recording of S8 data. Signals recorded include oximetry, chest effort, and body position. Airflow from the PAP device was used for scoring. RemLogic software was used for manual respiratory scoring. Apneas and hypopneas were manually scored according to the 2007 American Academy of Sleep Medicine guidelines, which included defining a hypopnea as being associated with a $\geq 4\%$ oxygen desaturation [8]. AutoSet respiratory events were autoscored by the device, and summary statistics were obtained within RemLogic. Manual scoring was blind to the AutoSet-scored respiratory events.

2.3. Data Analysis. Descriptive statistics (mean, median, and standard deviation and range) were calculated for the AHI, HI, and AI data. Scatterplots were generated to show the relationship between PAP-scored and manual-scored indices and included the line of identity. Spearman correlation coefficient was calculated. Wilcoxon signed rank test was used to test mean difference between the indices, and concordance correlation coefficients [9, 10] with 95% confidence interval (CI) were used to assess the agreement between PAP-scored and manual-scored indices. The concordance correlation coefficient less than 0.90 is interpreted as poor agreement, 0.90–0.95 as moderate, 0.95–0.99 as substantial, and greater than 0.99 as almost perfect [11]. Bland-Altman plots were created to provide a visualization of the bias and limits of agreement [12]. Data were analyzed using R [13].

3. Results

Participants were primarily overweight, middle aged, and sleepy with moderate to severe OSA (see Table 1) who had been using PAP on average 84.8 days prior to the PAP efficacy study. There were 6 men and 6 women.

Mean PAP-scored AHI was 14.2 ± 11.8 (median: 11.7, range: 3.9–46.3) and mean manual-scored AHI was 9.4 ± 10.2 (median: 7.7, range: 1.2–39.3) (see Table 2). Paired sample means testing found differences between the PAP AHI and manual AHI (mean difference = 4.84, median difference = 4.25, $P < 0.001$) and between the PAP HI and manual HI (mean difference = 4.45, median difference = 3.5, $P < 0.001$), but not between PAP AI and manual AI (mean difference = 0.21, median difference = 0.4, $P = 0.53$). The correlation coefficient between PAP-scored and manual-scored AHI, AI, and HI was 0.93 ($P < 0.001$), 0.92 ($P < 0.001$), and 0.87 ($P < 0.001$), respectively.

TABLE 2: OSA variables measured on efficacy study.

| | Manual scoring | | Autoscorer | |
|---------------------------------|----------------|----------|-----------------|-----------|
| | Mean \pm SD | Range | Mean \pm SD | Range |
| Apnea-Hypopnea Index (AHI) | 9.4 \pm 10.2 | 1.2–39.3 | 14.2 \pm 11.8 | 3.9–46.3 |
| Apnea Index (AI) | 5.1 \pm 7.9 | 0–28.4 | 5.3 \pm 7.7 | 0.20–27.6 |
| Central Apnea Index (CAI) | 2.6 \pm 6.6 | 0–23.5 | — | — |
| Obstructive Apnea Index (OAI) | 1.5 \pm 2.3 | 0–7.7 | — | — |
| Mixed Apnea Index (MAI) | 0.9 \pm 1.3 | 0–4.6 | — | — |
| Hypopnea Index (HI) | 4.3 \pm 3.0 | 0.5–10.9 | 8.8 \pm 4.9 | 3.1–18.7 |
| Oxygen Desaturation Index (ODI) | 7.7 \pm 7.4 | 0.6–27.2 | — | — |

Concordance correlation coefficient is 0.87 (95% CI: 0.71–0.95) for AHI, 0.994 (95% CI: 0.98–0.998) for AI, and 0.50 (95% CI: 0.23–0.70) for HI. Based on published guidelines, the agreement between PAP and manual scoring is considered “poor” in AHI and HI and is “almost perfect” for AI.

Ratios between the mean indices were calculated. The PAP-scored HI was 2.0 times higher than the manual-scored HI, the PAP-scored AHI was 1.5 times higher than the manual-scored AHI, and the PAP-scored AI was 1.04 of the manual-scored AI. It appears that the PAP device evaluated in this study, relative to manual scoring, only slightly overscored the number of apneas but significantly overscored the number of hypopneas. The difference in scoring of hypopneas seems to be the main contributor to the different AHI values between the PAP device and manual scoring.

Two graphical displays of the data were created. Figures 1(a)–1(c) show the scatterplots for the three indices, including the line of identity. In each case, the PAP-scored index was higher than the corresponding manual-scored index. Figures 2(a)–2(c) show the Bland-Altman plots. The mean bias (95% limit of agreement) is 4.84 (−0.95 to 10.6) for AHI, 0.28 (−1.50 to 1.91) for AI, and 4.45 (−0.17 to 9.07) for HI.

4. Discussion

The practice of sleep medicine is evolving, and ambulatory models of sleep apnea management using home sleep testing and APAP therapy are not only noninferior to traditional evaluations but are also gaining wider acceptance by sleep providers [14]. Home sleep testing, or cardiorespiratory polygraphy, is indicated for the diagnosis of OSA in patients with a high pretest probability of moderate to severe OSA, to monitor efficacy of non-PAP therapies for OSA, and may be indicated in those who would otherwise not be recommended for home evaluation but who cannot undergo in-laboratory diagnostic testing [15]. Emphasis is placed on reviewing the raw data from home sleep tests to ensure accurate diagnoses. Similarly, review of downloaded data from PAP machines is of great importance in determining efficacy of therapy and should guide decisions to change PAP settings. Thus, in an era of increasing dependence on efficacy and compliance information in the clinical management of sleep apnea patients, a greater understanding of how to interpret this information is needed.

In this study of home-based PAP efficacy, as measured by the S8 APAP device, the PAP-scored HI was on average more than double the manual-scored HI. Given the importance of PAP efficacy data in tracking treatment progress, it is important to recognize that this particular APAP device may overscore hypopneas. The most likely causes of this discrepancy are (a) the use of a proprietary algorithm and (b) the use of desaturations in manual hypopnea scoring. Because the number of apneas was underscored relative to manual scoring, the overall AHI does not appear to be different from manual scoring. This study and the evolving literature in this area suggest that it is important to understand how a specific PAP device identifies both apneas and hypopneas.

One previous study that used the S8 device also found relatively good apnea measurement but an overscoring of hypopneas [16]. That study found that the PAP HI was 3.3 times higher than the manual HI, and the resulting AHI was just over two times greater. Those values are slightly higher than the values found in the present study, but both speak to the importance of understanding the scoring algorithms for apneas and hypopneas of a specific PAP device so that treatment decisions are well informed. If it is found that, on average, a specific PAP device scores hypopneas at a rate of 2.0 times greater than manual scoring, then an adjustment can be made by the provider. For example, in the case where the measured HI is 20, the adjustment can be made by dividing 20 by the factor of 2 or an HI of 10 (which would theoretically be comparable to manual scoring).

Other studies in this area have utilized the RemStar autoadjusting PAP device by Philips Respironics. These study results show a different pattern, specifically that respiratory event detection varies based on the number of events. For example, RemStar-measured AHI tended to overestimate the AHI at lower AHI levels but underestimate the AHI at higher AHI levels [5, 17]. In short, it appears that AHI measurement is dependent on the specific APAP device used.

If there are systematic differences between PAP devices, it is important for the field to request that the manufacturers provide clinicians and researchers with clear information regarding what level of adjustment is necessary to allow for the most accurate interpretation of the PAP-scored apneas and hypopneas. The PAP-scored AHI value is a useful data point for gathering information on therapeutic efficacy. Previous studies have examined the percentage of patients that continue to have residual OSA even while using a PAP device.

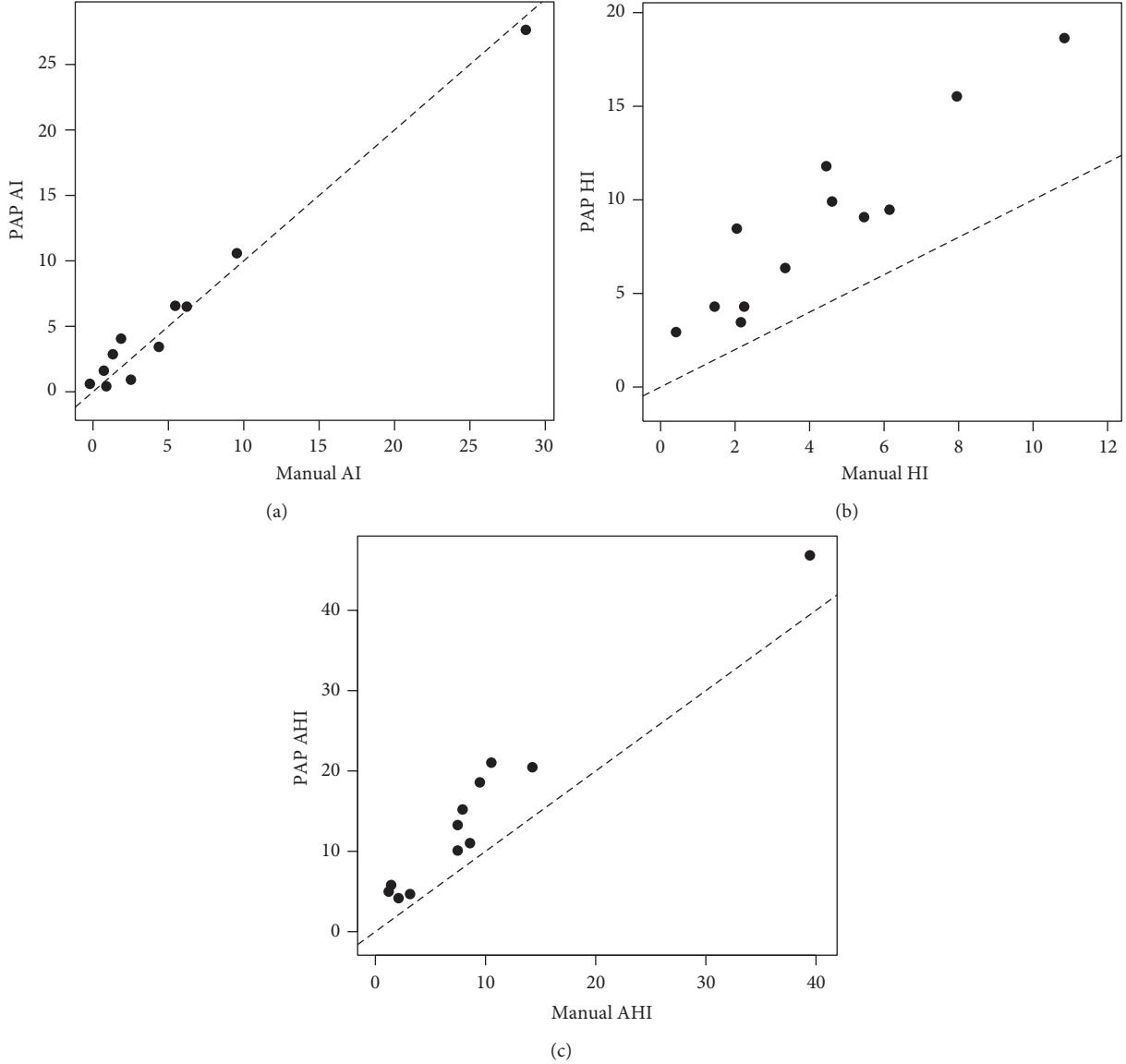


FIGURE 1: (a) Scatterplot of PAP-scored versus manual-scored Apnea Index. Each point represents one night. The diagonal line represents the line of identity. (b) Scatterplot of PAP-scored versus manual-scored Hypopnea Index. Each point represents one night. The diagonal line represents the line of identity. Note that all PAP HI values are greater than manual HI values. (c) Scatterplot of PAP-scored versus manual-scored Apnea-Hypopnea Index. Each point represents one night. The diagonal line represents the line of identity. Note that all values of PAP AHI are greater than manual AHI values.

In a study of patients using single-pressure CPAP, nearly 20% continued to have PAP AHI >10 after 3 months [18], while in another study of patients undergoing a home APAP trial, 29% had PAP AHI >10 [19]. The former study did not specify the CPAP device used, while the latter study used a ResMed AutoSet Spirit. Given the results of the current study and the associated literature, it would appear that the unique PAP device algorithms for automatic respiratory event detection affect the results of these and similar studies. Given the findings of the present study, it is possible that the study using the ResMed AutoSet has inflated AHI values, and therefore, the residual AHI in that study may be less than actually reported [19].

As per published clinical guidelines, the standard recommendation is that sleep monitoring is indicated for the assessment of treatment results on PAP therapy after (i) substantial weight loss (e.g., 10% of body weight) to ascertain whether PAP therapy is still needed at the prescribed pressure settings, (ii) substantial weight gain with return of symptoms (e.g., 10% of body weight) to ascertain whether pressure adjustments are needed, (iii) clinical response is insufficient (e.g., lack of symptom relief, above normal residual AHI, or poor adherence), or (iv) symptoms return despite a good initial response to CPAP [3, 20].

There are a number of potential study limitations. First, the number of participants is low relative to other studies

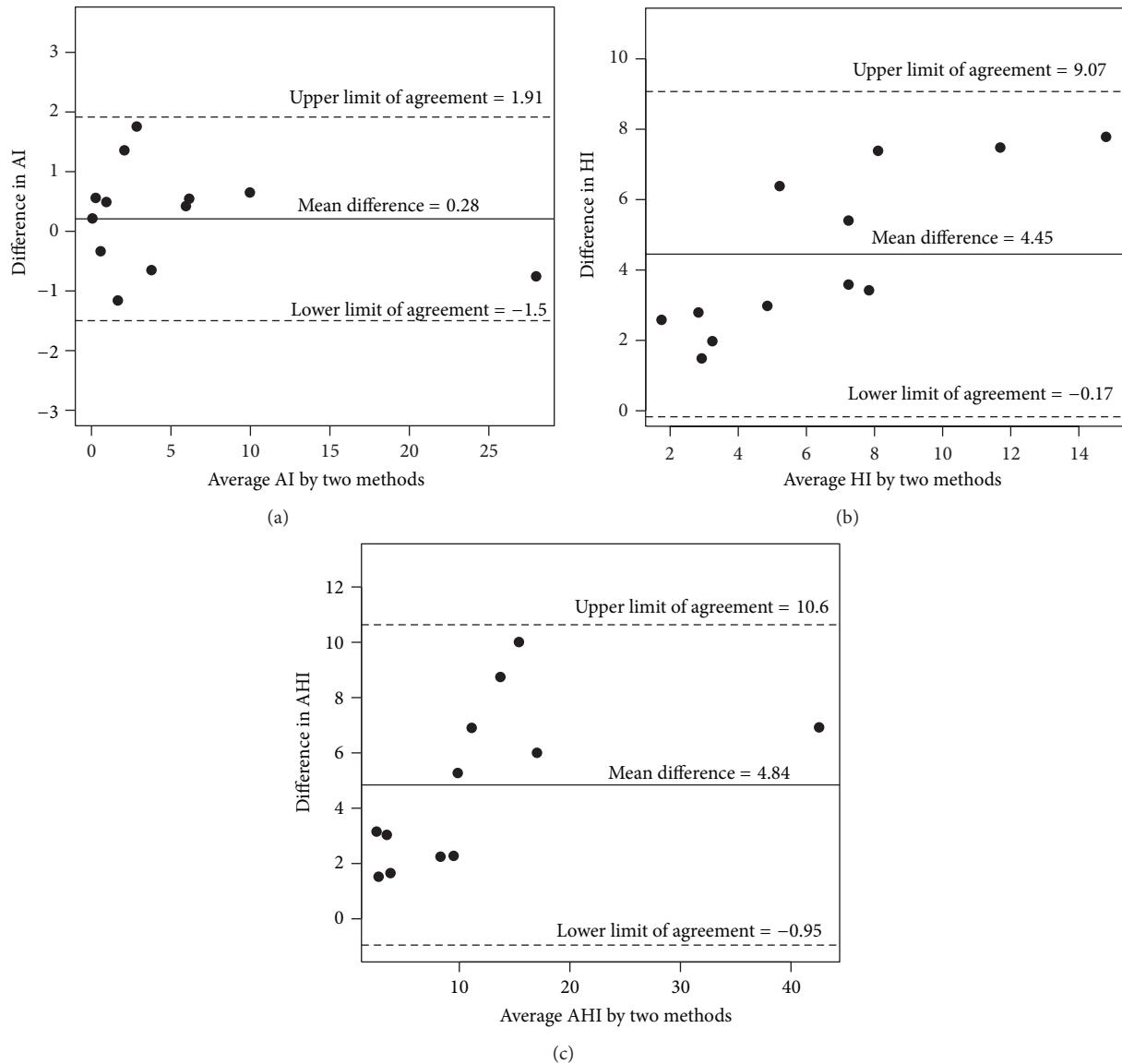


FIGURE 2: (a) Bland-Altman plot of PAP scoring versus manual scoring AI difference (PAP-scoring minus manual-scoring) by the mean. (b) Bland-Altman plot of PAP scoring versus manual-scoring HI difference (PAP scoring minus manual-scoring) by the mean. Note that all of the difference values were greater than 0, indicating that all PAP HI values were greater than the manual HI values. (c) Bland-Altman plot of PAP-scoring versus manual-scoring AHI difference (PAP scoring minus manual-scoring) by the mean. Note that all of the difference values were greater than 0, indicating that all PAP AHI values were greater than the manual AHI values.

in this area. That said, the results are consistent with the only other study that used the S8 APAP device. Second, the study was done in the home environment where there is less control than that in a laboratory environment. Third, there are limitations in scoring hypopneas with Type III home testing (events must meet flow and $\geq 4\%$ saturation criteria but cannot be scored based on an arousal, as sleep is not measured). Finally, newer generation of PAP devices can distinguish central and obstructive events. Future studies can utilize this newer technology to see how well PAP algorithms can distinguish these events and what role they have in overall AHI values.

In summary, PAP devices have automated, proprietary algorithms for respiratory event detection. When event

detection scoring is combined with PAP use duration in the denominator, a proxy AHI value is derived. Given the increased reliance on the PAP-scored events by both providers and patients, it is important to better understand the nuances of specific algorithms and how the PAP-scored AHI, HI, and AI values compare to those same values from manual scoring. Doing so is an important step toward making more informed treatment decisions.

Acknowledgments

This research was supported in part by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, and the Research Service of

the San Diego Veterans Affairs Healthcare System (VA HSRD IIR 02-275 and 07-163 and AHRQ R18 017246).

References

- [1] C. E. Sullivan, F. G. Issa, M. Berthon-Jones, and L. Eves, "Reversal of obstructive sleep apnoea by continuous positive airway pressure applied through the nares," *Lancet*, vol. 1, no. 8225, pp. 862–865, 1981.
- [2] T. I. Morgenthaler, R. N. Aurora, T. Brown et al., "Practice parameters for the use of autotitrating continuous positive airway pressure devices for titrating pressures and treating adult patients with obstructive sleep apnea syndrome: an update for 2007. An American Academy of Sleep Medicine Report," *Sleep*, vol. 31, no. 1, pp. 141–147, 2008.
- [3] L. J. Epstein, D. Kristo, P. J. Strollo Jr. et al., "Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults," *Journal of Clinical Sleep Medicine*, vol. 5, no. 3, pp. 263–276, 2009.
- [4] A. L. Denotti, K. K. Wong, G. C. Dungan II, J. W. Gilholme, N. S. Marshall, and R. R. Grunstein, "Residual sleep-disordered breathing during autotitrating continuous positive airway pressure therapy," *European Respiratory Journal*, vol. 39, pp. 1391–1397, 2012.
- [5] R. B. Berry, C. A. Kushida, M. H. Kryger, H. Soto-Calderon, B. Staley, and S. T. Kuna, "Respiratory event detection by a positive airway pressure device," *Sleep*, vol. 35, no. 3, pp. 361–367, 2012.
- [6] E. J. Topol, "Transforming medicine via digital innovation," *Science Translational Medicine*, vol. 2, no. 16, article 16cm4, 2010.
- [7] C. Stepnowsky, C. Edwards, T. Zamora, R. Barker, and Z. Agha, "Patient perspective on use of an interactive website for sleep apnea," *International Journal of Telemedicine and Applications*, vol. 2013, Article ID 239382, 10 pages, 2013.
- [8] C. Iber, S. Ancoli-Israel, A. Chesson, and S. F. Quan, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specification*, American Academy of Sleep Medicine, Westchester, Ill, USA, 2007.
- [9] L. I-Kuei Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, vol. 45, no. 1, pp. 255–268, 1989.
- [10] L. I. K. Lin, "A note on the concordance correlation coefficient," *Biometrics*, vol. 56, pp. 324–325, 2000.
- [11] G. B. McBride, "A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient," NIWA Client Report, National Institute of Water & Atmospheric Research, Ltd, Hamilton, New Zealand, 2005.
- [12] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 1, no. 8476, pp. 307–310, 1986.
- [13] R Development Core Team, *R: A Language and Environment For Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2009.
- [14] S. T. Kuna, I. Gurubhagavatula, G. Maislin et al., "Noninferiority of functional outcome in ambulatory management of obstructive sleep apnea," *American Journal of Respiratory and Critical Care Medicine*, vol. 183, no. 9, pp. 1238–1244, 2011.
- [15] N. A. Collop, W. M. Anderson, B. Boehlecke et al., "Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients," *Journal of Clinical Sleep Medicine*, vol. 3, no. 7, pp. 737–747, 2007.
- [16] K. Ueno, T. Kasai, G. Brewer et al., "Evaluation of the apnea-hypopnea index determined by the S8 auto-CPAP, a continuous positive airway pressure device, in patients with obstructive sleep apnea-hypopnea syndrome," *Journal of Clinical Sleep Medicine*, vol. 6, no. 2, pp. 146–151, 2010.
- [17] A. Cilli, R. Uzun, and U. Bilge, "The accuracy of autotitrating CPAP-determined residual apnea-hypopnea index," *Sleep and Breathing*, pp. 1–5, 2012.
- [18] M. A. Baltzan, I. Kassissia, O. Elkholy, M. Palayew, R. Dabrusin, and N. Wolkove, "Prevalence of persistent sleep apnea in patients treated with continuous positive airway pressure," *Sleep*, vol. 29, no. 4, pp. 557–563, 2006.
- [19] L. Torre-Bouscoulet, M. S. Meza-Vargas, A. Castorena-Maldonado, M. Reyes-Zúñiga, and R. Pérez-Padilla, "Autoadjusting positive pressure trial in adults with sleep apnea assessed by a simplified diagnostic approach," *Journal of Clinical Sleep Medicine*, vol. 4, no. 4, pp. 341–347, 2008.
- [20] C. A. Kushida, M. R. Littner, T. Morgenthaler et al., "Practice parameters for the indications for polysomnography and related procedures: an update for 2005," *Sleep*, vol. 28, no. 4, pp. 499–521, 2005.

