NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Validation of the STOP-BANG Questionnaire among Patients Referred for Suspected Obstructive Sleep Apnea

Grace Boynton[1], **Arshia Vahabzadeh**[1], **Sami Hammoud**[1], **Deborah L. Ruzicka**[1], and **Ronald D. Chervin**[1,*]

[1]Sleep Disorders Center and Department of Neurology, University of Michigan Health System, 1500 East Medical Center Drive, Ann Arbor, MI, USA

Abstract

Background—The STOP-BANG is a simple obstructive sleep apnea (OSA) screening tool, part questionnaire (STOP) and part demographic or physical measures (BANG), developed for use in preoperative surgical clinics. This study assessed sensitivity and specificity of the instrument among patients referred to a sleep disorders laboratory, and also its performance characteristics when BANG physical measures are patient-reported rather than measured.

Methods—Adults referred for diagnostic polysomnography completed the STOP questions and answered four yes/no questions (BANG self-reported) about their body mass index (weight and height), age, neck circumference, and gender, which were also assessed by laboratory technologists (BANG-measured).

Results—Among N=219 subjects (mean age 46.3 ± 13.9 [s.d.] years; 98 [44.8%] males) the sensitivity of the STOP-BANG measured for an apnea/hypopnea index (AHI, events per hour of sleep) >5, >15, and >30 was 82, 93, and 97% respectively. Corresponding negative predictive values were 44, 87, and 96%. Specificities were comparatively low (48, 40, and 33%). The STOP-BANG measured and STOP-BANG self-reported scores showed essentially equivalent test characteristics against polysomnography.

Conclusions—The STOP-BANG appears to have limited utility in a referred, sleep laboratory setting. Negative results help to identify some individuals as unlikely to have moderate-to-severe apnea, and may thereby prove useful in identification of patients who would benefit more from laboratory studies than home studies. A STOP-BANG in which all information is self-reported

*Corresponding author: Ronald D. Chervin, MD, MS, Michael S. Aldrich Sleep Disorders Laboratory, C728 Med Inn Building, 1500 East Medical Center Drive, Ann Arbor, MI 48109-5845, USA, Tel: 734 647 9064; Fax: 734 647 9065; chervin@umich.edu.

may be as effective as the original version, and has potential to facilitate research or community screening where good negative predictive value is required for an effective screening tool.

Keywords

Sleep apnea; Obstructive; STOP-BANG; Questionnaire; Polysomnography; Screening tests; Sleepiness; Sleep disordered breathing

Introduction

Obstructive sleep apnea (OSA) is characterized by repeated interruption of ventilation during sleep due to pharyngeal airway closure [1]. In association with apneic events, patients experience hypoxemia and hypercapnia [2], increased sympathetic activity, peripheral vasoconstriction [3], and swings in intrathoracic pressure [4]. OSA escalates risks of hypertension [5,6], coronary vascular disease [7], and stroke [8]. Moreover, moderate-to-severe OSA is independently associated with all-cause mortality [9,10].

Overt OSA has an estimated prevalence of 3–7% in men and 2–5% in women [11]. However, it is estimated that 80% of men and 93% of women remain undiagnosed [12]. One of the largest barriers to OSA diagnosis is that polysomnography, the gold standard diagnostic procedure, is time-consuming and costly. Additional obstacles in some countries include long wait times at sleep laboratories and insufficient numbers of sleep specialists. Several screening tests have been developed to identify high-risk patients who should undergo sleep studies, and low-risk patients for whom unnecessary testing can be avoided [13–18]. However, many of these screening tests are lengthy and complicated, making their use and interpretation less than convenient. Some studies that evaluated these questionnaires used methods that have been questioned, and reported accuracies of these instruments have varied across studies [18]. Home cardiopulmonary nocturnal recordings can be helpful for OSA diagnosis in some circumstances, but costs, required expertise, and availability may limit their use also.

In contrast to older screening questionnaires such as the Sleep Disorders Questionnaire (SDQ) [19] or the commonly used Berlin Questionnaire [20], the STOP-BANG is a shorter and more straight-forward instrument. It includes a subjective 4-item questionnaire (STOP) and a 4-item portion informed by demographics and measures (BANG) [21,22]. For comparison, the Sleep Disorders Questionnaire (SDQ) [19], Berlin Questionnaire [20], and American Society of Anesthesiologists' (ASA) checklist [23] contain 12, 10, and 12 OSA symptom-items respectively. The STOP-BANG questionnaire is significantly less complex to score than the SDQ, Berlin Questionnaire, and the ASA checklist, and in contrast to the ASA checklist does not require airway evaluation by healthcare staff. The STOP-BANG was developed to screen for OSA in preoperative surgical patients [21]. In that population, the sensitivities of the STOP-BANG for detection of OSA at an apnea/hypopnea index (AHI, number of events per hour sleep) >5, >15, and >30 were 83.6, 92.9, and 100% respectively [21]. Previous investigators did evaluate the STOP-BANG in a sleep laboratory setting retrospectively, by deriving the equivalent of STOP responses from other

questionnaires patients completed, and adding BANG information from medical records [24]. They found that the STOP-BANG had a sensitivity of 81.5% for detecting AHI 5.

The aim of present study was to prospectively assess the sensitivity and specificity of the STOP-BANG in the setting of a sleep laboratory, where a screen for severe OSA could help to prioritize laboratory studies or determine which patients may be good or poor candidates for an expedited home study. Additionally, this study was designed to assess sensitivity and specificity of the STOP-BANG when patients, instead of technologists, report values for BANG items (see below) that are usually measured. The ability to forego physical assessment would enhance the convenience of the STOP-BANG, facilitate its use in large-scale studies or busy clinic settings, and perhaps lead to community studies of its potential for public health screening.

Materials and Methods

Subjects

This prospective study was conducted at two University of Michigan sites, each accredited by the American Academy of Sleep Medicine: the Michael S. Aldrich Sleep Disorders Laboratory and the Sleep Disorders Laboratory -- South State Street. At these laboratories, approximately 60% of the patients are referred from non-sleep specialist faculty, who nonetheless see and treat patients with OSA, and the remainder is referred by faculty who are board-certified or eligible sleep specialists. The study was approved by the Institutional Review Board of the University of Michigan Medical School. Patients were eligible to participate if they were at least 18 years old, English-speaking, and referred for diagnostic, baseline polysomnography to assess for OSA. Patients were excluded if they were unable to read, sign, or understand an informed consent. Patients were also excluded if they had been previously diagnosed or treated for OSA. Subjects who met the inclusion criteria were approached by the research team and asked to complete both portions of the STOP-BANG on the evening of the sleep study. Neither sleep specialists nor non-sleep specialist faculty who refer to the sleep laboratories use the STOP-BANG in clinical practice and therefore the administration of the instrument in this study was likely to have been the first time each patient saw the instrument.

In addition to values reported by patients on the questionnaire, data on age and gender were also obtained from medical records. Patient weight and height were measured by sleep laboratory technologists on the night of the sleep study and body mass index (BMI) was subsequently calculated. Technologists also measured neck circumference on the night of the sleep study. Neck circumference was measured directly below the laryngeal prominence and measurements were made perpendicular to the vertical axis of the neck.

STOP-BANG Questionnaire

The STOP-BANG questionnaire used in this study included four yes/no questions: S- "Do you **S**nore loudly (louder than talking or loud enough to be heard through closed doors)", T- "Do you often feel **T**ired, fatigued, or sleepy during daytime?", O- "Has anyone **O**bserved you stop breathing during your sleep?", and P- "Do you have or are you being treated for

high blood **P**ressure?" [21] One point was given for each affirmative response. The BANG portion of the questionnaire asked patients to report their height and weight (from which **B**MI was calculated), **A**ge, **N**eck circumference or collar size, and **G**ender [21]. The BMIs, ages, genders, and neck circumferences of the patients were also separately measured as above or obtained from medical records. Gender was included among the other measured variables because the original STOP-BANG asks the technologist to record patient gender [21]. Patients received an additional point toward their STOP-BANG scores for the presence of each of the following clinical characteristics: BMI >35, age > 50, neck circumference > 40cm, and male gender. For the purposes of this research, patients were classified as having high risk for OSA if they had a total STOP-BANG score 3 points, out of a possible 8 points [21]. As both self-reported and measured or observed values for BMI, age, neck circumference, and gender were collected, two sets of scores were calculated. One STOP-BANG score (the "STOP-BANG self-reported score") was based entirely on patient responses to STOP questions and their self-reported BANG values. The second STOP-BANG score (the "STOP-BANG measured score") was based on patient responses to STOP questions and the BANG values that were measured by technicians or obtained from patient health records. This version of the questionnaire was completed by the research team after the study appointment.

Polysomnography

The results of a single nocturnal, laboratory-based sleep study were used to evaluate the utility of the STOP-BANG questionnaire. The diagnosis of OSA requires an apnea-hypopnea index greater than 5 events per hour of sleep, coupled with daytime sleepiness or symptoms of disturbed sleep [1]. Obstructive and central apneas were defined as the complete absence of airflow for at least 10 seconds, in the presence or absence of continued respiratory effort respectively. Hypopneas were defined as a 50% decrease in airflow followed by an arousal, awakening, or 3% desaturation from baseline levels, consistent with American Academy of Sleep Medicine guidelines available at the time the studies were performed [25]. A 2012 revision in these guidelines now allows hypopneas to be scored similarly, except that the decrement in airflow must only be 30% [26]. Guidelines of the American Academy of Sleep Medicine classify mild OSA as AHI> 5 through 15, moderate OSA as AHI>15 through 30, and severe OSA as AHI> 30 [1]. Both the technologists who scored the studies and the physicians who interpreted them were masked to STOP-BANG scores.

Data analysis

Data were double entered into Excel and verified for accuracy. Data analyses were performed using PASW Statistics 18 (SPSS Inc, Chicago, IL, USA). The OSA risk as determined by STOP-BANG was compared to polysomnographic results. Receiver operating characteristic (ROC) curve analyses were performed. Specificities, sensitivities, positive predictive values, negative predictive values, odds ratios, and areas under the curve were calculated to assess the effectiveness of the STOP-BANG measured score against polysomnography. The optimal point threshold for risk classification by the questionnaire was determined. Alternative STOP-BANG scoring models were also explored. To evaluate them, the sensitivities and specificities of each alternative model were compared to the test

characteristics of the original STOP-BANG measured model. Finally, patient self-reported values for weight, height, and calculated BMI were compared with values measured by technologists, and the statistical significance of the differences was assessed using paired t-tests. The level of significance was set at p <.05.

Results

Between May 2011 and September 2011, 219 patients consented to fill out the STOP-BANG questionnaire and underwent overnight diagnostic baseline polysomnography on the same night. The mean age was 46.3 ± 13.9 [s.d.] years, 98 (44.8%) were male, 121 (55.2%) were female, and the mean neck circumference was 39.9 ± 5.4 [s.d.] cm. These enrolled patients represented an estimated 90% of eligible patients who were asked to participate in the study. However, not all of the consented subjects completed the questionnaire in its entirety. The response rate for each item on the questionnaire is detailed in Table 1, along with the number and percentage of subjects who scored a point for each question-item.

The question-item with the lowest response rate was neck circumference, with 191 (87%) of the patients reporting a value. Self-reported values for neck circumference included length measurements and/or t-shirt collar sizes (eg: S, M, L, XL). All patients in the study reported their age and responded to the questions that inquired about daytime sleepiness and high blood pressure. The questions that received the highest numbers of affirmative responses were those addressing snoring and daytime sleepiness: 86% of patients in the study reported feeling tired or sleepy during the day and 54% indicated that they snored loudly.

When self-reported weights and heights (and the BMI values that were subsequently calculated from them) were compared to measured values (Table 2), patients under-reported their weights by only 1.7 kgs on average, and the difference was not statistically significant. Patients also tended to under-report their heights, on average by .84 centimeters, and this difference was statistically significant. The BMIs calculated from self-reported values were not significantly different from the BMIs calculated from technologist measurements.

The STOP-BANG measured score identified 165 (75%) of the 219 sleep-laboratory-referred patients to be at high risk for OSA (Table 1). Among these 165 patients, 91 (55.2%) were male and 74 (44.8%) were female. The average STOP-BANG measured score among all subjects was 3.87 ± 1.68 points out of a maximum of 8 points. The distribution of subject STOP-BANG measured scores is displayed in Figure 1. Polysomnography revealed that 169 (77%) of the subjects in the cohort had obstructive sleep apnea, as defined by an apnea-hypopnea index of at least 5. Among these subjects, 103 (47%) of the patients had an AHI>15 and 62 (28%) had an AHI>30.

The test characteristics of the STOP-BANG measured score, as compared to AHI levels of >5, >15, and >30, are shown in Table 3. The STOP-BANG cut-off used was 3 items positive. Receiver operating characteristic (ROC) curves for the STOP-BANG measured (shown in Figures 2–4) showed that this threshold optimized test characteristics. With a cut-off of 3 points, for AHI levels of >5, >15, and >30, respective sensitivities were 82.2, 93.2 and 96.8% and specificities were 48.0, 40.5, and 33.1%. For comparison, at a threshold of

2 points, STOP-BANG measured sensitivities, for AHI levels of >5, >15, and >30, were 94.7, 98.1, and 100% but specificities were 22.0, 13.8, and 11.5%. At a scoring cut-off of 4 points, sensitivities for AHI levels of >5, >15, and >30 were 65.1, 77.7, and 85.5% whereas specificities were 66.0, 59.5, and 52.9%.

The performance of the STOP-BANG measured score in prediction of minimum oxygen saturation values <85% and <70% was also explored. A scoring cut-off of 3 items positive was used. For minimum oxygen saturation values <85%, the STOP-BANG measured had a sensitivity of 93.4% and a specificity of 34.5%. For minimum oxygen saturation values of <70%, the STOP-BANG measured had a sensitivity of 88.9% and a specificity of 25.4%.

Alternative assessment and scoring models

The STOP-BANG self-reported model scores showed a distribution that was essentially identical to that created by the original, STOP-BANG measured model (Figure 5). Whereas the average STOP-BANG measured score was 3.87 ± 1.68 points, the average STOP-BANG self-reported score was 3.73 ± 1.69 points (paired t-test, t= −4.53, p <.0001). The test characteristics of the STOP-BANG self-reported model, as compared to AHI levels of >5, >15, and >30, are also shown in Table 3. For AHIs >5, >15, and >30, the STOP-BANG self-reported score demonstrated test characteristics that were not substantively inferior to those of the STOP-BANG measured model.

Finally, in exploratory analyses, abridged scoring models were also assessed. In particular, three models were examined: one looking only at STOP questions (i.e., without BANG), another looking only at BANG values that were obtained by technologists, and a third based on self-reported BANG values. For each model, test characteristics for the optimal cut-point were used. The scoring model based on only the STOP questions had inferior sensitivity and specificity when compared with the full STOP-BANG measured model (Table 4). In contrast, the sensitivities of the scoring models based on measured BANG values and self-reported BANG values (Table 4) were comparable to the sensitivity of the full STOP-BANG measured score. However, both of these BANG-only scores had compromised specificity in comparison to the full STOP-BANG measured model.

Discussion

This comparison of STOP-BANG scores to results from full polysomnography, among sleep-laboratory-referred patients, suggests potential value of the STOP-BANG questionnaire in identification of some individuals who are not likely to have moderate or severe OSA. Specifically, the STOP-BANG measured had negative predictive values of 87.0% and 96.3% for patients with AHI>15 and AHI>30, respectively. In contrast, identification of referred patients unlikely to have any OSA does not appear to be reliably enhanced by use of the STOP-BANG. Among the patients tested in this study 77.2% (169) had AHI>5 and the positive predictive value of the STOP-BANG questionnaire for AHI>5 was 84.2%. This suggests that the instrument fails to robustly increase the already-high likelihood, in this population, that OSA will be identified on testing.

The STOP-BANG questionnaire was originally validated for OSA screening in surgical patients at preoperative clinics. This study shows that utility of the questionnaire at a sleep disorders center may be more limited. Categorization of some patients as less likely to have prominent OSA could perhaps indicate a need for laboratory rather than home studies. Already at the time of this writing, some third-party-payors use STOP-BANG results to determine whether a patient has high enough pre-test suspicion for OSA to require a home study and treatment with auto-titrating positive airway pressure, rather than a more expensive in-laboratory polysomnogram followed by standard titration study. In this study, a positive STOP-BANG among referred patients did raise the pre-test probability of OSA (of any severity) from 77% to 84% (Table 3), nearly equivalent to the 85% threshold recommended for appropriate use of home studies to assess for OSA [27].

The results of this study demonstrate that the STOP-BANG can have a high sensitivity and negative predictive value, particularly for detection of severe (AHI>30) OSA. A recent study of the STOP-BANG in a sleep-clinic setting similarly found a high sensitivity (93.8%) and low specificity (33.3%) for patients with sleep-disordered breathing [28]. A retrospective study of the STOP-BANG in a sleep-laboratory setting reported a sensitivity of 85.1% for AHI 5 [24], close to the 82.2% we found. However, the authors found that the STOP-BANG numerical score rather than an outcome dichotomized at 3 may be useful to grade the probability of severe OSA and urgency for evaluation. Our results did not seem to support this conclusion. For example, STOP-BANG scores 6 rather than 3 increased the positive predictive value for AHI>30 from 36.4% to 57.5%, but also decreased the negative predictive value from 96.3% to 78.2%. Although a higher STOP-BANG score in our study did increase the probability of severe OSA, most patients with severe OSA had STOPBANG scores < 6. Triage of care based on high STOP-BANG scores would fail to identify many of our patients with severe disease.

In comparison to the STOP-BANG questionnaire, the Berlin questionnaire demonstrated less effectiveness among sleep laboratory patients, with sensitivities of 68% and 57% for respiratory disturbance indices of >5 and >15, and corresponding specificities of 49% and 41% [29]. Among a mixed population of sleep center patients and control subjects, the Sleep Disorders Questionnaire had sensitivities of 85% in men and 88% in women for screening for sleep apnea. Respective specificities were 76% and 81% [19]. In OSA screening among preoperative surgical patients, the ASA checklist had sensitivities of 72.1%, 78.6%, and 87.2% for AHIs >5, >15, and >30 respectively. Corresponding specificities were 38.2%, 37.4%, and 36.2% [23].

If results of the current study can be consistently replicated in primary care settings, the STOP-BANG questionnaire may prove useful for more widespread screening, in settings where lower prevalence of OSA would make the negative predictive value for the STOP-BANG higher than those we found at a sleep laboratory. The performance of the STOP-BANG questionnaire in identifying OSA patients within a general population setting was recently evaluated in São Paulo, Brazil [30]. In that population, the sensitivity of the STOP-BANG for either AHI>5 plus daytime sleepiness or AHI>15 was 83.8%, with a negative predictive value of 87.5%.

Practitioners may be able to rely on an instrument with good sensitivity to identify many patients as unlikely to have moderate-to-severe sleep apnea that carries highest risk for serious cardiovascular morbidity. However, despite good STOP-BANG negative predictive value for severe OSA, the instrument showed poor negative predictive value for mild levels of OSA that can still lead to excessive daytime sleepiness and diminished quality of life [31]. Therefore, a negative STOP-BANG does not rule out OSA as a potential contributor to those morbidities.

Although the STOP-BANG was more sensitive for severe OSA, its sensitivities for moderate (AHI>15) and mild (AHI>5) OSA are sufficient that it might still detect many community patients with these levels of pathology. The STOP-BANG is so simply, quickly, and inexpensively administered that its widespread use could potentially identify many affected patients, and direct them toward further investigation, rather than leaving symptoms ignored. Although many clinicians now recognize features of OSA captured more formally by the STOP-BANG, limited appointment times and focused visits may not provide opportunities for discussion of snoring and sleepiness unless the patient brings them up. Our clinical experience continues to suggest that most newly diagnosed OSA patients continue to report that they have had unaddressed snoring, sleepiness, and other relevant symptoms for many years.

Interestingly, use of self-reported rather than measured values for height, weight, gender, age, and neck circumference reduced the sensitivity of the STOP-BANG only to a negligible extent. The differences in sensitivities between the STOP-BANG measured and STOP-BANG self-reported were 2.9, 4.9, and 3.3 percentage points for AHI>5, >15, and >30 respectively. This suggests that use of self-reported values for patient clinical characteristics may be a satisfactory substitute for measurements by health care personnel. As a result, usefulness of the STOP-BANG may remain undiminished outside clinical settings, when self-report may be the only option. In screening programs administered to large numbers of people by mail or the internet, a negative score on the STOP-BANG may suggest that severe sleep apnea is highly unlikely and moderate OSA unlikely, whereas a positive screen could suggest that a more rigorous evaluation is warranted.

The self-reported BANG scoring model, perhaps even more so than the STOP or the STOP-BANG, may prove especially helpful at a community level. Some of the patients who participated in the study did not know whether they snored loudly at night or if anyone had observed them to stop breathing during their sleep. These patients were therefore unable to complete the STOP questions. Patients who sleep alone at night may not have a partner to alert them about OSA signs. These patients may benefit from using the self-reported BANG scoring model, because it does not take the STOP questions into account. Although this scoring model has lower specificity than the original STOP-BANG scoring model, it retains high sensitivity, so it could still be a useful tool for screening with reasonable negative predictive value for moderate or severe OSA.

Limitations of the current study include failure of all patients to provide all responses on the STOP-BANG questionnaires, though missing data were, in general, infrequent (Table 1). We elected to analyze all available data, in part so that results would be most generalizable

to clinic settings, where failure to complete all questionnaire items is also not uncommon. Another limitation is that the study was conducted at only two sites of one tertiary referral center, in one community, and therefore results could differ to some extent at other academic sleep centers or especially non-academic centers. However, at the two study sites, inclusion criteria were not restrictive and the participation rate, among patients approached, was high. In our sample, only 54% of the referred subjects endorsed loud snoring (Table 1). Sites at which referred patients more commonly admit to this symptom may experience different STOP-BANG performance, though our sample was not overall one that reflected primarily mild OSA: the mean apnea/hypopnea index was 24, with levels >30 suggesting severe OSA [1]. Our study defined OSA by the AHI, and this in itself is a limitation, if necessary for objective research purposes, because in practice OSA is diagnosed through a combination of symptoms, physical exam findings, associated morbidities, and objective findings including others besides the AHI. Finally, this study did not begin to explore all the possible uses that the STOP-BANG may have in the setting of a clinical sleep disorders center. For example, we speculate that its results could prove useful in providing advance information on levels of continuous positive airway pressure likely to be required during titration studies; on types of positive airway pressure devices likely to be necessary; or on likelihood of subsequent compliance with positive airway pressure at home.

The STOP-BANG is highly practical because it is short, has a seemingly easy-to-remember mnemonic, and has a straight-forward scoring system. As clear-cut thresholds exist for risk stratification, interpretation of the test is neither complicated nor user-dependent. Test outcomes are determined quickly without a calculator, and patients do not have to wait long for test results. Patients have relatively little difficulty in completing the STOP-BANG, and overall response rates are high. The instrument's low cost, ease of use, and high sensitivity to severe OSA in perioperative patients and now sleep-laboratory-referred patients also, suggest that the STOP-BANG or BANG alone merit further investigation as possible aids in public health efforts to facilitate diagnosis and treatment of a frequently unrecognized sleep disorder with serious associated morbidity and mortality.

Acknowledgments

References

1. The international classification of sleep disorders: diagnostic and coding manual. 2. American Academy of Sleep Medicine; Westchester: 2005.

2. Somers VK, Dyken ME, Clary MP, Abboud FM. Sympathetic neural mechanisms in obstructive sleep apnea. J Clin Invest. 1995; 96:1897–1904. [PubMed: 7560081]

3. Somers VK, Mark AL, Zavala DC, Abboud FM. Contrasting effects of hypoxia and hypercapnia on ventilation and sympathetic activity in humans. J Appl Physiol. 1989; 67:2101–2106. [PubMed: 2513316]

4. Kapur VK. Obstructive sleep apnea: diagnosis, epidemiology, and economics. Respir Care. 2010; 55:1155–1167. [PubMed: 20799998]

5. Hla KM, Young TB, Bidwell T, Palta M, Skatrud JB, et al. Sleep apnea and hypertension. A population-based study. Ann Intern Med. 1994; 120:382–388. [PubMed: 8304655]

6. Nieto FJ, Young TB, Lind BK, Shahar E, Samet JM, et al. Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. Sleep Heart Health Study. JAMA. 2000; 283:1829–1836. [PubMed: 10770144]

7. Schäfer H, Koehler U, Ewig S, Hasper E, Tasci S, et al. Obstructive sleep apnea as a risk marker in coronary artery disease. Cardiology. 1999; 92:79–84. [PubMed: 10702648]

8. Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, et al. Obstructive sleep apnea as a risk factor for stroke and death. N Engl J Med. 2005; 353:2034–2041. [PubMed: 16282178]

9. Marshall NS, Wong KK, Liu PY, Cullen SR, Knuiman MW, et al. Sleep apnea as an independent risk factor for all-cause mortality: the Busselton Health Study. Sleep. 2008; 31:1079–1085. [PubMed: 18714779]

10. Young T, Finn L, Peppard PE, Szklo-Coxe M, Austin D, et al. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. Sleep. 2008; 31:1071–1078. [PubMed: 18714778]

11. Punjabi NM. The epidemiology of adult obstructive sleep apnea. Proc Am Thorac Soc. 2008; 5:136–143. [PubMed: 18250205]

12. Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. Sleep. 1997; 20:705–706. [PubMed: 9406321]

13. Flemons WW, Whitelaw WA, Brant R, Remmers JE. Likelihood ratios for a sleep apnea clinical prediction rule. Am J Respir Crit Care Med. 1994; 150:1279–1285. [PubMed: 7952553]

14. Hoffstein V, Szalai JP. Predictive value of clinical features in diagnosing obstructive sleep apnea. Sleep. 1993; 16:118–122. [PubMed: 8446830]

15. Kump K, Whalen C, Tishler PV, Browner I, Ferrette V, et al. Assessment of the validity and utility of a sleep-symptom questionnaire. Am J Respir Crit Care Med. 1994; 150:735–741. [PubMed: 8087345]

16. Maislin G, Pack AI, Kribbs NB, Smith PL, Schwartz AR, et al. A survey screen for prediction of apnea. Sleep. 1995; 18:158–166. [PubMed: 7610311]

17. Viner S, Szalai JP, Hoffstein V. Are history and physical examination a good screening test for sleep apnea? Ann Intern Med. 1991; 115:356–359. [PubMed: 1863025]

18. Abrishami A, Khajehdehi A, Chung F. A systematic review of screening questionnaires for obstructive sleep apnea. Can J Anaesth. 2010; 57:423–438. [PubMed: 20143278]

19. Douglass AB, Bornstein R, Nino-Murcia G, Keenan S, Miles L, et al. The Sleep Disorders Questionnaire. I: Creation and multivariate structure of SDQ. Sleep. 1994; 17:160–167. [PubMed: 8036370]

20. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med. 1999; 131:485–491. [PubMed: 10507956]

21. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. Anesthesiology. 2008; 108:812–821. [PubMed: 18431116]

22. Chung F, Subramanyam R, Liao P, Sasaki E, Shapiro C, et al. High STOP-Bang score indicates a high probability of obstructive sleep apnoea. Br J Anaesth. 2012; 108:768–775. [PubMed: 22401881]

23. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, et al. Validation of the Berlin questionnaire and American Society of Anesthesiologists checklist as screening tools for obstructive sleep apnea in surgical patients. Anesthesiology. 2008; 108:822–830. [PubMed: 18431117]

24. Farney RJ, Walker BS, Farney RM, Snow GL, Walker JM. 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. 2011; 7:459–465B. [PubMed: 22003340]

25. The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specifications. American Academy of Sleep Medicine; Westchester, IL: 2007.

26. Berry, RB.; Brooks, R.; Gamaldo, CE.; Harding, SM.; Marcus, CL., et al. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.0. American Academy of Sleep Medicine; Darien, Illinois: 2012.

27. Collop NA, Tracy SL, Kapur V, Mehra R, Kuhlmann D, et al. Obstructive sleep apnea devices for out-of-center (OOC) testing: technology evaluation. J Clin Sleep Med. 2011; 7:531–548. [PubMed: 22003351]

28. Vana KD, Silva GE, Goldberg R. Predictive abilities of the STOP-Bang and Epworth Sleepiness Scale in identifying sleep clinic patients at high risk for obstructive sleep apnea. Res Nurs Health. 2013; 36:84–94. [PubMed: 23007730]

29. Ahmadi N, Chung SA, Gibbs A, Shapiro CM. The Berlin questionnaire for sleep apnea in a sleep clinic population: relationship to polysomnographic measurement of respiratory disturbance. Sleep Breath. 2008; 12:39–45. [PubMed: 17684781]

30. Coelho FM, Pradella-Hallinan M, Palombini L, Tufik S, Bittencourt LR. The STOP-BANG questionnaire was a useful tool to identify OSA during epidemiological study in São Paulo (Brazil). Sleep Med. 2012; 13:450–451. [PubMed: 22386388]

31. Guilleminault C, Stoohs R, Clerk A, Cetel M, Maistros P. A cause of excessive daytime sleepiness. The upper airway resistance syndrome. Chest. 1993; 104:781–787. [PubMed: 8365289]
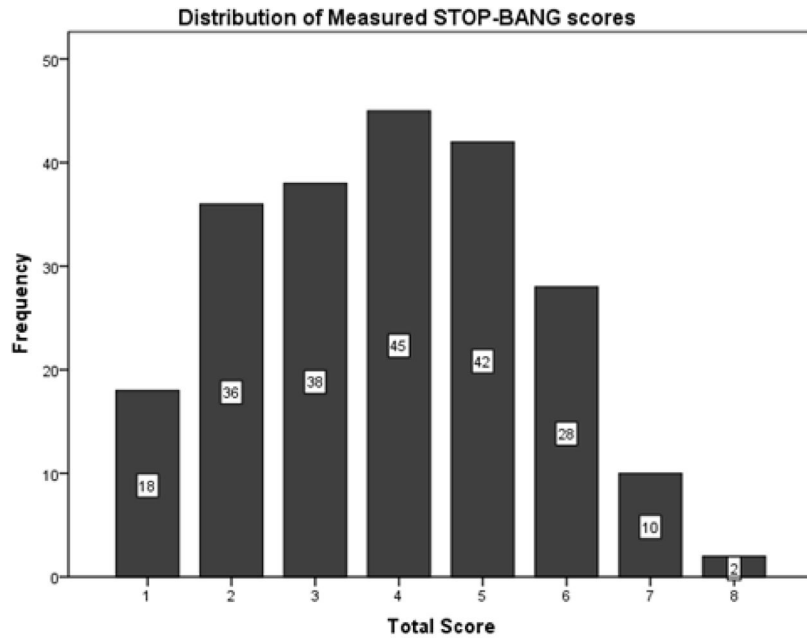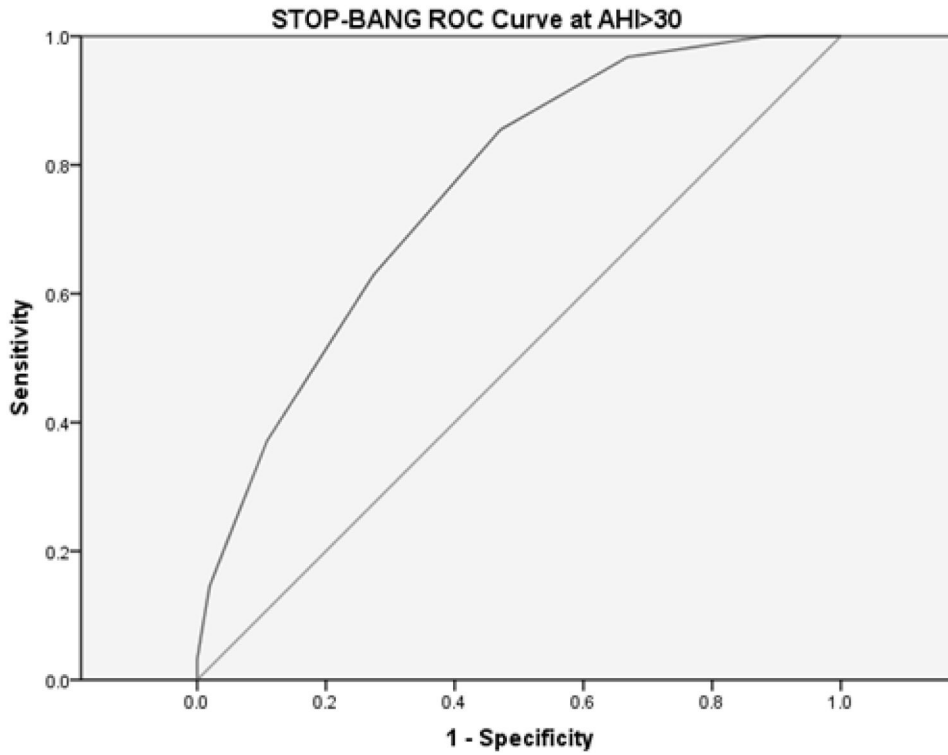
Figure 1.
Histogram showing distribution of measured STOP-BANG scores.
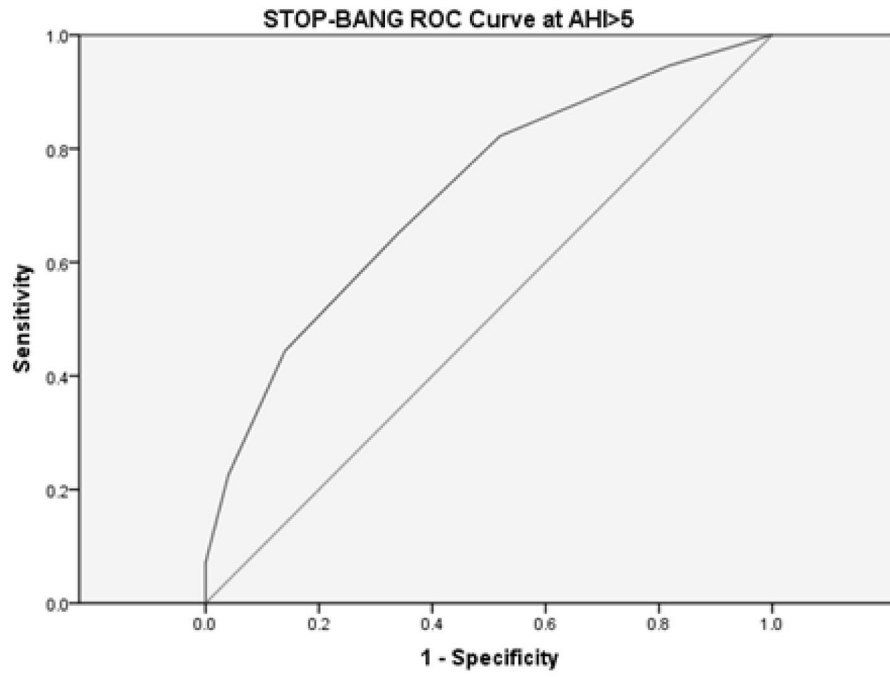
Figure 2.
STOP-BANG measured ROC curve for AHI>5.

The area under the curve for the STOP-BANG questionnaire at AHI>5 was .722 (95% CI [.645, .799]).
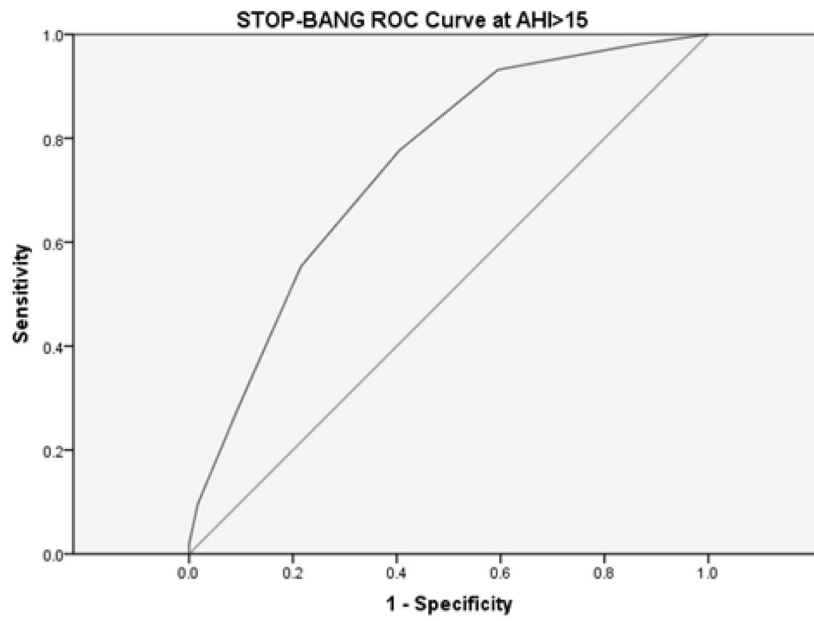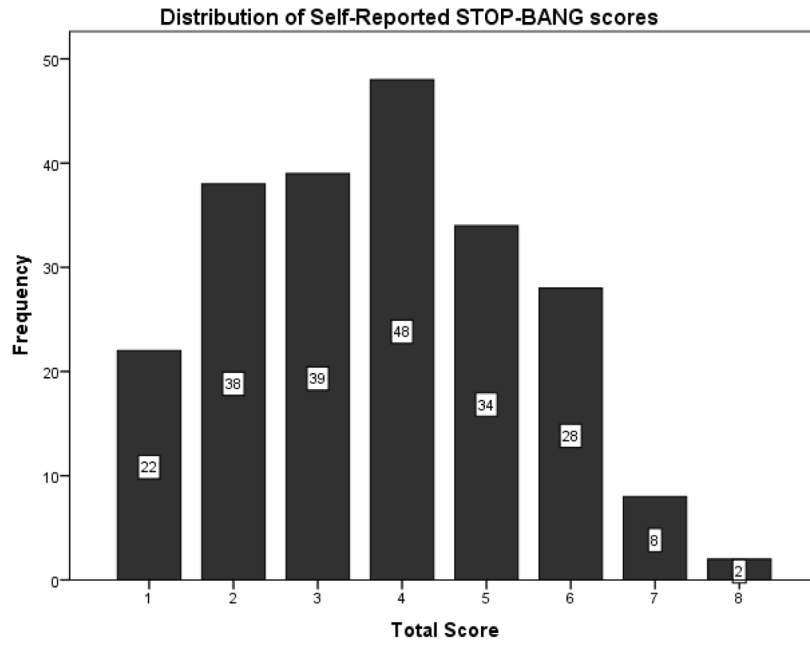
Figure 3.
STOP-BANG measured ROC curve for AHI>15.
The area under the curve for the STOP-BANG questionnaire at AHI>15 was .746 (95% CI [.682, .811]).

Figure 4.
STOP-BANG measured ROC curve for AHI>30.
The area under the curve for the STOP-BANG questionnaire at AHI>30 was .762 (95% CI[.696,.827]).

Figure 5.
Histogram showing distribution of STOP-BANG self-reported scores.

Table 1

Responses to STOP-BANG questionnaire items[†].

	No. (%) of 219 patients who responded to question	No. (%) of 219 patients who endorsed each risk factor
Do you snore loudly?	208 (95.0)	119 (54.3)
Do you often feel tired, fatigued, or sleepy during daytime?	219 (100)	189 (86.3)
Has anyone observed you stop breathing during your sleep?	214 (97.7)	90 (41.1)
Do you have high blood pressure?	219 (100)	75 (34.3)
BMI (>35)	216 (98.6)	83 (37.9)
Age (>50)	219 (100)	88 (40.2)
Neck circumference (>40cm)	191 (87.2)	106 (48.4)
Gender (male)	216 (98.6)	98 (44.8)
No. of patients classified as high risk (3 points on total questionnaire)	---	165 (75.3)

[†]BANG items (BMI, Age, Neck circumference, and Gender) listed here reflect subject response rates on questionnaires, rather than demographic or objective measures

Table 2

STOP-BANG Self-Reported *vs.* STOP-BANG Measured Clinical Characteristics.

	Mean ± SD		Average Difference between self-reported and measured means	Paired samples T-test	
	Self-reported	Measured		t statistic	p-value
Weight (kg)	96.74 ± 33.17	98.44 ± 29.76	−1.70 ± 14.43	−1.738	.084
Height (cm)	170.43 ± 11.02	171.27 ± 10.16	−.84 ± 5.28	−2.368	.019
Body Mass Index (BMI, kg/m^2)	33.32 ± 10.27	33.43 ± 8.76	−.11 ± 5.62	−.298	.766

Table 3

Test characteristics (95% confidence intervals) for the STOP-BANG measured and STOP-BANG self-reported scores.

	AHI>5	AHI>15	AHI>30
STOP-BANG measured			
Sensitivity (95%CI)	82.2% (78.6–85.7)	93.2% (87.3–96.9)	96.8% (88.8–99.4)
Specificity (95%CI)	48.0% (35.8–59.8)	40.5% (35.3–43.8)	33.1% (30.0–34.2)
PPV (95%CI)	84.2% (80.5–87.8)	58.2% (54.5–60.5)	36.4% (33.3–37.4)
NPV (95%CI)	44.4% (33.1–55.4)	87.0% (75.9–94.0)	96.3% (87.1–99.4)
Odds ratio (95%CI)	4.28 (2.05–8.95)	9.34 (3.77–24.17)	14.86 (3.37–91.50)
Area under ROC curve (95%CI)	.722 (.645–.799)	.746 (.682–.811)	.762 (.696–.827)
STOP-BANG self-reported			
Sensitivity (95%CI)	79.3% (75.6–82.8)	88.3% (81.9–93.2)	93.5% (84.7–97.9)
Specificity (95%CI)	50.0% (37.6–62.0)	41.4% (35.7–45.7)	35.7% (32.2–37.4)
PPV (95%CI)	84.3% (80.4–88.1)	57.2% (53.1–60.4)	36.5% (33.0–38.2)
NPV (95%CI)	41.7% (31.3–51.7)	80.0% (69.0–88.4)	93.3% (84.2–97.8)
Odds ratio (95%CI)	3.83 (1.87–7.89)	5.35 (2.52–11.57)	8.04 (2.62–27.55)
Area under ROC curve (95%CI)	.714 (.637–.792)	.736 (.670–.802)	.758 (.690–.826)

Table 4

Test characteristics (95% confidence intervals) for the STOP score (without BANG) [†], the BANG measured score[‡], and the BANG self-reported score[§].

	AHI>5	AHI>15	AHI>30
STOP			
Sensitivity (95%CI)	74.6% (71.2–78.3)	80.6% (73.8–86.6)	83.9% (73.7–91.3)
Specificity (95%CI)	34.0% (22.6–46.6)	34.5% (28.5–39.8)	31.8% (27.8–34.8)
PPV (95%CI)	79.2% (75.7–83.2)	52.2% (47.8–56.1)	32.7% (28.8–35.6)
NPV (95%CI)	28.3% (18.9–38.8)	66.7% (55.1–76.9)	83.3% (72.9–91.0)
Odds ratio (95%CI)	1.51 (.72–3.14)	2.18 (1.13–4.26)	2.43 (1.08–5.57)
Area under ROC curve (95%CI)	.607 (.524–.690)	.629 (.555–.703)	.649 (.569–.728)
BANG measured			
Sensitivity (95%CI)	89.9% (86.7–93.0)	94.2% (88.8–97.5)	98.4% (91.2–99.9)
Specificity (95%CI)	36.0% (25.1–46.5)	25.0% (20.2–28.0)	21.7% (18.8–22.3)
PPV (95%CI)	82.6% (79.7–85.5)	52.7% (49.7–54.6)	33.2% (30.7–33.7)
NPV (95%CI)	51.4% (35.9–66.4)	82.9% (67.0–92.7)	97.1% (84.3–99.9)
Odds ratio (95%CI)	5.03 (2.19–11.59)	5.39 (2.01–15.25)	16.86 (2.38–339.04)
Area under ROC curve (95%CI)	.735 (.658–.813)	.765 (.701–.828)	.772 (.706–.837)
BANG self-reported			
Sensitivity (95%CI)	88.8% (85.4–92.0)	93.2% (87.6–96.9)	96.8% (89.0–99.4)
Specificity (95%CI)	40.0% (28.7–50.8)	27.6% (22.6–30.8)	23.6% (20.5–24.6)
PPV (95%CI)	83.3% (80.2–86.3)	53.3% (50.1–55.4)	33.3% (30.6–34.2)
NPV (95%CI)	51.3% (36.8–65.1)	82.1% (67.3–91.7)	94.9% (82.4–99.1)
Odds ratio (95%CI)	5.26 (2.36–11.80)	5.22 (2.07–13.76)	9.25 (2.08–57.49)
Area under ROC curve (95%CI)	.726 (.648–.805)	.751 (.686–.816)	.770 (.704–.837)

[†] For this model, subjects were considered to be at high risk for OSA if they answered "yes" to 2 of the 4 STOP questions

[‡] For this model, subjects were considered to be at high risk for OSA if they received 1 point out of the 4 possible points on the BANG portion of STOP-BANG measured score

[§] For this model, subjects were considered to be at high risk for OSA if they received 1 point out of the 4 possible points on the BANG portion of the STOP-BANG self-reported score