JCSM | Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Performance of peripheral arterial tonometry–based testing for the diagnosis of obstructive sleep apnea in a large sleep clinic cohort

Octavian C. Ioachimescu, MD, PhD[1,2]; J. Shirine Allam, MD[1,2]; Arash Samarghandi, MD[1]; Neesha Anand, MD[1]; Barry G. Fields, MD, MSEd[1,2]; Swapan A. Dholakia, MD[1,2]; Saiprakash B. Venkateshiah, MD[1,2]; Rina Eisenstein, MD[1,2]; Mary-Margaret Ciavatta, PAC[2]; Nancy A. Collop, MD[1] for the Pulse Arterial Tonometry Evaluation of Reliability (PATER) study investigators

[1]Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Emory University School of Medicine, Atlanta, Georgia; [2]Department of Medicine, Sleep Medicine Center, Atlanta VA Medical Center, Atlanta, Georgia

Study Objectives: Peripheral arterial tonometry (PAT)–based technology represents a validated portable monitoring modality for the diagnosis of OSA. We assessed the diagnostic accuracy of PAT-based technology in a large point-of-care cohort of patients studied with concurrent polysomnography (PSG).
Methods: During study enrollment, all participants suspected to have OSA and tested by in-laboratory PSG underwent concurrent PAT device recordings.
Results: Five hundred concomitant PSG and WatchPat tests were analyzed. Median (interquartile range) PSG AHI was 18 (8–37) events/h and PAT $AHI_{3\%}$ was 25 (12–46) events/h. Average bias was + 4 events/h. Diagnostic concordance was found in 42%, 41%, and 83% of mild, moderate, and severe OSA, respectively (accuracy = 53%). Among patients with PAT diagnoses of moderate or severe OSA, 5% did not have OSA and 19% had mild OSA; in those with mild OSA, PSG showed moderate or severe disease in 20% and no OSA in 30% of patients (accuracy = 69%). On average, using a 3% desaturation threshold, WatchPat overestimated disease prevalence and severity (mean + 4 events/h) and the 4% threshold underestimated disease prevalence and severity by −6 events/h.
Conclusions: Although there was an overall tendency to overestimate the severity of OSA, a significant percentage of patients had clinically relevant misclassifications. As such, we recommend that patients without OSA or with mild disease assessed by PAT undergo repeat in-laboratory PSG. Optimized clinical pathways are urgently needed to minimize therapeutic decisions instituted in the presence of diagnostic uncertainty.
Keywords: home sleep apnea testing, polysomnography, OSA, sleep testing, WatchPat, peripheral arterial tonometry
Citation: Ioachimescu OC, Allam JS, Samarghandi A, et al. Performance of peripheral arterial tonometry–based testing for the diagnosis of obstructive sleep apnea in a large sleep clinic cohort. *J Clin Sleep Med.* 2020;16(10):1663–1674.

BRIEF SUMMARY
Current Knowledge/Study Rationale: Peripheral arterial tonometry (WatchPat 200) technology is a validated modality for the diagnosis of OSA, yet its performance and validity in a large clinic population have not been tested in a systematic fashion.
Study Impact: The WatchPat 200 devices both under- and overestimated the presence and the severity of OSA, generating diagnostic misclassification in approximately 30%–50% of patients; if peripheral arterial tonometry–based testing suggests no or mild OSA, then we recommend repeat testing with gold standard polysomnography. Although insomnia symptoms did not alter the performance of the WatchPat-based testing, we recommend developing specific assessment pathways and improvement algorithms that mitigate the inherent diagnostic uncertainty.

INTRODUCTION

The sleep disorder OSA is associated with major neurocognitive and cardiovascular complications and affects approximately 1 billion people worldwide.[1] Although the condition's prevalence may exceed 50% in some countries,[1] a significant undiagnosed disease burden remains in the general population, and effective and efficacious diagnostic and therapeutic strategies are urgently needed.[2] The gold standard diagnostic procedure for diagnosing OSA, polysomnography (PSG), is laborious, costly, and resource intensive. Home sleep apnea testing (HSAT), also known as portable monitoring,[3] out-of-center testing,[4] or oligosomnography,[5] has provided a major advance in diagnosing OSA with reasonable accuracy, lower cost, and greater accessibility. Most HSAT devices use airflow and effort monitoring for the diagnosis of respiratory events, although sleep or sleep stages are not often

evaluated. Other technologies such as WatchPat devices (Itamar Medical Ltd, Caesarea, Israel) define respiratory events and sleep stages based on proprietary algorithms that incorporate peripheral arterial tonometry (PAT) variability as a surrogate measurement of autonomic sympathetic tone. Although WatchPat devices have been validated and approved for diagnostic use in OSA, systematic large-scale validation of these devices and assessment of diagnostic uncertainty by using this technology have not been performed to date. The current study aims to assess the performance of the PAT-based HSAT wrist-worn portable devices (WatchPat 200) in a large sleep clinic–based cohort of patients.

METHODS

The study included 500 consecutive patients with valid tests who were evaluated in the Atlanta Veteran Affairs Sleep

Medicine Center and referred for PSG testing. Participants were enrolled during 2 distinct, predetermined periods of time: between August 2018 and March 2019 (part 1, n = 240), and between June 2019 and February 2020 (part 2, n = 260). All participants were evaluated concurrently with in-laboratory PSG and PAT-based HSAT devices (WatchPat 200). By design, during part 1 of enrollment, sleep testing was ordered after an initial intake evaluation that included a battery of questionnaires such as the Berlin Questionnaire, the Epworth Sleepiness Scale, the Fatigue Severity Scale, and the Insomnia Severity Scale (ISI). If the index of suspicion for OSA was high (based on the Berlin Questionnaire or a prior diagnosis of OSA) and ISI was abnormal, then a PSG was ordered (and consequently the participant would be included in part 1 of the study); if the ISI was normal, then the patient underwent an HSAT session using WatchPat 200 or other devices, outside the purview of this study. During part 2 of the study, to eliminate any source of laboratory referral bias or triage-induced distortion based on insomnia symptoms (given that by American Academy of Sleep Medicine [AASM] recommendations, patients with insomnia should preferably undergo PSG), all participants with a high index of suspicion for OSA, irrespective of their ISI scores, were referred for in-laboratory parallel PSG and PAT device testing. In addition, all 260 participants from part 2 had additional PAT reports generated using a 4% desaturation threshold to compare the performance of various threshold-dependent testing parameters such as PAT-based AHI (pAHI) and PAT-based respiratory disturbance index (pRDI).

Definitions and criteria used in PSG interpretation were based on the *International Classification of Sleep Disorders,* third edition and the AASM practice parameters.[6–9] Per the AASM scoring manual,[6] a nasal pressure transducer, an oronasal thermistor, and respiratory inductance plethysmography effort belts were used in all patients. Apnea was defined as a near-complete cessation of airflow (≥ 90% reduction from baseline), lasting for ≥ 10 seconds. Hypopnea was defined polysomnographically as an amplitude reduction in airflow or a respiratory effort of 30%–90% from baseline, lasting ≥ 10 seconds and associated with either a ≥ 3% oxygen desaturation or an arousal.

Signals recorded by WatchPat devices include wrist activity (actigraphy), PAT signal and pulse rate (PAT probe), pulse oximetry–based oxyhemoglobin saturation (SpO_2), and snoring (microphone). The WatchPat 200 devices define respiratory events by pulse oximetry desaturations (using 3% or 4% thresholds) and sympathetic discharges, the latter being defined by a PAT amplitude reduction and concomitant increases in heart rate (measured using a proprietary software algorithm). Further, the WatchPat software employs a proprietary automated algorithm for defining sleep and wake states based on movements and their patterns (sporadic or periodic) and for sleep stage differentiation (REM sleep vs non-REM sleep) based on the spectral and temporal components of the PAT signals.

To assess the diagnostic reliability of PAT devices in several special situations, we did not exclude patients with atrial fibrillation or congestive heart failure or those on alpha-adrenergic blockers. By default, and unless stated otherwise, pAHI and pRDI were computed using a 3% desaturation threshold. For the purpose of this study, we used the proprietary and validated automated WatchPat scoring and reporting—ie, we used no manual overscoring. The interpretation of all studies was blinded and completed by board-certified sleep physicians.

Descriptive analyses of the study variables were performed. Categorical variables were presented as frequencies or percentages. Continuous variables were described as means ± standard deviations (if normally distributed) or medians, 25th–75th interquartile ranges (IQR), and ranges (R; where relevant, if nonnormally distributed). Distribution normality fitting was evaluated using Shapiro-Wilk and Anderson-Darling tests. The Student *t* test and analysis of variance were used to compare mean values, and categorical variables were compared using the χ^2 (likelihood ratio) test. Tukey-Kramer honestly significant difference and Games-Howell (Tukey-Kramer honestly significant difference with Welch correction)[10] tests were used to compare means among pairs when the variances were similar or dissimilar, respectively. Agreement between results derived from PAT and PSG was determined by Pearson correlation coefficients and by the Bland-Altman method.[11] Misclassification of OSA severity (none, mild, moderate, or severe) was evaluated using contingency tables and percent agreement.

A priori sample size and power calculations were performed using various scenarios, with $P = .05–.001$, power = 0.80–0.90, a standard deviation of 12–16, and a minimally significant bias or difference between pAHI and AHI of 5 events/h or a misclassification rate > 10%. As such, the target number of valid studies (able to reach statistical significance, given the above assumptions) was set at 500.

Statistical significance was predefined as $P < .05$. Analyses were performed using JMPPro15 statistical software (SAS Institute, Cary, NC). Institutional research approvals were obtained to conduct the study (Emory University Institutional Review Board No. 00049576; Atlanta VA Research and Development Committee No. 0002). Preliminary analyses of the current study were presented in late-breaking abstract format during the SLEEP international meeting in 2019.[12]

RESULTS

The study included 500 consecutive valid tests that were evaluated after excluding 31 tests with a poor PAT and/or SpO_2 signal (defined as a substantial portion being unusable for scoring sleep, respiratory events, or pulse oximetry; **Figure 1**). Baseline characteristics of the study participants are shown in **Table 1.** Median (IQR; R) age was 52.5 (41.8–62.5; 24–92) years. Eighty percent of the participants were men; 26% were self-identified as white or Caucasian and 72% as black or African American. Tested participants were significantly symptomatic: 71% of them had complaints of excessive daytime sleepiness, ie, an Epworth Sleepiness Scale score ≥ 10. Approximately 72% of the participants had difficulty initiating or maintaining sleep, as illustrated by an ISI ≥ 8; among those, half

Figure 1—Study flow chart.
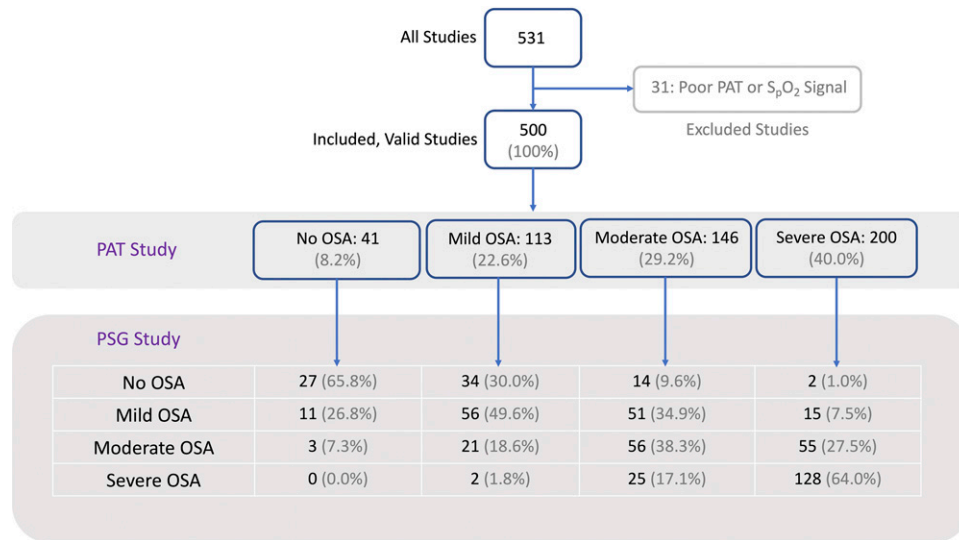
PAT = peripheral arterial tonometry, PSG = polysomnography, SpO$_2$ = pulse oximetry–based oxyhemoglobin saturation.

Table 1—Baseline characteristics of the study group.

Characteristic and Measurement	Part 1, n = 240	Part 2, n = 260	All, n = 500
Age (y)			
Median	51.8	52.7	52.5
IQR	41.4–61.8	42.4–62.8	41.8–62.5
Sex (%)			
Male	80	80	80
Female	20	20	20
Race or ethnicity (%)			
White or Caucasian	28	23	26
Black or African American	68	76	72
BMI (kg/m^2)			
Median	32.0	31.3	31.6
IQR	28.3–36.4	27.8–35.6	28.0–35.9
ESS			
Median	13	13	13
IQR	9–17	9–17	9–17
ISI			
Median	21	20	20
IQR	16–25	15–24	15–25
BQ (%)			
Positive	95	95	95
Negative	5	5	5

BMI = body mass index, BQ = Berlin Questionnaire, ESS = Epworth Sleepiness Scale, IQR = interquartile ratio, ISI = insomnia severity index.

had ISI-defined subthreshold insomnia, and half had moderate or severe insomnia. The Berlin Questionnaire was classified as "positive," indicating a high clinical suspicion for OSA in 95% of participants, during both periods of enrollment in the study (**Table 1**). The PSG-based diagnosis of OSA was present in 85% of participants, and OSA syndrome (OSA and an Epworth Sleepiness Scale score ≥ 10) was found in 70% of participants (**Table 2**). The median (IQR; R) of AHI and nadir SpO$_2$ were 18.4 (7.6–36.7; 0.4–145.6) events/h and 83% (76–88; 51–95), respectively; the central apnea index was 0.2 (0–0.8; 0–53).

Table 2—PSG test parameters.

Parameter and Measurement	Part 1, n = 240	Part 2, n = 260	All, n = 500
OSA (%)			
Present	85	85	85
Absent	15	15	15
OSA syndrome (%)			
Present	70	70	70
Absent	30	30	30
TST (min)			
Median	322	329	327
IQR	263–358	286–363	278–361
AHI			
Median	19.0	17.6	18.4
IQR	7.8–39.3	7.5–34.7	7.6–36.7
REM AHI			
Median	24.0	30.5	26.0
IQR	9.3–55.9	9.0–53.3	9.0–54.0
$ODI_{3\%}$			
Median	2.0	2.9	2.5
IQR	0.3–8.8	0.4–11.7	0.4–10.2
$ODI_{4\%}$			
Median	0.2	2.9	0.5
IQR	0.1–0.6	0.4–10.8	0.1–4.0
Hypoxic burden (% TST with SpO_2 < 90%, %)			
Median	2	2	2
IQR	0-9	0-10	0-10
Heart rate, min (bpm)			
Median	55	54	54
IQR	49–61	48–60	48–60
Heart rate, mean (bpm)			
Median	66	65	65
IQR	60–73	57–71	59–72
Heart rate, max (bpm)			
Median	86	86	86
IQR	79–94	79–93	79–93

bpm = beats per minute, IQR = interquartile ratio; ODI = oxygen desaturation index, $ODI_{3\%}$ = oxygen desaturation index using 3% desaturation threshold, $ODI_{4\%}$ = oxygen desaturation index using 4% desaturation threshold, PSG = polysomnography, SpO_2 = pulse oximetry–based oxyhemoglobin saturation, TST = total sleep time.

Based on the standard cutoffs of AHI 5, 15, and 30 events/h, approximately 27%, 27%, and 31% of participants had mild, moderate, and severe OSA, respectively. **Table 3** shows the PAT-based functional parameters and diagnostic classifications.

Thirty participants (6%) had a prior chart-adjudicated diagnosis of congestive heart failure: 14 (45%) had systolic dysfunction, with a median (IQR) left ventricular ejection fraction of 40% (30%–60%) on the last echocardiogram, and 17 (55%) had a diagnosis of diastolic dysfunction. The electrocardiogram on the PSG showed sinus rhythm in 98% of participants and atrial fibrillation was found in 2% of participants (8/9 with continuous arrhythmia), all in the moderate or severe

OSA categories. Approximately 3% and 6% of patients had a central apnea index > 10 and > 5, respectively. Periodic breathing was noted in 8 patients (1.6%).

In the entire study cohort, the mean residual $pAHI_{3\%}$-AHI or bias was 4.2 (95% confidence interval [CI], 2.8–5.5), with a median (IQR; R) of 3.7 (−3 to 12; −61 to 62; see **Figure 2**). In part 1 of the study, these differences were on average 5 (95% CI, 3–7), with a median (IQR; R) of 5 (−2 to 13; −52 to 62), and in part 2 the differences were 4 (95% CI, 2–5), with a median (IQR; R) of 3 (−3 to 11; −61 to 54). These differences were not explained on univariate and multivariate analyses by the Epworth Sleepiness Scale, the ISI, the Berlin Questionnaire, the

Table 3—PAT device test parameters.

Parameter and Measurement	Part 1, n = 240	Part 2, n = 260	All, n = 500
OSA (%)			
Present	93	91	92
Absent	7	9	8
OSA syndrome (%)			
Present	74	71	73
Absent	26	28	27
pTST (min)			
Median	343	349	348
IQR	305–369	313–380	309–378
$pAHI_{3\%}$			
Median	25.8	25.2	25.3
IQR	14.9–48.6	10.9–42.9	11.9–46.2
$pAHI_{4\%}$			
Median	-	13.7	13.7
IQR	-	3.6–29.6	3.6–29.6
$pRDI_{3\%}$			
Median	28.7	28.1	28.3
IQR	18.2–49.6	14.8–46.3	16.4–47.9
$pRDI_{4\%}$			
Median	-	17.4	17.4
IQR	-	9.6–31.6	9.6–31.6
$pODI_{4\%}$			
Median	11.3	10.8	10.9
IQR	5.6–25.7	3.2–23.6	3.9–25.1
Hypoxic burden (% TST with SpO_2 < 90%, %)			
Median	0.1	0.3	0.2
IQR	0–1.9	0–2.0	0–2
Pulse rate, min (bpm)			
Median	47	46	46
IQR	41–54	41–53	41–53
Pulse rate, mean (bpm)			
Median	68	66	67
IQR	62–75	60–72	61–74
Pulse rate, max (bpm)			
Median	101	100	100
IQR	95–111	93–109	94–111

bpm = beats per minute, IQR = interquartile ratio, $pAHI_{3\%}$ = peripheral arterial tonometry–based AHI using 3% desaturation threshold, $pAHI_{4\%}$ = peripheral arterial tonometry–based AHI using 4% desaturation threshold, PAT = peripheral arterial tonometry, $pODI_{4\%}$ = peripheral arterial tonometry–based oxygen desaturation index using 4% desaturation threshold, $pRDI_{3\%}$ = peripheral arterial tonometry–based respiratory distress index using 3% desaturation threshold, $pRDI_{4\%}$ = peripheral arterial tonometry–based respiratory distress index using 4% desaturation threshold, pTST = peripheral arterial tonometry–based total sleep time, SpO_2 = pulse oximetry–based oxyhemoglobin saturation, TST = total sleep time.

presence of insomnia or OSA syndrome as categorical variables (because these were potentially confounded by the testing triage algorithm during part 1), or any other baseline characteristics of the study participants. The only differentiating feature that explained this performance asymmetry was the use of different desaturation thresholds for pAHI. As such, in the 3% desaturation threshold-based group (**Figure 3**, panel A,

panel A', and panel A"), the PAT devices significantly overestimated the severity of OSA against PSG, whereas in the 4% group (**Figure 3**, panel B, panel B', and panel B"), the PAT devices tended to underestimate on average the severity of sleep-disordered breathing.

Because polysomnographic hypopneas were defined by a 30%–90% flow amplitude reduction with associated arousal or

Figure 2—Linear fit of pAHI vs AHI, Bland-Altman diagram of residual pAHI-AHI vs the average values, and shadowgram of residuals.
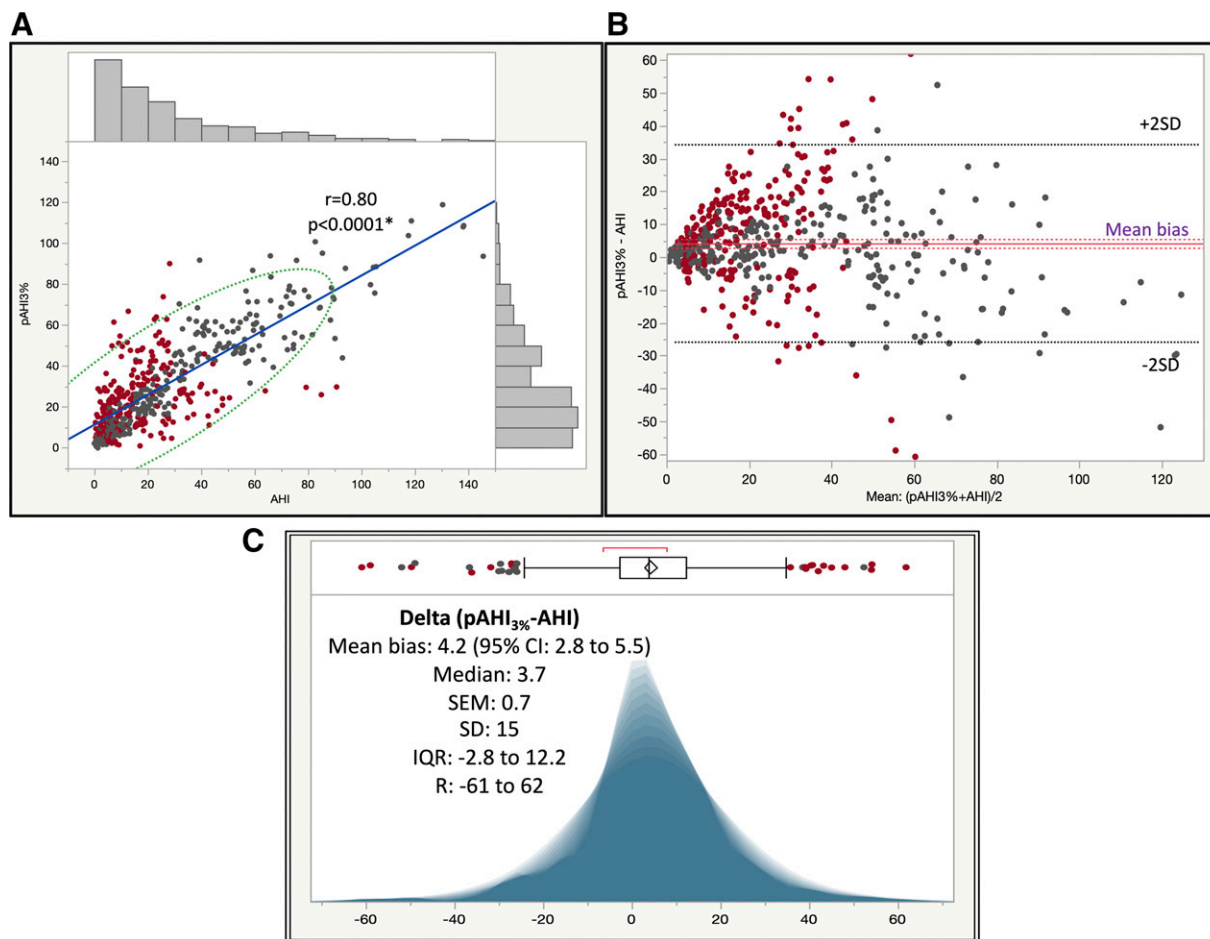
(A) Linear fit of pAHI vs AHI (*r* = .80, blue line), with side histograms and 95% CI normal ellipse (green). **(B)** Bland-Altman diagram of residual pAHI-AHI vs the average values, showing mean bias and 95% CI (purple) and ± 2 SD lines (black). **(C)** Shadowgram of residuals (pAHI$_{3\%}$-AHI). Above the graph is shown the outlier box plot, delineated by the 25th and 75th quartiles (IQR). Vertical line in the middle of the rectangle = median; diamond = mean and 95% CI; whiskers = 1.5 × IQR; red bracket: the shortest or the densest 50% of the distribution of observations. A shadowgram is a graphic representation of all representative histograms with different bin widths on the x axis (it overcomes distortions related to the bin width; as such, dominant features of a distribution are less transparent). Marker color code: dark gray = concordant; red = discordant diagnoses of no, mild, moderate, or severe OSA. CI = confidence interval, IQR = interquartile range, pAHI = peripheral arterial tonometry–based AHI, pAHI$_{3\%}$ = peripheral arterial tonometry–based AHI using 3% desaturation threshold, R = range, SD = standard deviation.

an SpO$_2$ drop of ≥ 3%, we also compared AHI with pRDI$_{3\%}$. In our analyses, we found that the mean bias pRDI$_{3\%}$-AHI was 7.3 (95% CI, 5.8–8.8) events/h (R^2 = .63; *P* < .0001)—ie, larger than the pAHI-AHI differences; the median (IQR, R) of the pRDI$_{3\%}$-AHI residuals was 7.3 (0.2 to 16.8, −59.7 to 61.9). **Figure S1** in the supplemental material shows shadowgrams of pAHI$_{3\%}$-AHI (blue) and pRDI$_{3\%}$-AHI (green, panel A), and pAHI$_{4\%}$-AHI (blue) and pRDI$_{4\%}$-AHI (green, panel B).

Figure 4 shows a mosaic plot of PSG diagnostic categories (absent, mild, moderate, and severe OSA) vs the same diagnoses made by PAT device on the *x* axis (*P* < .0001; degree of agreement kappa = 0.36). The positive and negative predictive values of the PAT test showing a diagnosis of no OSA were 66% and 89%, respectively, and the sensitivity and specificity were 35% and 97%, respectively. The overall concordance or accuracy rate—ie, the same diagnoses by both diagnostic

modalities—was 53.4% (**Figure 4**). In part 1 the concordance rate was 51.7%, and in part 2 it was only slightly better at 55%.

Conversely, a diagnosis of moderate or severe OSA by WatchPat (**Figure 5**) had positive and negative predictive values of 76% and 83%, respectively, and the sensitivity and specificity were 91% and 61%, respectively. We show in **Figure 5** a mosaic plot of diagnostic categories established by PSG (absent, mild, moderate, or severe OSA) against diagnoses made by the PAT devices on the *x* axis (*P* < .0001; degree of agreement kappa = 0.10). The overall relevant concordance or accuracy rate (ie, the diagnostic categories with distinct therapeutic implications) was 69.4%. In part 1 the relevant concordance rate was 69.6%, and in part 2 it was 69.2%.

Further, we found that the concordance rate was lower in the 3% desaturation threshold-defined group (52.6%; **Figure 6A**) than in the 4% desaturation group (56.1%; **Figure 6B**).

Figure 3—Shadowgrams and Bland-Altman diagrams of residuals, and linear correlations.
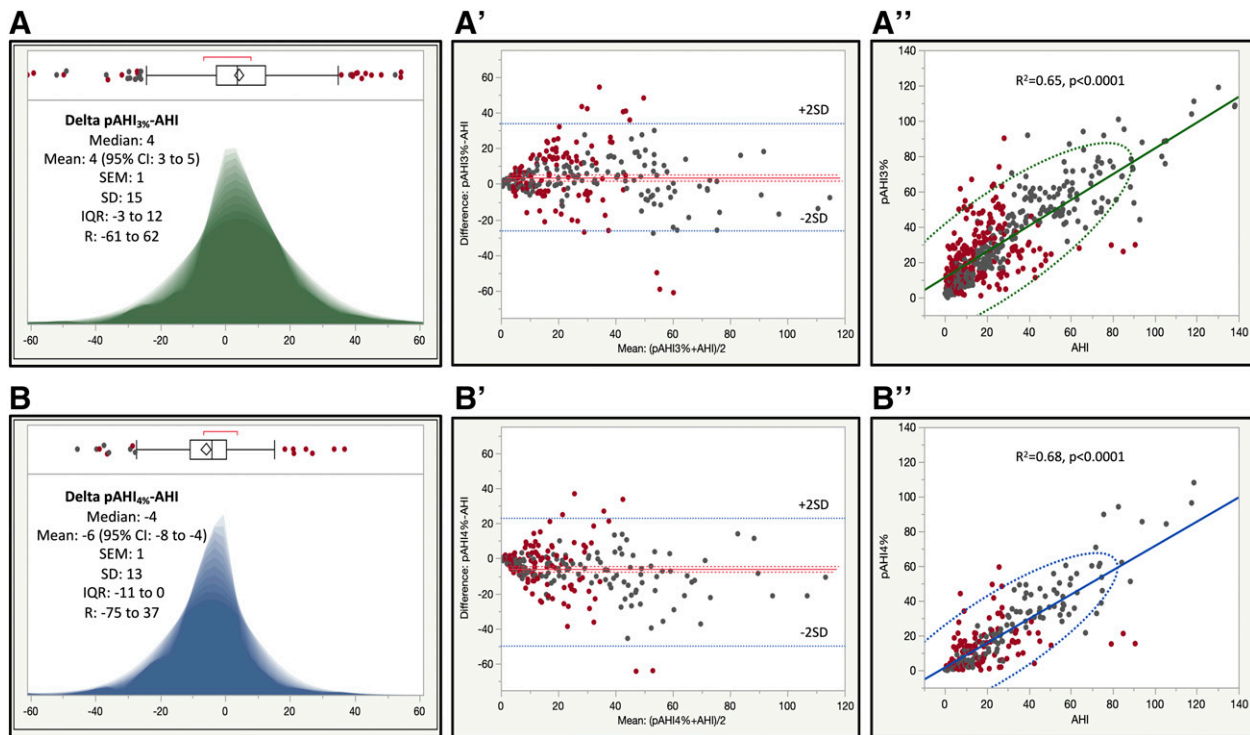
Left: shadowgrams of residual pAHI₃%-AHI **(panel A, green)** and pAHI₄%-AHI **(panel B, blue)**. Center: Bland-Altman diagrams of residual pAHI₃%-AHI **(panel A')** and pAHI₄%-AHI **(panel B')**. Right: Linear correlation and 95% CI ellipses for pAHI₃% vs AHI **(panel A")** and pAHI₄% vs AHI **(panel B")**. Marker color code: dark gray = concordant; red = discordant diagnoses of no, mild, moderate, or severe OSA. CI = confidence interval, IQR = interquartile range, pAHI = peripheral arterial tonometry–based AHI, pAHI₃% = peripheral arterial tonometry–based AHI using 3% desaturation threshold, pAHI₄% = peripheral arterial tonometry–based AHI using 4% desaturation threshold, R = range, SD = standard deviation.

Similarly, when we analyzed the relevant or therapeutically significant misclassifications (ie, between the no OSA, mild OSA, or moderate-severe OSA groups), the accuracy rate was 68.9% in the 3% oxyhemoglobin desaturation-based PAT report set (**Figure 7A**) and 71.0% in the 4% desaturation threshold-based set (**Figure 7B**).

Other significant comorbidities or current treatments in our study participants included asthma, chronic obstructive pulmonary disease, and alpha and beta-blocker medication use, which were found in 4.6%, 5.0%, 16.6%, and 17.3% of participants, respectively. Twenty-eight participants (5.6%) were on at least 1 narcotic medication at the time of the study. In our analyses, none of the associated comorbidities (including atrial fibrillation and congestive heart failure) or pharmacologic therapies influenced the performance of the PAT-based testing.

DISCUSSION

In this study on 500 consecutive patients evaluated with WatchPat 200 devices, we found that the PAT-based testing presented high rates of diagnostic misclassification (30%–50%) against concomitant gold standard PSG. The PAT-based diagnostic classifications were both under- and overestimations. We also found that when using a 3% desaturation threshold for

pAHI, WatchPat 200 tended to overestimate the prevalence and severity of OSA (on average by + 4 events/h), whereas the 4% threshold seemed to underestimate sleep-disordered breathing (on average by −6 respiratory events/h).

Several prior studies evaluated the reliability of WatchPat testing, showing strong correlations between AHI and pAHI, in the range of 0.85–0.90.[13–23] First, although the overall correlation coefficient and markers of central tendency such as mean bias and its 95% CI are important, some of the previous publications did not explore or present in detail the high dispersion of the residuals and their practical implications. Second, when evaluating a new device or method of testing, what is of significance to the clinician is not the combination of sensitivity and specificity, but rather the positive and negative predictive values. In other words, clinicians are more interested in the proportion of real positive and real negatives among patients who test positive or negative by the new method, test, or device, not in the proportion among those with gold standard–based disease, which in practice is not known. One should also remember that positive and negative predictive values are dependent on disease prevalence in the patient population applied, although this is typically not a significant issue in sleep clinic–based cohorts.

In our analyses, we showed that whereas the mean bias was relatively small, the dispersion of the residual pAHI-AHI was quite significant, which may be why the Pearson correlation
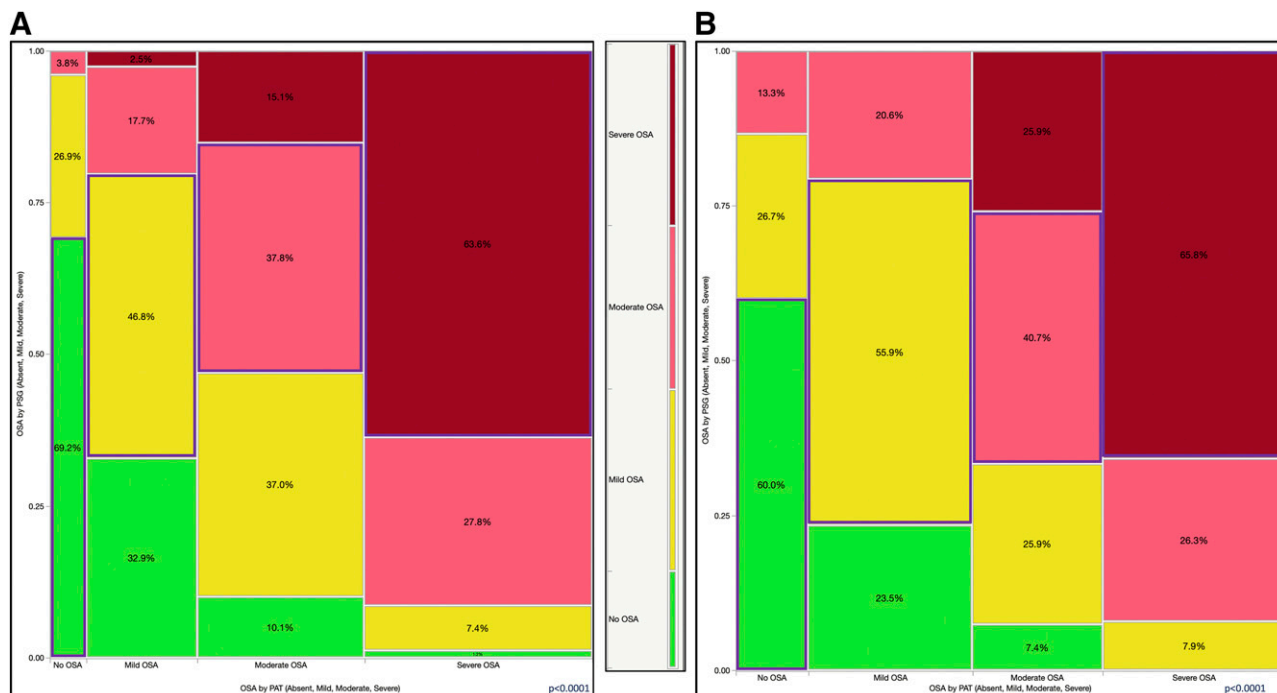
Figure 4—Mosaic plot showing a contingency analysis of OSA diagnosed by PSG (absent, mild, moderate, severe) vs the same diagnoses made by PAT device.
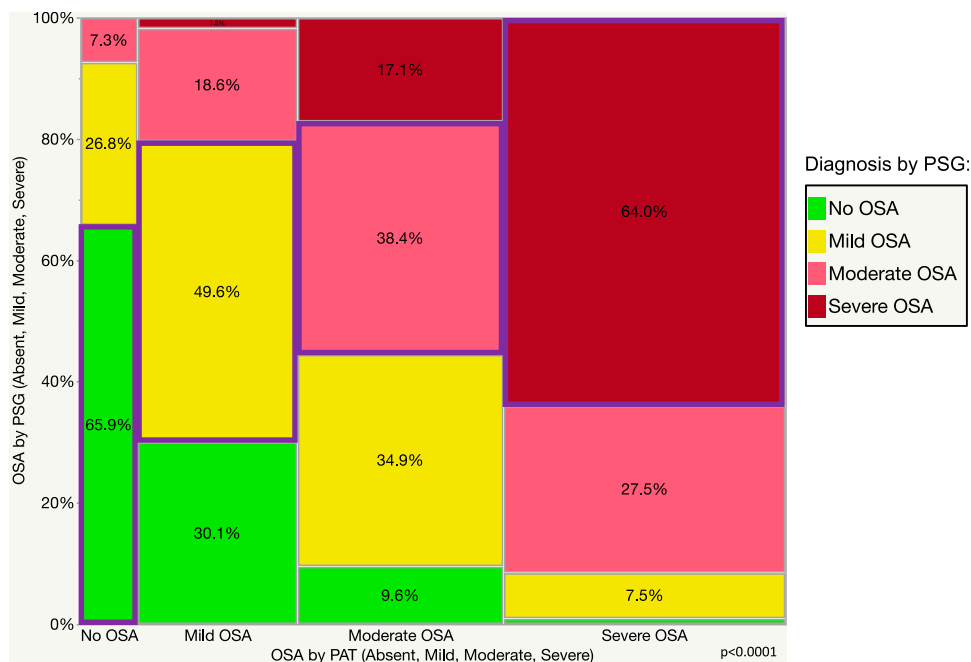
All categories were diagnosed by PSG, not only severe OSA. Overall accuracy or concordance rate was 53.4%. Color code: green (bottom) = no OSA; yellow (next to bottom) = mild OSA; pink (next to top) = moderate OSA; dark red (top) = severe OSA (by PSG); blue outlined rectangles = concordant categories. PAT = peripheral arterial tonometry, PSG = polysomnography.

**Figure 5**—Mosaic plot showing a contingency analysis of OSA diagnosed by PSG (absent, mild, moderate/severe) vs same diagnostic categories by PAT.
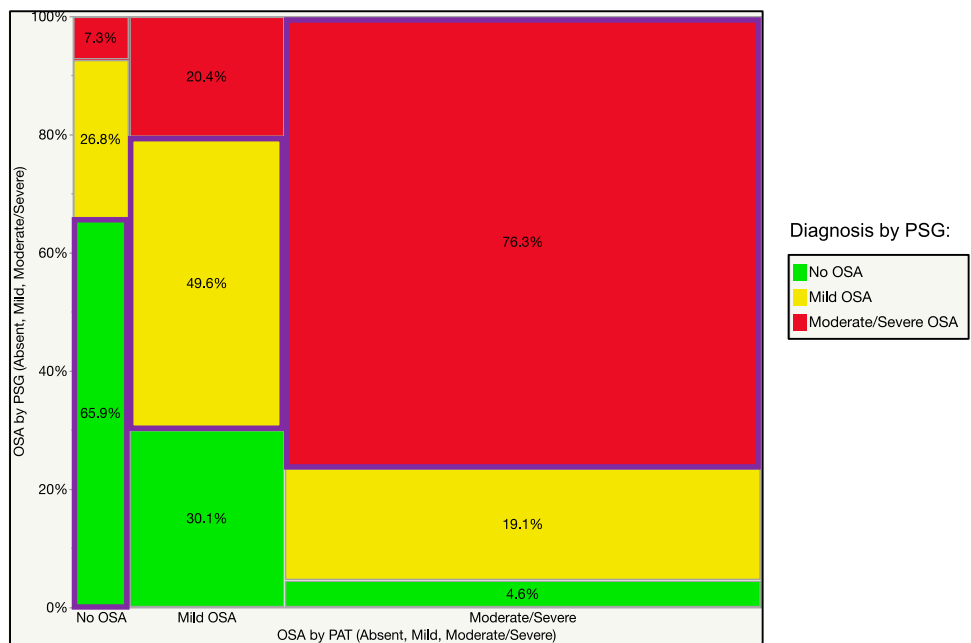
All categories were diagnosed by PSG, not only severe OSA. Relevant accuracy or concordance rate was 69.4%. Color code: green (bottom) = no OSA; yellow (middle) = mild OSA; bright red (top) = moderate/severe OSA (by PSG); blue outlined rectangles = concordant categories. PAT = peripheral arterial tonometry, PSG = polysomnography.

Figure 6—Mosaic plot showing a contingency analysis of OSA diagnosed by PSG (absent, mild, moderate, severe) vs the same diagnostic categories by PAT.

All categories were diagnosed by PSG, not only severe OSA. **(A)** The 3% desaturation threshold group. Accuracy or concordance rate was 52.6%. **(B)** The 4% desaturation threshold group. Accuracy or concordance rate was 56.1%. Color code: green (bottom) = no OSA, yellow (next to bottom) = mild OSA; pink (next to top) = moderate OSA; dark red (top) = severe OSA (by PSG); blue outlined rectangles = concordant categories. PAT = peripheral arterial tonometry, PSG = polysomnography.

Figure 7—Mosaic plot showing a contingency analysis of OSA diagnosed by PSG (absent, mild, moderate/severe) vs same diagnostic categories by PAT.
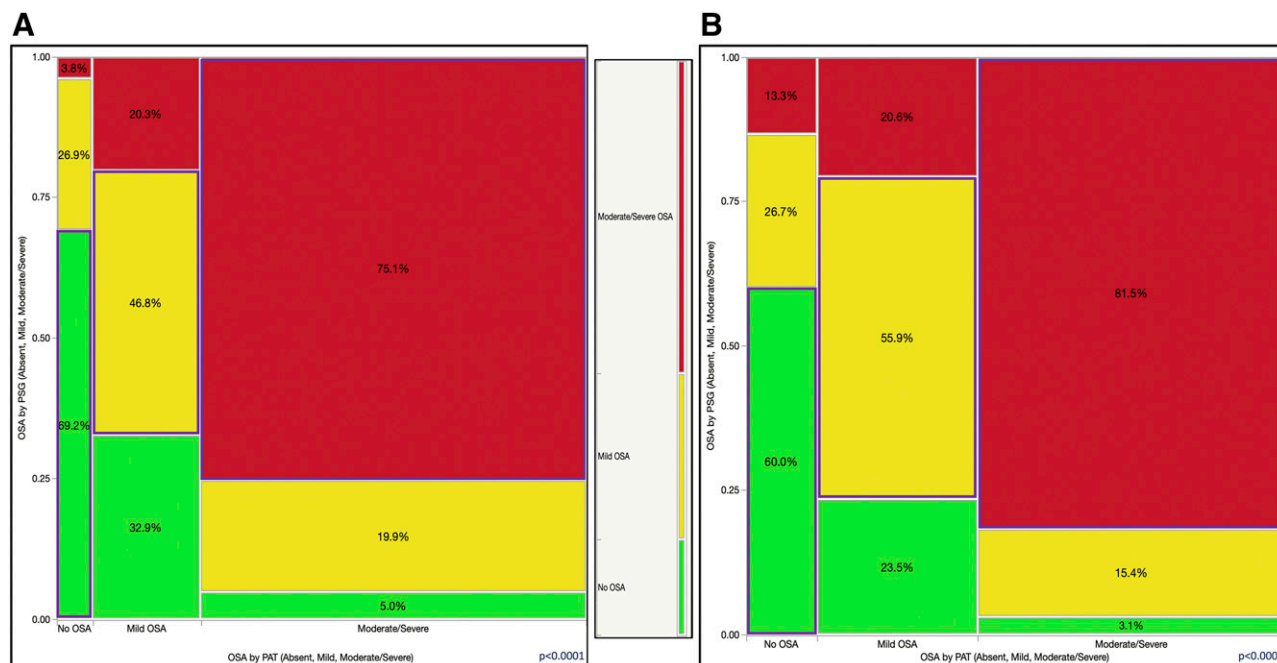
All categories were diagnosed by PSG, not only severe OSA. **(A)** The 3% desaturation threshold group. Accuracy or concordance rate was 68.9%. **(B)** The 4% desaturation threshold group. Accuracy or concordance rate was 71.0%. Color code: green (bottom) = no OSA; yellow (middle) = mild OSA; bright red (top) = moderate/severe OSA (by PSG); blue rectangles = concordant categories. PAT = peripheral arterial tonometry, PSG = polysomnography.

coefficient of 0.80 and the graph (**Figure 2A**) was misleading; however, the Bland-Altman diagram of the residuals (**Figure 2B**) and the histogram of the residuals (**Figure 2C**) tell another side of the story. To illustrate the point, we color-coded each marker **Figure 2A, Figure 2B,** and **Figure 2C** (red = discordant diagnoses, gray = concordant diagnoses of no OSA, mild OSA, or moderate or severe OSA). As seen in **Figure 2**, there are many large residuals that do not lead necessarily to reclassification or misclassification and small residuals that do change the discrete diagnostic category. This is important because some of these reclassifications pose significantly different therapeutic implications, and the average residual simply does not capture these implications. Further, the fact that the bias was actually larger when $pRDI_{3\%}$ was used instead of $pAHI_{3\%}$ in comparison with AHI (**Figure S1**) points toward the circumstance that an increase in heart rate concomitant with a reduction in PAT amplitude does not always equate to an arousal—perhaps an intrinsic limitation of the PAT-based algorithms that define pRDI. This discrepancy may also indicate pulse oximetry signal imprecisions, and indeed, the correlations between the oxygen desaturation index (ODI) and the PAT-based oxygen desaturation index (data not shown) had Pearson correlation coefficients < 0.81. At the same time, the polysomnographic definition of cortical arousal and its low interrater reliability in visual scoring may induce additional imprecision.

We found that among patients with PAT-diagnosed severe OSA, only 1.1% had no OSA and 7.5% had mild OSA by concomitant PSG (**Figure 4**), reassuring us that the negative predictive value of the test is high for severe OSA (91%). The positive predictive value of PAT for diagnosing moderate or severe OSA was 76%, and its negative predictive value was 83%. Similarly, as **Figure 5** shows, only 4.6% of patients with PAT-diagnosed moderate or severe OSA did not actually have OSA, and approximately 19.1% in fact had mild disease. Conversely, the negative predictive value of PAT for a diagnosis of OSA of any severity was only 66%, reinforcing the current practice that a negative or inconclusive PAT test should be followed by PSG in the sleep laboratory. What is new in our study is that the diagnostic accuracy was close to only 49.6% in those deemed to have mild OSA by PAT, whereas 30.1% did not have OSA at all and 20.4% had in fact moderate or severe OSA (**Figure 4** and **Figure 5**, middle columns). As such, the logical recommendation in this situation is that diagnoses of no OSA or mild OSA by PAT technology in patients with high pretest probability should be followed by gold standard PSG testing.

During part 1 of the study, we directed participants with significant insomnia complaints preferentially toward sleep laboratory testing and referred those without any difficulty initiating or maintaining sleep to undergo HSAT with various available portable monitors. These procedures could have potentially skewed the patient population included in part 1 toward a patient category in which the WatchPat 200 devices were not officially recommended (as per AASM guidelines) and could possibly underperform. Such a possibility was the main reason to follow this phase of the investigation with the part 2 enrollment, in which no referral bias was present, as all participants with high clinical suspicion for OSA were referred to undergo PSG (with concurrent PAT testing). Nevertheless,

we found that the triage algorithm employed in part 1 produced similar results to those in the part 2 enrollment period.

The use of a more conservative desaturation threshold (4%; **Figure 6B**) led to a higher specificity of the PAT study overall. As such, only 25.9% of participants with moderate OSA as diagnosed by PAT had mild disease as diagnosed by PSG, and 7.4% had no OSA. Among those with severe OSA, only 7.9% had mild sleep-disordered breathing; none of these participants were deemed to have no OSA (**Figure 6B**). It is also important to remember that hypopneas on PSG were defined based on a 30%–90% flow amplitude reduction and a 3% desaturation or an arousal. Similarly, in the 4% threshold PAT-based reports, when we combined PAT diagnoses of moderate and severe OSA into the same category (with similar therapeutic implication), only 18.5% of the results were discordant (15.4% mild OSA and 3.1% no OSA as diagnosed by PSG; **Figure 7B**). As such, for a diagnosis of moderate or severe OSA, the 4% PAT report had 56% sensitivity and 94% specificity and 92% positive and 64% negative predictive values, respectively. The 3% PAT report had 65% sensitivity and 91% specificity and 91% positive and 64% negative predictive values, respectively, for moderate and severe disease combined. Even in the 4% threshold–based group, it is important to recognize that PAT-defined mild OSA included 20.6% of patients diagnosed with moderate or severe disease and approximately 23.5% of patients diagnosed as having no OSA by PSG (**Figure 7B**).

As such, it is our recommendation that the 4% desaturation threshold be used in PAT studies to improve the specificity and negative predictive value of the tests showing moderate or severe OSA, and negative and mild OSA results should prompt clinicians to re-evaluate with a follow-up PSG.

Other groups have recognized the potential limitations of PAT-based sleep testing technology (such as PAT amplitude changes and heart rate changes independent of each other; arousals potentially leading to PAT amplitude changes, heart rate changes, or both; and the potential need to use different SpO_2 thresholds in REM sleep vs non-REM sleep), possible pulse oximeter artifacts, and proposed corrective strategies for use in manual scoring.[24]

Discrepancies between PSG- and PAT-based diagnoses (including the extreme bidirectional reclassifications between the no OSA and severe OSA diagnoses) may be explained by several synergistic effects. First, the misestimation of the total sleep time by the WatchPat 200 proprietary algorithm is responsible for some error. As such, the mean difference ± standard deviation between the PAT- and PSG-based total sleep time was 26 (95% CI, 20–31) ± 63 minutes and the REM sleep difference between PAT and PSG was 23 (95% CI, 20–26) ± 32 minutes. A second type of error that we observed was induced by pulse oximetry artifacts. As shown in **Table 2** and **Table 3**, PAT-based ODI significantly overestimated ODI (on average): The mean difference ± SD between the PAT-based $ODI_{4\%}$ and $ODI_{4\%}$ was 12 ± 17 desaturations per hour. Last, the PAT-based specific definition of respiratory events is based on opposite-direction changes in PAT signal amplitude and pulse rate, thus allowing pulse detection artifacts to induce further errors. This measurement imprecision was in fact the lowest: the difference ± standard deviation between the mean pulse as measured by

WatchPat and the mean heart rate as measured by PSG was 1 ± 3 (R, −19 to 31) beats per minute.

For the purposes of this study, we did not use manual scoring, because our intent was to (1) recognize the extent of diagnostic uncertainty and misclassification by using traditional WatchPat testing and reporting and (2) evaluate alternative pathways to improve the precision of the test without resorting to manual scoring.

Although the AASM clinical guidelines[3] clearly recommend sleep testing interpretation to be based on manual scoring and review, there is a clear, concurrent, opposite, and likely equally valid trend—ie, that of developing new artificial intelligence–based approaches that will reduce the labor intensity and error propensity of standard testing and interpretation.[25]

Our study has several strengths. First, the study is a large point-of-care investigation that systematically evaluated, without significant missingness and with blinded interpretations, concurrent PSG- and PAT-based testing (ie, no night-to-night variability bias) in a sleep clinic-based patient population with a high clinical probability of OSA and a high prevalence of the condition. Second, we also evaluated a systematic triage-based approach (part 1) vs a consecutive "all referrals" type of approach (part 2). Third, we comparatively assessed the impact of 3% vs 4% desaturation thresholds for OSA diagnosis and severity stratification. Potential weaknesses of our investigation are related to the single-center nature of the study, which was conducted on a population of mostly male, African American or Black military veterans with significant comorbidity burden, including insomnia (all potentially limiting the generalizability of the findings), the point-of-care design (no randomization) and the lack of manual scoring, review and adjudication of all the respiratory events and/or sleep stages (as reported by the WatchPat software). In addition, although the prevalence of heart failure in our study was higher than the one in the general population, it was likely lower than the actual disease frequency in our veterans. The use of patient care overflow mechanisms designed to improve access for veterans may explain, at least in part, the latter observation, as patients with cardiovascular comorbidities may want to have their sleep studies expedited. While patients occasionally do opt for "outside" VA sleep studies, in our center and during this period of time we found that less than 10% used the overflow pathways. The majority of them were HSATs and were performed within similar timeframes.

CONCLUSIONS

In this large point-of-care, clinic-based study on patients with various sleep complaints and a high pretest probability of OSA who were evaluated with WatchPat 200 devices and concomitant PSG, we found that the PAT-based testing presented high rates of diagnostic misclassification of sleep-disordered breathing presence or severity. The PAT-based diagnostic misclassifications were both under- and overestimations. This result suggests that in patients with a high probability for the condition and with no or mild OSA as diagnosed by PAT, a repeat in-laboratory PSG is warranted. Even when moderate or severe OSA is diagnosed, the

possibility of overestimating or underestimating the disease severity is significant, a fact that could adversely influence the therapeutic recommendations in significant proportions of patients. We also found that when using a 3% desaturation threshold for pAHI, WatchPat tended to overestimate the prevalence and severity of OSA (on average by 4 events/h) and the 4% threshold seemed to underestimate it (on average by −6 respiratory events/h).

ABBREVIATIONS

AASM, American Academy of Sleep Medicine
BQ, Berlin Questionnaire
AHI, apnea-hypopnea index
BMI, body mass index
CI, confidence interval
ESS, Epworth Sleepiness Scale
HSAT, home sleep apnea test
IQR, interquartile range
ISI, insomnia severity index
ODI, oxygen desaturation index
PAT, peripheral arterial tonometry
pAHI, peripheral arterial tonometry–based AHI
$pAHI_{3\%}$, peripheral arterial tonometry–based AHI using 3% desaturation threshold
$pAHI_{4\%}$, peripheral arterial tonometry-based AHI using 4% desaturation threshold
$pODI_{4\%}$, peripheral arterial tonometry–based oxygen desaturation index using 4% desaturation threshold
pRDI, peripheral arterial tonometry–based respiratory distress index
PSG, polysomnography
pTST, peripheral arterial tonometry–based total sleep time
R, range
SpO_2, pulse oximetry–based oxyhemoglobin saturation
TST, total sleep time

REFERENCES

1. Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med.* 2019;7(8):687–698.

2. Kundel V, Shah N. Impact of portable sleep testing. *Sleep Med Clin.* 2017;12(1): 137–147.

3. Collop NA, Anderson WM, Boehlecke B, et al. 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.

4. Collop NA, Tracy SL, Kapur V, Mehra R, Kuhlmann D, Fleishman SA, et al. Obstructive sleep apnea devices for out-of-center (OOC) testing: technology evaluation. *J Clin Sleep Med.* 2011;7(5):531–548.

5. Ioachimescu OC, Collop NA. Sleep-disordered breathing. *Neurol Clin.* 2012; 30(4):1095–1136.

6. 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.

7. American Academy of Sleep Medicine. *International Classification of Sleep Disorders.* 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014.

8. Morgenthaler TI, Aurora RN, Brown T, 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*. 2008;31(1):141–147.

9. Kushida CA, Littner MR, Morgenthaler T, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep*. 2005;28(4):499–523.

10. Games PA, Howell JF. Pairwise multiple comparison procedures with unequal n's and/or variances: a Monte Carlo study. *J Educ Stat*. 1976;1(2):113–125.

11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.

12. Zahiruddin F, Chada A, Allam JS, et al. Point of care validation study of WatchPat testing for the diagnosis of obstructive sleep apnea in a cohort of 250 sleep laboratory patients. Abstract [Paper] [Poster] presented at: SLEEP meeting; June 8–12, 2019; San Antonio, TX.

13. Weimin L, Rongguang W, Dongyan H, Xiaoli L, Wei J, Shiming Y. Assessment of a portable monitoring device WatchPAT 200 in the diagnosis of obstructive sleep apnea. *Eur Arch Otorhinolaryngol*. 2013;270(12):3099–3105.

14. Pang KP, Gourin CG, Terris DJ. A comparison of polysomnography and the WatchPAT in the diagnosis of obstructive sleep apnea. *Otolaryngol Head Neck Surg*. 2007;137(4):665–668.

15. Jen R, Orr JE, Li Y, DeYoung P, Smales E, Malhotra A, et al. Accuracy of WatchPAT for the diagnosis of obstructive sleep apnea in patients with chronic obstructive pulmonary disease. *COPD*. 2020;17(1):34–39.

16. Gan YJ, Lim L, Chong YK. Validation study of WatchPat 200 for diagnosis of OSA in an Asian cohort. *Eur Arch Otorhinolaryngol*. 2017;274(3):1741–1745.

17. Hedner J, White DP, Malhotra A, et al. Sleep staging based on autonomic signals: a multi-center validation study. *J Clin Sleep Med*. 2011;7(3):301–306.

18. White DP, Gibb TJ, Wall JM, Westbrook PR. Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing in the home. *Sleep*. 1995;18(2):115–126.

19. Ayas NT, Pittman S, MacDonald M, White DP. Assessment of a wrist-worn device in the detection of obstructive sleep apnea. *Sleep Med*. 2003;4(5):435–442.

20. Zou D, Grote L, Peker Y, Lindblad U, Hedner J. Validation of a portable monitoring device for sleep apnea diagnosis in a population based cohort using synchronized home polysomnography. *Sleep*. 2006;29(3):367–374.

21. O'Brien LM, Bullough AS, Shelgikar AV, Chames MC, Armitage R, Chervin RD. Validation of Watch-PAT-200 against polysomnography during pregnancy. *J Clin Sleep Med*. 2012;8(3):287–294.

22. Park CY, Hong JH, Lee JH, et al. Clinical usefulness of Watch-PAT for assessing the surgical results of obstructive sleep apnea syndrome. *J Clin Sleep Med*. 2014; 10(1):43–47.

23. Yalamanchali S, Farajian V, Hamilton C, Pott TR, Samuelson CG, Friedman M. Diagnosis of obstructive sleep apnea by peripheral arterial tonometry: meta-analysis. *JAMA Otolaryngol Head Neck Surg*. 2013;139(12):1343–1350.

24. Zhang Z, Sowho M, Otvos T, et al. A comparison of automated and manual sleep staging and respiratory event recognition in a portable sleep diagnostic device with in-lab sleep study. *J Clin Sleep Med*. 2020;16(4):563–573.

25. Goldstein CA, Berry RB, Kent DT, et al. Artificial intelligence in sleep medicine: an American Academy of Sleep Medicine position statement. *J Clin Sleep Med*. 2020;16(4):605–607.

ACKNOWLEDGMENTS

SUBMISSION & CORRESPONDENCE INFORMATION

DISCLOSURE STATEMENT

All authors have seen and approved the manuscript. Work for this study was performed at the Atlanta Veteran Affairs Sleep Medicine Center. The authors report no conflicts of interest.