Sleep and Breathing (2019) 23:65–75
https://doi.org/10.1007/s11325-018-1658-y

SLEEP BREATHING PHYSIOLOGY AND DISORDERS • ORIGINAL ARTICLE

# The discriminative power of STOP-Bang as a screening tool for suspected obstructive sleep apnea in clinically referred patients: considering gender differences

Jin Mou[1] · Bethann M. Pflugeisen[1] · Brian A. Crick[2] · Paul J. Amoroso[1] · Kirk T. Harmon[3] · Stephen F. Tarnoczy[4] · S. Shirley Ho[4] · Kimberly A. Mebust[4]

Received: 2 November 2017 / Revised: 7 February 2018 / Accepted: 4 April 2018 / Published online: 24 April 2018


## Abstract

**Purpose** Obstructive sleep apnea (OSA) is the most commonly seen clinical sleep disorder. STOP-Bang, a widely used screening tool, yields a composite score based on eight dichotomized items including male gender. This study was designed to validate STOP-Bang among clinically referred patients and tested alternative scoring designs on tool performance, with a focus on gender differences in OSA.

**Method** STOP-Bang was administered to 403 female and 532 male subjects, followed by comprehensive sleep evaluation that included measurement of apnea-hypopnea indexes. Gender differences in STOP-Bang scores, OSA diagnosis, and severities were explored, and gender-specific alternative score cutoffs evaluated. Optimal operating points (OOP) were tested for female body mass index (BMI) and male neck circumference to inform STOP-Bang threshold refinement. Receiver operating characteristic curves were used to compare conventional and modified STOP-Bang.

**Results** STOP-Bang performance by gender showed extremely low specificity in males at the recommended cutoff of ≥3. Better utility was presented at a cutoff of 4 or 5 among clinically referred patients irrespective of gender differences. Screening performance was improved by modifying BMI and/or neck circumference thresholds using gender-triaged OOP estimation. Three gender-based model revisions outperformed conventional STOP-Bang.

**Conclusion** Our study suggests that gender-specific consideration needs to be incorporated into the application of STOP-Bang in a clinically referred patient population with a higher risk of OSA. Alternative scoring systems may improve predictive performance of STOP-Bang.

**Keywords** Obstructive sleep apnea · STOP-Bang · Screening · Gender disparity · Sleep disorder

## Introduction

Sleep disorders are common health conditions that affect both men and women. Obstructive sleep apnea (OSA) is the most commonly seen sleep disorder and is associated with many chronic health conditions including insomnia [1], hypertension [2], cardiovascular disease [3], cerebrovascular disease [4], myocardial infarction [5], atrial fibrillation [6], type 2 diabetes [7], depression [8], anxiety [9], and cognitive impairment [10]. Evidence has shown that OSA increases all-cause mortality [4]. Twenty-five percent of the US adult population is thought to be at high-risk for OSA, and the estimated OSA prevalence is 20% [11]. Due to the negative mental and physical effects of low-quality sleep, patients' health-related quality of life is frequently compromised [10]. Clinically, OSA is defined by the occurrence of excessive daytime sleepiness, loud snoring, witnessed breathing interruptions during sleep, or awakenings due to gasping or choking, with at least five obstructive respiratory events per hour of sleep [12].

✉ Jin Mou
jin.mou@multicare.org

[1] MultiCare Institute for Research & Innovation, MultiCare Health System, Ste 402, 314 MLK Jr. Way, Tacoma, WA 98405, USA

[2] Pulse Heart Institute, MultiCare Health System, Tacoma, WA 98405, USA

[3] MultiCare Centers of Occupational Medicine, Fife, WA 98424, USA

[4] MultiCare Sleep Medicine Center, MultiCare Neuroscience Center of Washington, Tacoma, WA 98405, USA

OSA severity must be established in order to allow for appropriate treatment, although a substantially large percentage of the afflicted stay undiagnosed and untreated [13]. OSA diagnosis can be established based on a comprehensive sleep evaluation (CSE) in conjunction with diagnostic testing including either laboratory-based polysomnography (PSG)/respiratory polygraphy (RP) or home testing with portable monitors (PM), guided by the frequency of obstructive events, reported as an apnea-hypopnea index (AHI), respiratory disturbance index (RDI) [14] (for in-lab PSG based test), or respiratory event index (REI) (for home sleep testing).

The American Academy of Sleep Medicine (AASM) recommends that OSA screening should be incorporated into routine health evaluations [15]. Screening methodology in this field has been developed quickly and includes questionnaires and clinical prediction models [16]. The Epworth Sleepiness Scale (ESS) [17], the Berlin Questionnaire (BQ), the American Society of Anesthesiologists' Checklist [18], and the STOP-Bang questionnaire [19] are examples of screening tools with varied applicability. STOP-Bang combines subjective self-reported items with objective demographic and anthropometric variables [20]. It is easy to use and has been validated across different sub-populations [21–23]. However, concerns have been raised about its specificity and the use of the conventional cutoff of ≥3 [24]. This may relate to the fact that STOP-Bang was originally created by anesthesiologists as a clinical instrument to rule out OSA in a lower-to-average risk population, without necessitating confirmatory PSG. STOP-Bang has a higher sensitivity to predict moderate-to-severe and severe OSA than mild OSA in population-based studies [25]. This might be attributable to the gender imbalance associated with under-diagnosis and OSA prevalence, especially if there are, in fact, a higher percentage of females with mild OSA. The fact that OSA is much more likely to be underdiagnosed in women raises the critical issue of clinically significant gender differences in OSA disease manifestation [26, 27]. Gender differences in sleep become apparent as early as the onset of puberty. Menstrual cycles, pregnancy, menopause, and hormone replacement therapy [28] can all alter sleep and breathing. Gender-related differences in prevalence, pathophysiology, and clinical presentation are also observed in sleep disorders. OSA gender disparities need to be considered in screening beyond a "yes-or-no" distinction, particularly in the scoring and result interpretation. The item dichotomization in STOP-Bang may have oversimplified the gender-specific nuances of OSA development, diminishing its total discriminative power. Currently available prediction models for OSA are not gender specific and use variables derived from initial studies of predominantly male patients. Also, no gender-specific prediction models exist for clinically referred high-risk patients. Improvements to the tool have been proposed, including alternative scoring systems [29], weighted models [30], the STOP-Bang equivalent model [31], and a two-step approach [32]. However, none of these proposed modifications consider gender as a confounding parameter.

The purpose of this study is to validate the application of STOP-Bang in a clinically referred high-risk patient population, and to evaluate alternative STOP-Bang scoring systems, with a focus on gender differences in OSA. We aim to test gender-specific models for total score cutoffs, evaluate BMI and neck circumference as the most critical independent parameters in predicting OSA across gender, and redefine the scoring system based on gender-specific attributes. The study will compare the discriminative power in OSA prediction between conventional STOP-Bang, and modified models.

## Methods

### STOP-Bang questionnaire

The STOP-Bang was developed in Canada by Chung et al. [24]. The two sections of this tool are STOP (***S***noring, ***T***iredness, ***O***bserved apnea, and high blood ***P***ressure) and Bang (***B***ody mass index over 35 kg/m$^2$, ***A***ge over 50 years, ***N***eck circumference over 40 cm/15.75 in., and male ***G***ender). Each affirmative response/characteristic scores one point. Negative responses/characteristics are scored as zero. The item scores are summed. The conventional recommendation is for any patient with a score ≥ 3 to be referred for OSA testing.

### Subjects

All patients were referred by non-sleep-medicine providers to the MultiCare Sleep Medicine Centers for comprehensive sleep disordered breathing evaluation during 2015–2016. This study presented data from adult patients 18 years or older at the time of referral.

### Sleep study

CSE is established according to the AASM Clinical Practice Guidelines on adult OSA diagnosis [23, 33, 34], and supervised by a board-certified sleep medicine practitioner. In-lab/center PSG and home sleep testing (HST) PM approaches were offered to patients, with recommendation prioritized for in-lab PSG, particularly for patients who had chronic comorbidities [23].

Patients undergoing in-center PSG were evaluated with recommended recording parameters outlined by the AASM. During the study, the registered polysomnographic technologist (RPSGT) attached sensors and electrodes to record the various parameters: electroencephalogram (EEG), electro-oculogram (EOG), submental electromyogram (EMG), limb muscle movements, lead II electrocardiogram

(ECG), finger pulse oximetry for oxygen saturation (SpO2), snore sensor, nasal-oral thermistor, nasal pressure transducer, and thoraco-abdominal plethysmography for respiration (RIP). The minimal digital PSG recording duration was 6 h, which was displayed in 30-s epochs displaying 22 channels of data on a computer screen. During the night, the PSG was staged and partially scored to calculate AHI criteria for split night evaluation after 2–3 h of sleep, if indicated. The AHI is the number of apneas or hypopneas recorded during the study per hour of sleep. It is generally expressed as the number of events per hour.

Home sleep studies were conducted using one of two types of home test devices: The Philips Respironics Alice PDx or the Itamar Medical WatchPAT. An RPSGT instructed each patient on the proper use of the prescribed device. The Alice PDx measures oral-nasal airflow and pressure via cannula, respiratory effort via abdominal and chest RIP belts, and arterial oxygen saturation level via a pulse oximeter probe to measure SpO2 and pulse rate. It also detects body position. Home-based Alice PDx shared the same AASM hypopnea criteria with PSG and showed a high level of diagnostic agreement with a simultaneous PSG [35, 36]. The Itamar WatchPAT detects the peripheral artery tone (PAT) by recording finger arterial volume changes, SpO2 and pulse rate. Sleep apnea and sleep disordered breathing are associated with an increase in heart rate, blood pressure, and sympathetic activation. WatchPAT works by correlating patient's low oxygen (oxygen desaturation) with patient's sympathetic tone (high, being a stressful obstructive event, and low, being a central apnea) and whether the patient is sleeping or awake (actigraph). PAT detects the change in volume in peripheral arteries that results from pauses in blood flow. The PAT signal is used as a proxy for respiratory disturbances and is automatically analyzed utilizing a clinically validated algorithm that can generate sleep disordered breathing parameters including AHI and/or RDI, based on specific signal patterns. PAT technology has been evaluated to be an accurate and clinically effective PM method for OSA diagnosis [37] and has been recently re-evaluated and addressed by AASM, considered to be "technically adequate" to diagnose OSA if "used in combination with oximetry and actigraphy" [23]. Throughout the night, data was recorded and stored on a removable storage card, before the device was returned and data was downloaded for viewing by a technologist and confirmed by a physician. Manual scoring/editing of the WatchPAT signals is possible and can be easily performed [38]. Both types of HST used 3% oxygen desaturation as the threshold for event definition.

REI is defined as "the total number of respiratory events scored × 60 divided by monitoring time (MT)".

We adopted the AASM criteria and scored a respiratory event as a hypopnea if ALL of the following criteria are met: a. The peak signal excursions drop by ≥ 30% of pre-event baseline using nasal pressure (diagnostic study), PAP device flow (titration study), or an alternative hypopnea sensor (diagnostic study); b. The duration of the ≥ 30% drop in signal excursion is ≥ 10 s; and c. There is a ≥ 3% oxygen desaturation from pre-event baseline and/or the event is associated with an arousal.

After the results of the study were completed, the patients were scheduled to meet with their provider for follow-up assessment and further testing or treatment if necessary. For highly susceptible and symptomatic patients with negative HST results, in-lab PSG may be recommended. For patients who went through both in-lab PSG and HST, only PSG-calculated AHI was recorded.

## OSA diagnosis and severity of OSA

A physician diplomat of the American Board of Sleep Medicine (ABSM), interpreted the PSG study or HST, and generated or dictated an Electronic Medical Record Report. A copy of the report was sent to the providers of record. Before an OSA diagnosis is finally made, a multidisciplinary clinical team, led by a board-certified sleep medicine specialist comprehensively evaluates the patients' clinical information and test results. Once an OSA diagnosis is confirmed, it is further characterized as mild OSA (AHI/REI ≥ 5–14), moderate OSA (AHI/REI ≥ 15–29), or severe OSA (AHI/REI ≥ 30). All the RPSGT and physicians participated in the AASM Inter-scorer Reliability Program (IRP) to ensure scoring quality.

## Data collection and statistics

The height, weight, and neck circumference were taken from office visits. The neck circumference was measured utilizing a retractable metal measuring tape. The BMI was calculated using weight in pounds × 703/height in inches squared.

In addition to descriptive statistics, combined responsiveness of the dichotomized BMI and neck circumference based on conventional STOP-Bang item cutoffs were mapped to OSA levels. Patient characteristics of the false negative cases were examined. Receiver operating characteristic (ROC) analysis was used to generate area under the curve (AUC) to measure tool performance. Gender-specific optimal operating points (OOP) were calculated for categories of OSA severity that were statistically equivalent to or better than conventional STOP-Bang. OOP was defined as the Youden Index ($J$), the maximum vertical distance between the ROC curve, and the diagonal line ($J$ = maximum (sensitivity + specificity − 1)) [39–41]. Modifications using the new thresholds for BMI and neck circumference were made in alternative systems and AUC were recalculated to compare with the conventional STOP-Bang utility, as well as continuous BMI or neck circumference. Statistical analyses were performed in R (Vienna, Austria) and Stata 14.1 (College Station, TX).

**Table 1** Age and neck circumference by BMI and gender

| Gender | BMI < 35 | BMI ≥ 35 | Estimate | 95% CI | *p* |
|---|---|---|---|---|---|
| Female (*n* = 403) | 204 (50.6%) | 199 (49.4%) | – | – | – |
| Age [mean (SD)] | 55.1 (15.6) | 48.8 (14.6) | 6.3 | 3.4, 9.3 | < .001 |
| Neck circumference [mean (SD)] | 14.4 (1.2) | 16.1 (1.4) | − 1.7 | − 2.0, − 1.5 | < .001 |
| % Age over 50 years | 64.7 | 50.3 | 14.4 | 4.9, 24.0 | .003 |
| % Neck size over 15.75 in. | 13.2 | 62.8 | − 49.6 | − 57.7, − 41.4 | < .001 |
| Male (*n* = 532) | 312 (58.6%) | 220 (41.4%) | – | – | – |
| Age [mean (SD)] | 51.5 (15.8) | 48.9 (15.0) | 2.6 | − .04, 5.3 | .05 |
| Neck circumference [mean (SD)] | 16.6 (1.2) | 18.6 (1.5) | − 2 | − 2.3, − 1.8 | < .001 |
| % age over 50 years | 51.0 | 45.5 | 5.5 | − 3.1, 14.1 | .21 |
| % neck size over 15.75 in. | 80.4 | 99.5 | − 19.1 | − 23.6, − 14.6 | < .001 |

This study was approved as a quality improvement initiative by the MultiCare Institutional Review Board and thus considered exempt from full board review.

## Results

### Patient characteristics

Of the 935 patients in the study, 43.1% were female. Ages ranged from 18 to 93 years old, with more women above 50 years old than men (57.6 vs. 48.7%, $p = .007$). Mean BMI was 35.7 ± 9.5 with small gender differences ($F$ 36.8 ± 10.4 vs. $M$ 34.8 ± 8.6, $p = .001$). There were more females with BMI > 35 than males ($p = .02$) (Table 1). Using the World Health Organization (WHO) standards [42], there were 22.1 and 69.5% of women who fell into the overweight (BMI 25–29.9) and obese BMI (≥ 30) categories, respectively, compared to 23.5 and 69.5% for males ($p = .65$). Female and male average neck circumference was significantly different ($F$ 15.2 ± 1.6 vs. $M$ 17.4 ± 1.7, $p < .001$). 37.7% of the female patients had a neck circumference ≥ 15.75 in. (88.2% for male, $p < .001$). BMI and neck circumference were significantly correlated for both genders, and BMI correlated with age for females only (Table 2). Analysis of the gender difference across the conventional thresholds shows the gaps in responsiveness of these variables after dichotomization, with a majority of women falling above the BMI threshold and men below, and a majority of men falling above the neck circumference threshold (Figs. 1 and 2).

**Table 2** Correlation of BMI, age, and neck circumference by gender

| Gender | Variables | Correlation | 95% CI | *p* |
|---|---|---|---|---|
| Female | BMI~age | − .18 | − .27, − .08 | < .001 |
| | BMI~neck circumference | .67 | .61, .72 | < .001 |
| | Age~neck circumference | − .10 | − .19, .00 | .05 |
| Male | BMI~age | − .08 | − .16, .01 | .08 |
| | BMI~neck circumference | .72 | .68, .76 | < .001 |
| | Age~neck circumference | − .03 | − .12, .05 | .42 |

Data comparison was done between patients who went through PSG based in-lab test (PSG group, $n = 695$) and those who accepted PM-based HST (HST group, $n = 240$) using chi square or $t$ tests, and no significant differences were found between them: (1) Demographics: percentage of male: 56.4 vs. 59.1%, $p = .442$; age: 50.8 years old vs. 52.2 years old, $p = .235$; 2) BMI: 35.8 vs. 35.2, $p = .370$; weight: 231.0 vs. 230.7 lb, $p = .953$; neck circumference: 41.9 vs. 42.1 cm, $p = .550$; 3) AHI/REI score: 25.1 vs. 25.7, $p = .746$; 4) STOP-Bang score: 4.60 vs. 4.46, $p = .267$; 5) OSA diagnosis and severity: 23.96% (no OSA), 27.59% (mild OSA), 19.52% (moderate OSA), 28.94% (severe OSA) vs. 17.12, 27.24, 22.57, and 33.07%, $p = .112$.

### STOP-Bang and AHI/REI results

There were 819 patients (87.6%) with STOP-Bang total score ≥ 3. The mean total was 4.6 ± 1.7 (STOP 2.6 ± 1.1; Bang 2.0 ± 1.1). After extracting gender from "Bang," the mean Ban score was 1.4 ± 0.9. These scores showed statistically significant gender differences (STOP: $F$ 2.43 ± 1.06 vs. $M$ 2.67 ± 1.09, $p = .0007$; STOP-Bang: $F$ 3.75 ± 1.56 vs. $M$ 5.15 ± 1.51, $p < .001$; Ban: $F$ 1.33 ± 0.93 vs. $M$ 1.48 ± 0.90, $p = .01$). Sleep study AHI/REI scores differed significantly by gender ($F$ 18.9 ± 23.7 vs. $M$ 30.3 ± 28.6, $p < .001$). Of all, 208 (22.3%) were free of OSA ($F$ 30.8%; $M$ 15.8%), 258 (27.6%) had mild OSA ($F$ 31.8% vs. $M$ 24.4%), 189 (20.2%) had moderate OSA ($F$ 17.9 vs. $M$ 22.0%) and 280 (30.0%) had severe OSA ($F$ 19.6 vs. $M$ 37.8%). The male/female ratio for any OSA was 1.2:1. A higher percentage of male than female patients with moderate or severe OSA was seen (59.8 vs. 37.5%, $p < .001$).

Using STOP-Bang cutoff of ≥ 3 or 4, feature of the false negative cases was analyzed. These missed patients varied in BMI and age and neck circumference, revealing statistically significant gender-based differences (Table 3).

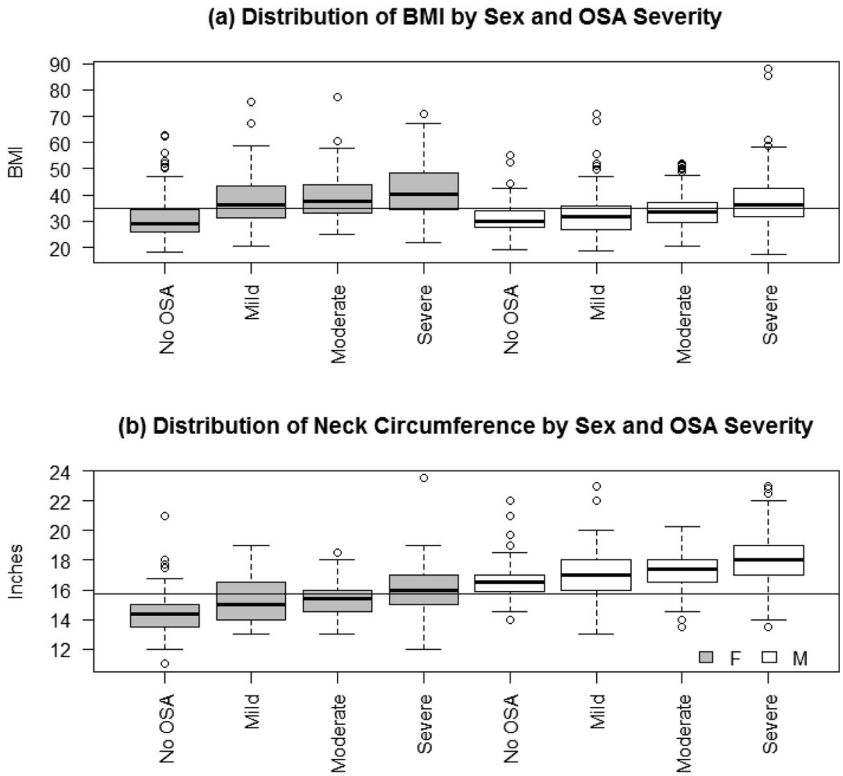


**Fig. 1** Box plots for BMI and neck circumference (in inches) by gender and OSA diagnosis with conventional STOP-Bang thresholds represented by horizontal lines

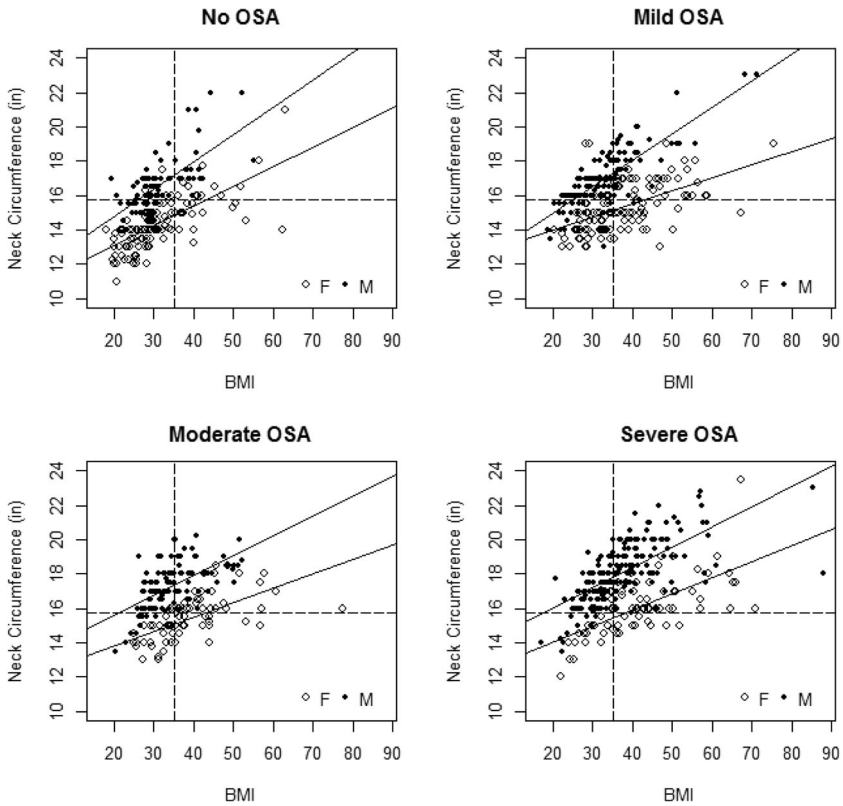


**Fig. 2** Gender difference across the thresholds of neck circumference ≥ 15.75 in. and BMI ≥ 35

**Table 3** Characteristics of false negatives (STOP-Bang < 3 or 4 with some type of OSA/AHI/REI > 5)

| Cutoff | Variable | Female | Male | Estimate | 95% CI | *p* |
|---|---|---|---|---|---|---|
| STOP-Bang ≥ 4 as the cutoff for sleep study referral | False negative *n* (%) | 93 (23.1%) | 47 (8.8%) | 14.3% | 9.5, 19.0% | < .001 |
| | Mild OSA | 54 | 23 | | | |
| | Moderate OSA | 22 | 14 | | | |
| | Severe OSA | 17 | 10 | | | |
| | STOP | 1.6 (.7) | 1.0 (.7) | .6 | .3, .8 | < .001 |
| | BMI | 35.1 (8.4) | 30.3 (6.5) | 4.8 | 2.2, 7.3 | < .001 |
| | Age | 52.4 (15.6) | 47.7 (17.8) | 4.7 | − 1.4, 10.7 | .13 |
| | Neck Circumference | 14.9 (1.3) | 16.6 (1.6) | − 1.7 | − 2.2, − 1.1 | < .001 |
| STOP-Bang ≥ 3 as the cutoff for sleep study referral | False negative *n* (%) | 30 (7.4%) | 15 (2.8%) | 4.6% | 1.7, 7.5 | .001 |
| | Mild OSA | 22 | 7 | | | |
| | Moderate OSA | 6 | 6 | | | |
| | Severe OSA | 8 | 2 | | | |
| | STOP | 1.1 (.6) | .5 (.5) | .6 | .2, .9 | .004 |
| | BMI | 32.5 (6.2) | 30.4 (6.2) | 2.1 | − 1.8, 6.0 | .28 |
| | Age | 50.3 (16.0) | 39.3 (11.5) | 11.0 | 2.9, 19.0 | .009 |
| | Neck circumference | 14.8 (1.5) | 16.7 (1.3) | − 1.9 | − 2.7, − 1.1 | < .001 |

## ROC analyses on conventional STOP-Bang

Using the cutoff at ≥ 3, the specificity of STOP-Bang was low at all OSA levels, particularly for males (Table 4). As AHI/REI increases, specificity improved for females but worsened for males. For both genders, STOP-Bang performed the best at ≥ 5 for AHI/REI ≥ 5. For AHI/REI ≥ 15, the female group showed the best utility, using *J* as the standard, at ≥ 4 while the optimal cutoff for males was ≥ 5. For severe OSA only, the best cut-point was ≥ 5 for both genders.

## ROC analyses to set item OOPs

After age was weighted, BMI outperformed STOP-Bang (AUC .75 ± .004 vs. .74 ± .004, $\chi^2 = 8.5$, $p = .004$) at all OSA levels in females. The BMI optimal operating point for any OSA in females was 31.1 ($J = .44$, 74.4% correctly predicted, 77.9% sensitivity, 65.7% specificity).

For males, neck circumference outperformed STOP-Bang at AHI/REI ≥ 5 (AUC .69 ± .005 vs. .67 ± .005, $\chi^2 = 8.2$, $p = .004$), AHI/REI ≥ 15 (AUC .70 ± .002 vs. .66 ± .003, $\chi^2 = 83.4$, $p < .001$), and AHI/REI ≥ 30 (.70 ± .003 vs. .64 ± .003, $\chi^2 = 244.2$, $p < .001$). The neck circumference OOP was 17.1 in. for male patients in predicting any OSA ($J = .34$, 56.4% correctly predicted, 52.6% sensitivity, 81.8% specificity), moderate/severe OSA ($J = .35$, 65.8% correctly predicted, 61.0% sensitivity, 74.3% specificity), and severe OSA ($J = 0.33$, 66.1% correctly predicted, 67.6% sensitivity, 65.0% specificity).

## Alternative scoring systems and ROC analyses

Four modifications to the conventional STOP-Bang scoring algorithm were devised based on the above findings: (1) BMI ≥ 30 as alternative cutoff (SB_B30) (WHO obesity cutoff), (2) BMI ≥ 31 as alternative cutoff (SB_B31) (rounded female OOP for any OSA), (3) neck size ≥ 17 as alternative cutoff (SB_N17) (rounded male OOP for any OSA), and (4) both BMI ≥ 30 and neck size ≥ 17 (SB_B30-N17). These revised models were tested against conventional STOP-Bang across gender and OSA categories with or without weighted age. Revised models that outperformed conventional STOP-Bang using unweighted models (Table 5) or age-weighted models (Table 6), with statistical significance, are presented. Females were most responsive to modification of the BMI threshold, while males were most responsive to modifications of the neck circumference threshold.

Comparison of *J* values between STOP-Bang and SB_B30 (cutoff at ≥ 3 and 4, respectively) within the female group for any OSA showed improved results. At a cutoff value of ≥ 3, *J* for STOP-Bang was .16 compared to .30 for SB_B30, while at a cutoff of ≥ 4, *J* values were .32 and .34, respectively. For males, the *J* value comparison was significantly different between conventional STOP-Bang and SB_N17 at three OSA levels. For any OSA, *J* values for SB_N17 at the ≥ 3 and ≥ 4 cutoff points were .14 and .29, while for STOP-Bang they were .09 and .23 respectively. Within the male patient group, at moderate or severe OSA levels and using ≥ 3 or ≥ 4 as the cutoff, *J* values for SB_N17 also surpassed STOP-Bang at .08 and .29 vs. .05 and .16. At the severe OSA level we evaluated performance based on cutoffs of ≥ 4 and ≥ 5. In this case, the improvement for SB_N17 still held at .17 and .23 (vs. .11 and .19 for STOP-Bang for both genders).

Given that the female and male patients differed significantly in age distributions, we weighted age to compare alternative scoring systems against the conventional STOP-Bang. The new systems using neck circumference or

**Table 4** ROC analysis against conventional STOP-Bang score by gender and OSA severity

| | Sensitivity (%) | | Specificity (%) | | Correctly classified (%)* | | Youden Index (*J*) | |
|---|---|---|---|---|---|---|---|---|
| Cut-point | *F* | *M* | *F* | *M* | *F* | *M* | *F* | *M* |
| Any OSA (AHI/REI ≥ 5) | | | | | | | | |
| ≥ 1 | 100.0 | 100.0 | 0.0 | 0.0 | 69.2 | 84.2 | 0.00 | 0.00 |
| ≥ 2 | 100.0 | 99.6 | 0.8 | 1.2 | 69.5 | 84.0 | 0.01 | 0.01 |
| ≥ 3 | 96.8 | 96.7 | 19.4 | 11.9 | 73.0 | 83.3 | 0.16 | 0.09 |
| ≥ 4 | 87.1 | 89.5 | 44.4 | 33.3 | 75.0 | 80.6 | 0.32 | 0.23 |
| *≥ 5* | *66.7* | *71.7* | *69.4* | *61.9* | *67.5* | *70.1* | *0.36* | *0.34* |
| ≥ 6 | 43.4 | 48.0 | 88.7 | 79.8 | 57.3 | 53.0 | 0.32 | 0.28 |
| ≥ 7 | 18.3 | 21.2 | 97.6 | 90.5 | 42.7 | 32.1 | 0.16 | 0.12 |
| > 7 | 5.7 | 6.0 | 100.0 | 97.6 | 34.7 | 20.5 | 0.06 | 0.04 |
| AUC | .75 ± .03 | .70 ± .03 | | | | | | |
| Moderate/Severe OSA (AHI/REI ≥ 15) | | | | | | | | |
| ≥ 1 | 100.0 | 100.0 | 0.40 | 0.0 | 37.5 | 59.8 | 0.00 | 0.00 |
| ≥ 2 | 98.0 | 99.7 | 11.9 | 0.9 | 37.7 | 60.0 | 0.10 | 0.01 |
| ≥ 3 | 90.7 | 97.5 | 30.6 | 7.9 | 44.2 | 61.5 | 0.21 | 0.05 |
| *≥ 4* | *74.2* | 92.5 | *55.6* | 23.8 | *53.1* | *64.9* | *0.30* | 0.16 |
| *≥ 5* | 51.0 | *77.7* | 77.0 | *50.5* | *62.5* | *66.7* | 0.28 | *0.28* |
| ≥ 6 | 24.5 | 53.5 | 93.3 | 71.0 | 67.3 | 60.5 | 0.18 | 0.25 |
| ≥ 7 | 6.6 | 26.1 | 97.6 | 90.7 | 67.5 | 52.1 | 0.04 | 0.17 |
| > 7 | 0.0 | 7.9 | 100.0 | 98.1 | 63.5 | 44.2 | 0.00 | 0.06 |
| AUC | .70 ± .03 | .68 ± .02 | | | | | | |
| Severe OSA (AHI/REI ≥ 30) | | | | | | | | |
| ≥ 1 | 100.0 | 100.0 | 0.3 | 0.0 | 19.9 | 37.8 | 0.00 | 0.00 |
| ≥ 2 | 98.7 | 100.0 | 9.9 | 0.9 | 27.3 | 38.4 | 0.09 | 0.01 |
| ≥ 3 | 89.9 | 99.0 | 25.6 | 7.0 | 38.2 | 41.7 | 0.16 | 0.06 |
| ≥ 4 | 78.5 | 95.0 | 50.0 | 19.6 | 55.6 | 48.1 | 0.29 | 0.15 |
| *≥ 5* | *59.5* | *80.6* | *72.8* | *42.3* | *70.2* | *56.8* | *0.32* | *0.23* |
| ≥ 6 | 30.4 | 57.7 | 90.7 | 65.0 | 78.9 | 62.2 | 0.21 | 0.23 |
| ≥ 7 | 10.1 | 29.9 | 97.5 | 87.0 | 80.4 | 65.4 | 0.08 | 0.17 |
| > 7 | 0.00 | 9.5 | 100.0 | 97.0 | 80.4 | 63.9 | 0.00 | 0.07 |
| AUC | .71 ± .03 | .67 ± .02 | | | | | | |

Data with achieved highest J index are shown in italics

*Correctly classified (%) = (true positivity + true negativity)/(true positivity + true negativity + false positivity + false negativity)

SB_N17 performed better than the conventional STOP-Bang in predicting moderate/severe OSA or severe-only OSA groups in male patients. Alternative systems using SB_B30/31 demonstrated better utility in predicting any OSA in female patients. SB_B30-N17 which incorporates both changes in BMI and neck circumference cutoffs worked better for both male and female patients for all level of OSA severity against STOP-Bang. This shows that (1) BMI of 30 or 31 gives better discriminative power in predicting any OSA in females, and (2) neck circumference of 17 in. yields a better prediction for severe only or moderate/severe OSA in male patients. A combined alternative scoring system with BMI at 30 and neck size at 17 in. are better than STOP-Bang for both genders at all OSA levels (Table 6).

## Discussion

Previously existing findings demonstrate gender differences in the prevalence of sleep apnea that are independent of respiratory disturbance index, age, ethnicity, and various comorbidities. The impact of pathophysiologic differences between genders on OSA is poorly understood. Conventional STOP-Bang inflates scores for males, making them more likely to be sent for follow-up testing under the paradigm of a universal cutoff (and hence the lower specificity). Although it gives the questionnaire a simple way of considering the male predominance of OSA, it neglects the gender-specific nuances of the disorder, oversimplifying the disease gender discrepancies.

**Table 5** AUC analyses comparing alternative BMI and neck circumference thresholds age not weighted

| | AUC Conventional | AUC Alternative | $\chi^2$ | $p$ |
|---|---|---|---|---|
| Female | BMI ≥ 35 | BMI ≥ 30 | | |
| AHI/REI ≥ 5 | .739 ± .004 | .742 ± .004 | 9.3 | < .001 |
| Male | Neck ≥ 15.75 in. | Neck ≥ 17 in. | | |
| AHI/REI ≥ 5 | .671 ± .005 | .679 ± .005 | 32.3 | < .001 |
| AHI/REI ≥ 15 | .660 ± .003 | .673 ± .003 | 181.5 | < .001 |
| AHI/REI ≥ 30 | .638 ± .003 | .651 ± .003 | 209.1 | < .001 |
| Both genders | BMI ≥ 35 and neck ≥ 15.75 in. | BMI ≥ 30 and neck ≥ 17 in. | | |
| AHI/REI ≥ 5 | .742 ± .003 | .747 ± .003 | 19.1 | < .001 |
| AHI/REI ≥ 15 | .707 ± .002 | .711 ± .002 | 10.0 | .002 |
| AHI/REI ≥ 30 | .696 ± .003 | .704 ± .002 | 35.5 | < .001 |

Evidence has suggested that the gender disparities in OSA may be reflected in under-diagnosis of OSA in women during the early stages of the disease, as initial presentation of female patients with OSA has often been interpreted as depression and/or insomnia [43]. However, the severity of certain OSA-associated comorbidities is more likely or worse in women than in men upon initial OSA diagnosis [27]. For example, female patients with moderate OSA have demonstrated more severe endothelial dysfunction than male patients, suggesting that women might be more vulnerable to the effect of OSA on the cardiovascular system [44].

This study validated the application of STOP-Bang among clinically referred patients in a community healthcare context. The results support the use of alternative scoring systems for OSA screening using STOP-Bang for higher-risk patients. The validation outcomes may well result in redesign of the scoring system to improve predictive accuracy. Our study presented reference for how modification of a questionnaire-based screening tool can be achieved among patients with higher OSA risks. Pragmatically, test cutoffs may be

**Table 6** ROC comparisons between conventional STOP-Bang and alternative scoring systems with different item scoring designs on the predictive utility by OSA level for all and by gender (age weighted)

| AHI/REI Category | Scoring System | AUC | SE | $\chi^2$ | $p$ |
|---|---|---|---|---|---|
| AHI/REI ≥ 5 | Conventional | 0.7418 | 0.0029 | 19.1189 | 0.0000 |
| For both genders | STOP-Bang SB_B30-N17 | 0.7474 | 0.0028 | | |
| AHI/REI ≥ 5 | Conventional | 0.7385 | 0.0036 | ref | |
| For females only | STOP-Bang SB_B30 | 0.7418 | 0.0036 | 9.2480 | 0.0024 |
| | SB_B31 | 0.7502 | 0.0036 | 56.5260 | 0.0000 |
| | BMI | 0.7511 | 0.0038 | 8.5157 | 0.0035 |
| | SB_B30-N17 | 0.7417 | 0.0036 | 3.8072 | 0.0510 |
| AHI/REI ≥ 15 | Conventional | 0.7070 | 0.0023 | 10.0160 | 0.0016 |
| For both genders | STOP-Bang SB_B30-N17 | 0.7107 | 0.0023 | | |
| AHI/REI ≥ 15 | Conventional | 0.6599 | 0.0034 | Ref | |
| For males only | STOP-Bang Neck circumference (inches) | 0.6953 | 0.0033 | 83.3810 | 0.0000 |
| | SB_N17 | 0.6728 | 0.0034 | 181.5353 | 0.0000 |
| | SB_B30-N17 | 0.6748 | 0.0034 | 58.5801 | 0.0000 |
| AHI/REI ≥ 30 | Conventional | 0.6961 | 0.0025 | 35.5229 | 0.0000 |
| For both genders | STOP-Bang SB_B30-N17 | 0.7038 | 0.0024 | | |
| AHI/REI ≥ 30 | Conventional | 0.6384 | 0.0033 | Ref | |
| For males only | STOP-Bang Neck circumference (inches) | 0.6970 | 0.0033 | 244.1327 | 0.0000 |
| | SB_N17 | 0.6514 | 0.0033 | 209.1142 | 0.0000 |
| | SB_B30-N17 | 0.6533 | 0.0033 | 63.1352 | 0.0000 |

determined based on the needs of the clinician using the test, rather than by using any particular method.

By providing both *J* and correctly classified, our evaluation presents alternative angles to appreciate the gender differences in the application of STOP-Bang and how the gender disparity varies as OSA severity changes. ROC analysis on test utility, however, would not be affected by the change in cutoffs produced by these different methods [45].

The association between obesity and OSA is well recognized [46]. However, how obesity interacts with gender in OSA and influences its severity has not been fully elucidated. We found that the predictive performance of the tool in female patients was influenced more by BMI, while for men, neck circumference might be most relevant. The validation and modifications can potentially encourage use of STOP-Bang as an effective screening tool and simplify workflow for sleep study referral from non-sleep-medicine providers. Our study suggests that gender-triaged BMI and neck circumference thresholds may be considered in high-risk patient cohorts. The study also suggests revised higher total score thresholds for males and females across OSA severity levels instead of a unified ≥ 3.

This study is based on data from clinically referred high-risk patients, so the results do not necessarily apply directly to pre-operative or primary care populations.

## Limitations

The study could be criticized because both in-lab PSG and two types of PM HSTs were used. Although evidence showed that the two approaches could be comparable if implemented properly [47], some inconsistencies may still exist. CSE, sleep testing, and guideline-led good clinical practice were aligned to ensure the gold standard for OSA diagnosis. The AHI of an in-lab study, even if not 100% equal to the REI from a home sleep study, is nonetheless highly correlated with the latter. Further, no evidence shows that home sleep studies systematically over- or underestimate AHI scores. The AASM standards by defining the REI both acknowledges and adjusts for these differences. This study fairly represents the real-world scenario seen in most community sleep medicine centers, but it may not provide the most accurate screening performance estimates, particularly when patients' preference, physical and financial accessibility, or pre-existing conditions may have the potential to confound patients' choice for PSG vs. HST studies. Due to the study design, no data regarding patients' OSA-relevant comorbidities and referral reasons had been collected. Further studies are warranted to rectify issues as such.

## Conclusion

By applying gender-specific cutoffs at ≥ 4 or ≥ 5 for the STOP-Bang total score, the tool performed better than at a unified cutoff of 3 in clinically referred high-risk patients. Incorporating a BMI cutoff of ≥30 or ≥ 31 helped improve the clinical utility of STOP-Bang in predicting any OSA in females, while incorporating a neck circumference cutoff of ≥ 17 in. improved the discriminative power for male patients for moderate and severe OSA. The alternative scoring system using neck size of ≥ 17 in. and BMI ≥ 30 significantly improved screening utility for all at any OSA severity level.

A better utility in predicting females with AHI/REI ≥ 5 has significant clinical import for referral in women and could reduce or eliminate delayed diagnosis of the disorder.

Our study used a gender-specific approach to examine STOP-Bang utility in a group of clinically suspected patients and tested alternative gender-sensitive scoring approaches, which can be generalized for application in other sleep medicine clinics with similar mixed referral patterns and patient characteristics. Considering substitution of the conventional item scoring methods with a gender triaging may reduce providers' work, simplify screening procedures, and improve care quality.

**Author contributions** J.M. helped design the study, implement statistical analysis, and write/revise the manuscript; B.M.P. set statistical analytical frameworks, wrote and did data management and statistics; B.A.C. and S.S.H. helped with study implementation, quality control, sleep study, writing up the sleep study methodology; P.J.A. and K.T.H. assisted with collaborative activities, study design, data collection, and manuscript revision; S.F.T. helped with writing and revision of the whole manuscript; K.A.M. was responsible for the whole study design, implementation, quality assurance of the sleep studies, and manuscript revision. K.A.M. is taking responsibility for the integrity of the work, from inception to publication.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Informed consent** For this type of study (retrospective study), formal consent is not required. This study was evaluated by the MultiCare IRB and deemed to be a quality improvement project not requiring board review and approval.

## References

1. Beneto A, Gomez-Siurana E, Rubio-Sanchez P (2009) Comorbidity between sleep apnea and insomnia. Sleep Med Rev 13:287–293
2. Torres G, Sanchez-de-la-Torre M, Barbe F (2015) Relationship between OSA and hypertension. Chest 148:824–832
3. Wang X, Ouyang Y, Wang Z, Zhao G, Liu L, Bi Y (2013) Obstructive sleep apnea and risk of cardiovascular disease and

all-cause mortality: a meta-analysis of prospective cohort studies. Int J Cardiol 169:207–214
4. Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, Mohsenin V (2005) Obstructive sleep apnea as a risk factor for stroke and death. N Engl J Med 353:2034–2041
5. Shah N, Redline S, Yaggi HK, Wu R, Zhao CG, Ostfeld R et al (2013) Obstructive sleep apnea and acute myocardial infarction severity: ischemic preconditioning? Sleep Breath 17:819–826
6. Cadby G, McArdle N, Briffa T, Hillman DR, Simpson L, Knuiman M et al (2015) Severity of OSA is an independent predictor of incident atrial fibrillation hospitalization in a large sleep-clinic cohort. Chest 148:945–952
7. Kendzerska T, Gershon AS, Hawker G, Tomlinson G, Leung RS (2014) Obstructive sleep apnea and incident diabetes. A historical cohort study. Am J Respir Crit Care Med 190:218–225
8. Kerner NA, Roose SP (2016) Obstructive sleep apnea is linked to depression and cognitive impairment: evidence and potential mechanisms. Am J Geriatr Psychiatry 24:496–508
9. Kunik ME, Roundy K, Veazey C, Souchek J, Richardson P, Wray NP et al (2005) Surprisingly high prevalence of anxiety and depression in chronic breathing disorders. Chest 127:1205–1211
10. Antic NA, Catcheside P, Buchan C, Hensley M, Naughton MT, Rowland S et al (2011) The effect of CPAP in normalizing daytime sleepiness, quality of life, and neurocognitive function in patients with moderate to severe OSA. Sleep 34:111–119
11. Hiestand DM, Britz P, Goldman M, Phillips B (2006) Prevalence of symptoms and risk of sleep apnea in the US population: results from the national sleep foundation sleep in America 2005 poll. Chest 130:780–786
12. Deary V, Ellis JG, Wilson JA, Coulter C, Barclay NL (2014) Simple snoring: not quite so simple after all? Sleep Med Rev 18:453–462
13. Shahrabani S, Tzischinsky O, Givati G, Dagan Y (2014) Factors affecting the intention and decision to be treated for obstructive sleep apnea disorder. Sleep Breath 18:857–868
14. BaHammam AS, Obeidat A, Barataman K, Bahammam SA, Olaish AH, Sharif MM (2014) A comparison between the AASM 2012 and 2007 definitions for detecting hypopnea. Sleep Breath 18:767–773
15. Epstein LJ, Kristo D, Strollo PJ Jr, Friedman N, Malhotra A, Patil SP et al (2009) Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. J Clin Sleep Med 5:263–276
16. Ramachandran SK, Josephs LA (2009) A meta-analysis of clinical screening tests for obstructive sleep apnea. Anesthesiology 110: 928–939
17. Johns MW (1993) Daytime sleepiness, snoring, and obstructive sleep apnea: the Epworth Sleepiness Scale. Chest 103:30–36
18. Abrishami A, Khajehdehi A, Chung F (2010) A systematic review of screening questionnaires for obstructive sleep apnea. Can J Anaesth 57:423–438
19. Chung F, Yang Y, Liao P (2013) Predictive performance of the STOP-Bang score for identifying obstructive sleep apnea in obese patients. Obes Surg 23:2050–2057
20. Chung F, Subramanyam R, Liao P, Sasaki E, Shapiro C, Sun Y (2012) High STOP-Bang score indicates a high probability of obstructive sleep apnoea. Br J Anaesth 108:768–775
21. Ong TH, Raudha S, Fook-Chong S, Lew N, Hsu AA (2010) Simplifying STOP-BANG: use of a simple questionnaire to screen for OSA in an Asian population. Sleep Breath 14:371–376
22. Nagappa M, Liao P, Wong J, Auckley D, Ramachandran SK, Memtsoudis S et al (2015) Validation of the STOP-Bang questionnaire as a screening tool for obstructive sleep apnea among different populations: a systematic review and meta-analysis. PLoS One 10: e0143697
23. Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K et al (2017) Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 13:479–504
24. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S et al (2008) STOP questionnaire: a tool to screen patients for obstructive sleep apnea. Anesthesiology 108:812–821
25. Silva GE, Vana KD, Goodwin JL, Sherrill DL, Quan SF (2011) Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth Sleepiness Scales. J Clin Sleep Med 7:467–472
26. YOUNG T, PEPPARD PE (2005) Clinical presentation of OSAS: gender does matter. Sleep 28:293–295
27. Ye L, Pien GW, Weaver TE (2009) Gender differences in the clinical manifestation of obstructive sleep apnea. Sleep Med 10:1075–1084
28. Bixler EO, Vgontzas AN, Lin HM, Ten Have T, Rein J, Vela-Bueno A et al (2001) Prevalence of sleep-disordered breathing in women: effects of gender. Am J Respir Crit Care Med 163:608–613
29. Chung F, Yang Y, Brown R, Liao P (2014) Alternative scoring models of STOP-bang questionnaire improve specificity to detect undiagnosed obstructive sleep apnea. J Clin Sleep Med 10:951–958
30. Nahapetian R, Silva GE, Vana KD, Parthasarathy S, Quan SF (2016) Weighted STOP-Bang and screening for sleep-disordered breathing. Sleep Breath 20:597–603
31. Farney RJ, Walker BS, Farney RM, Snow GL, Walker JM (2011) The STOP-Bang equivalent model and prediction of severity of obstructive sleep apnea: relation to polysomnographic measurements of the apnea/hypopnea index. J Clin Sleep Med 7:459–65B
32. Chung F, Abdullah HR, Liao P (2016) STOP-Bang questionnaire: a practical approach to screen for obstructive sleep apnea. Chest 149: 631–638
33. Kushida CA, Littner MR, Morgenthaler T, Alessi CA, Bailey D, Coleman J Jr et al (2005) Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 28:499–523
34. Collop NA, Anderson WM, Boehlecke B, Claman D, Goldberg R, Gottlieb DJ et al (2007) Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. J Clin Sleep Med Off Public Am Acad Sleep Med 3:737–747
35. Nilius G, Domanski U, Schroeder M, Franke K-J, Hogrebe A, Margarit L et al (2017) A randomized controlled trial to validate the Alice PDX ambulatory device. Nat Sci Sleep 9:171
36. Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK et al (2012) Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine. J Clin Sleep Med Off Public Am Acad Sleep Med 8:597
37. Choi JH, Kim EJ, Kim YS, Choi J, Kim TH, Kwon SY et al (2010) Validation study of portable device for the diagnosis of obstructive sleep apnea according to the new AASM scoring criteria: Watch-PAT 100. Acta Otolaryngol 130:838–843
38. White DP (2008) Monitoring peripheral arterial tone (PAT) to diagnose sleep apnea in the home. J Clin Sleep Med 4:73
39. Akobeng AK (2007) Understanding diagnostic tests 3: receiver operating characteristic curves. Acta Paediatr 96:644–647
40. Perkins NJ, Schisterman EF (2006) The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 163:670–675
41. Fluss R, Faraggi D, Reiser B (2005) Estimation of the Youden Index and its associated cutoff point. Biom J 47:458–472
42. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM (2006) Prevalence of overweight and obesity in the United States, 1999-2004. JAMA 295:1549–1555

43. McKinney J, Ortiz-Young D, Jefferson F (2015) Gender differences in obstructive sleep apnea and the associated public health burden. Sleep Biol Rhythms 13:196–209
44. Faulx MD, Larkin EK, Hoit BD, Aylor JE, Wright AT, Redline S (2004) Sex influences endothelial function in sleep-disordered breathing. Sleep 27:1113–1120
45. Larner AJ (2015) Optimising the cutoffs of cognitive screening instruments in pragmatic diagnostic accuracy studies: maximising accuracy or the Youden index? Dement Geriatr Cogn Disord 39:167–175
46. Schwartz AR, Patil SP, Laffan AM, Polotsky V, Schneider H, Smith PL (2008) Obesity and obstructive sleep apnea: pathogenic mechanisms and therapeutic approaches. Proc Am Thorac Soc 5:185–192
47. Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus C, Vaughn B (2012) The AASM manual for the scoring of sleep and associated events: Rules, Terminology and Technical Specifications, American Academy of Sleep Medicine, Darien