

STOP Questionnaire

A Tool to Screen Patients for Obstructive Sleep Apnea

Frances Chung, F.R.C.P.C.,* Balaji Yegneswaran, M.B.B.S.,† Pu Liao, M.D.,‡ Sharon A. Chung, Ph.D.,§
Santhira Vairavanathan, M.B.B.S.,|| Sazzadul Islam, M.Sc.,|| Ali Khajehdehi, M.D.,† Colin M. Shapiro, F.R.C.P.C.#

Background: Obstructive sleep apnea (OSA) is a major risk factor for perioperative adverse events. However, no screening tool for OSA has been validated in surgical patients. This study was conducted to develop and validate a concise and easy-to-use questionnaire for OSA screening in surgical patients.

Methods: After hospital ethics approval, preoperative patients aged 18 yr or older and without previously diagnosed OSA were recruited. After a factor analysis, reliability check, and pilot study; four yes/no questions were used to develop this screening tool. The four questions were respectively related to snoring, tiredness during daytime, observed apnea, and high blood pressure (STOP). For validation, the score from the STOP questionnaire was evaluated *versus* the apnea-hypopnea index from monitored polysomnography.

Results: The STOP questionnaire was given to 2,467 patients, 27.5% classified as being at high risk of OSA. Two hundred eleven patients underwent polysomnography, 34 for the pilot test and 177 for validation. In the validation group, the apnea-hypopnea index was 20 ± 6 . The sensitivities of the STOP questionnaire with apnea-hypopnea index greater than 5, greater than 15, and greater than 30 as cutoffs were 65.6, 74.3, and 79.5%, respectively. When incorporating body mass index, age, neck circumference, and gender into the STOP questionnaire, sensitivities were increased to 83.6, 92.9, and 100% with the same apnea-hypopnea index cutoffs.

Conclusions: The STOP questionnaire is a concise and easy-to-use screening tool for OSA. It has been developed and validated in surgical patients at preoperative clinics. Combined with body mass index, age, neck size, and gender, it had a high sensitivity, especially for patients with moderate to severe OSA.

OBSTRUCTIVE sleep apnea (OSA) is the most prevalent breathing disturbance in sleep,¹ affecting 2–26% of the general population depending on sex, age, and the definition of criteria.² OSA is associated with significant morbidity, including excessive daytime sleepiness, loud snoring during sleep, refractory hypertension, and impaired quality of life. Studies have also shown that OSA

is associated with a high risk for traffic accidents and cardiovascular disease.^{3,4}

The prevalence of OSA in the surgical population is higher than in the general population and varies with different surgical populations. In particular, approximately 7 of every 10 patients undergoing bariatric surgery were found to have OSA,⁵ presumably because of the high level of obesity in this surgical population. Of even greater concern, despite OSA being present in the majority of patients presenting for bariatric surgery,^{5,6} most cases were not diagnosed, and careful screening was not implemented before surgery.⁶ One of the barriers to study the prevalence of OSA in surgical patients is the difficulty with recruiting patients to undergo polysomnography before surgery. Fidan *et al.*⁷ screened 433 surgical patients; only 18 of 41 invited patients agreed to undergo polysomnographic testing, and 14 patients (3.2% of all screened patients) were diagnosed with OSA. In another study conducted by Chung *et al.*,⁸ 24% of 305 surgical patients were classified as being at high risk of having OSA using the Berlin questionnaire, and 13 patients were confirmed as having OSA by polysomnography, 4.2% of the total number of patients screened.

It is estimated that nearly 80% of men and 93% of women with moderate to severe sleep apnea are undiagnosed.⁹ Undiagnosed OSA may pose a variety of problems for anesthesiologists. A number of case reports have documented an increase in the incidence of postoperative complications and deaths among patients suspected of having OSA.¹⁰ Untreated OSA patients are known to have a higher incidence of difficult intubation, postoperative complications, increased intensive care unit admissions, and greater duration of hospital stay.^{11–13} Identifying patients with OSA is the first step in preventing postoperative complications due to OSA.

In-laboratory polysomnography is the accepted standard for diagnosing OSA.¹⁴ However, polysomnography is a time-consuming and costly procedure. Further, the growing awareness of sleep apnea has exacerbated the long waiting list in many sleep laboratories.¹⁵ To deal with this issue, a number of screening questionnaires and clinical screening models have been developed to help identify patients with OSA.^{16–25} However, a significant limitation to the aforementioned studies is that patients were preselected because most studies were conducted in the sleep laboratory setting.^{16,18,19,26} Furthermore, clinical models designed for OSA screening usually require the assistance of a computer and may not be suitable for clinical practice. One of the most widely

This article is featured in "This Month in Anesthesiology."
Please see this issue of ANESTHESIOLOGY, page 5A.

* Professor, Department of Anesthesia, # Professor, Department of Psychiatry, University of Toronto, Toronto Western Hospital, University Health Network. † Research Fellow, ‡ Research Data Analyst, || Research Coordinator, Department of Anesthesia, § Staff Scientist, Department of Psychiatry, Toronto Western Hospital, University Health Network.

Received from the Department of Anesthesia, University Health Network, University of Toronto, Toronto, Ontario, Canada. Submitted for publication August 28, 2007. Accepted for publication December 14, 2007. Supported by a grant from Physician Services Incorporated Foundation, Toronto, Ontario, Canada, and University Health Network Foundation, Toronto, Ontario, Canada.

Address correspondence to Dr. Chung: Room 405, 2McL, Department of Anesthesia, 399 Bathurst Street, Toronto, Ontario, Canada M5T 2S8. frances.chung@uhn.on.ca. This article may be accessed for personal use at no charge through the Journal Web site, www.anesthesiology.org.

used questionnaires, the Berlin questionnaire, has not been validated as a screening instrument in surgical patients. The American Society of Anesthesiologists (ASA) checklist, a screening instrument recommended by the ASA Task Force on Perioperative Management of Patients with Obstructive Sleep Apnea, has not been validated. Besides the fact that no predictive model or questionnaire for identifying OSA has been validated in the surgical patient population, most questionnaires have numerous items with a confusing scoring system. As a result, they are not suitable for a busy clinical setting, such as preoperative clinics.

The purpose of the study was to develop and validate a concise and easy-to-use questionnaire for OSA screening in surgical patients.

Materials and Methods

Patient Population of the Study

The study was conducted in the preoperative clinics of Toronto Western Hospital and Mount Sinai Hospital, Toronto, Ontario, Canada. Ethics approval was obtained from both institutions. Patients aged 18 yr or older who had an ASA physical status of I-IV and were scheduled to undergo elective procedures in general surgery, gynecology, orthopedics, urology, plastic surgery, ophthalmology, or neurosurgery were included in the study. Patients who were unwilling or unable to give informed consent, patients previously diagnosed with OSA or any other sleep breathing disorder, or patients who were expected to have abnormal electroencephalographic findings (e.g., brain tumor, epilepsy surgery, patients with deep brain stimulator) were excluded. All patients who visited the preoperative clinics for their scheduled surgery and met the inclusion criteria were approached by the research staff. After informed consent was obtained, patients were asked to complete a questionnaire and were invited to undergo an overnight polysomnographic study.

Development of the STOP Questionnaire

To keep the questionnaire concise and easy to use, the questions were designed in yes/no format. Based on our previous work with the Berlin questionnaire,⁸ consensus from a group of anesthesiologists and sleep specialists, and a literature review, four questions (STOP Q1-4) related to snoring, tiredness during the daytime, stopped breathing during sleep, and hypertension were designed. They were phrased in English at a fifth-grade reading level by using the Flesch-Kincaid reading-level determination method built into Microsoft Word (Microsoft, Redmond, WA).

To examine the association of the questions with the underlying constructs that the questions were designed to reflect, these four yes/no questions were combined with items 1-10 (Berlin Q1-10) from the Berlin ques-

tionnaire to make a question list consisting of 14 questions. The question list was administered to 278 patients to answer. Of these patients, 254 answered all of the questions. Factor analysis with the SAS procedure Factor was based on the responses from these 254 patients. After a significant level of association was demonstrated, these four yes/no questions were accepted to form the STOP questionnaire (appendix 1). The four-item STOP questionnaire is a self-report, forced-choice (yes/no), paper-and-pencil scale that takes approximately 1 min to complete. It consists of the following four questions: S—"Do you snore loudly (louder than talking or loud enough to be heard through closed doors)?" T—"Do you often feel tired, fatigued, or sleepy during daytime?" O—"Has anyone observed you stop breathing during your sleep?" P—"Do you have or are you being treated for high blood pressure?"

The STOP questionnaire was given to 592 preoperative clinic patients as a pilot study. All patients who answered the STOP questionnaire were invited to undergo an overnight, technician-supervised polysomnographic study. According to the interim analysis of the data from pilot study, the cutoff point of the STOP questionnaire had been decided and the sample size had also been adjusted.

To check the reliability of the questionnaire, 55 patients answered the STOP questionnaire twice at different time intervals of 1-27 days (median, 8 days). Because these four questions reflected four different dimensions of OSA morbidity, internal consistency checking was not applicable.

Validation of the STOP Questionnaire

After the pilot study, 1,875 patients were screened and asked to complete the STOP questionnaire. All patients, regardless of their score on the STOP questionnaire, were invited to undergo an overnight polysomnographic study. The data from patients who completed the polysomnographic study was used to validate the STOP questionnaire versus the apnea-hypopnea index (AHI) obtained from polysomnography in all patients of the validation group and in subgroups with different clinical characteristics—such as body mass index (BMI), age, neck circumference, and gender—were analyzed. An alternative scoring model incorporating BMI, age, neck circumference, and gender into the STOP questionnaire, termed the STOP-Bang (appendix 2), was also presented.

Sleep Study

A one-night, in-laboratory polysomnographic study was conducted before surgery at Toronto Western Hospital Sleep Laboratory. The result of polysomnography was used to evaluate the validity of the STOP questionnaire. During the overnight polysomnographic study, every patient went to bed at his or her usual bedtime.

Collection of continuous sleep architectural data was accomplished using a standard electroencephalographic montage consisting of an electroencephalogram, electrooculogram, submental electromyogram, and electrocardiogram using surface electrodes. Ancillary channels were used to specifically record respiratory parameters, including respiratory effort by thoracoabdominal excursion, respiratory inductive plethysmography, and oronasal airflow by nasal airflow pressure. Oxygen saturation was measured with a pulse oximeter.

One certified polysomnographic technologist with 10 yr of experience scored all of the polysomnographic recordings. Her scoring was under the supervision of a sleep physician (C.M.S.). The reports had to be assessed and approved by the sleep physician (C.M.S.). The certified technologist was blinded to the results of the STOP questionnaire (*i.e.*, whether patients were at high or low risk of having OSA) and clinical information of the patients. Sleep stages and the AHI were scored according to standard criteria.^{27,28} To avoid bias and inaccuracy from polysomnographic scoring, the polysomnographic recording of 10 randomly selected patients was rescored by another experienced certified polysomnographic technologist, who was blinded to the scores of other technologist. The scores from two technologists for the same patient were almost identical ($r = 0.984$, $P < 0.0001$).

The clinical diagnosis of OSA was defined as AHI greater than 5 with fragmented sleep and daytime sleepiness. According to the American Academy of Sleep Medicine practice guideline, the severity of OSA is determined by the AHI: 5–15, mild; greater than 15–30, moderate; greater than 30, severe.²⁷ After polysomnography, patients were scheduled to meet with a sleep specialist (C.M.S.) for follow-up assessment and clinical management, where necessary.

Data Analysis and Statistics

Sample Size Estimation. The calculation of sample size was performed according to the method reported by Obuchowski.²⁹ Briefly, the two separate calculations of sample size were performed based on either estimated sensitivity, the precision (potential error) of sensitivity, expected power, a type I error, and estimated prevalence; or specificity, the precision of specificity, expected power, type I error, and prevalence. The bigger number of the two is chosen as the sample size. Based on the literature on the Berlin questionnaire^{30–32} and the prevalence of OSA,^{33,34} a sensitivity of 0.88, a precision of 0.09, an OSA prevalence rate of 24%, a type I error of 0.05, and a power of 0.8 were used to calculate sample size. The result was 208. The number calculated based on a specificity of 0.80 was much smaller than 208. So 208 was initially chosen as the sample size. From the pilot study data, the sensitivity was 0.72 and the prevalence was 0.7. If the other parameters were kept the same, the sample size would be 137. However,

a prevalence of 0.7 is very high. It may be biased because of the small number of patients in the pilot study. If, for safety, 0.55 were taken as the prevalence, the adjusted sample size would be 170.

For factor analysis, the minimum requirement for sample size is the bigger of 100 respondents or 5 times the number of variables. In our study, we had 14 questions (variables), so we needed at least 100 complete respondents. The list of 14 questions was given to 278 patients; 254 patients who answered all of the questions were used for the factor analysis.

Data Analysis. Data were entered into a specifically designed Microsoft Access database and checked for possible errors. SAS 9.1 for Windows (SAS Institute, Cary, NC) was used for data analysis. Categorical data were presented as frequency and percentage with 95% confidence interval (CI). The statistical significance was checked by chi-square test or Fisher exact test. Resampling with bootstrap was used to calculate the CI of the likelihood ratios. A logistic regression procedure was used to calculate the odds ratio and area under the receiver operating characteristic curve. Continuous data were presented as mean \pm SD, and the Student *t* test or analysis of variance was used to calculate the *P* value. $P < 0.05$ was defined as significant. The SAS procedure Factor was used for factor analysis. The report from the principal components analysis with varimax rotations was presented. Factors with an eigenvalue greater than average were retained. Questions with factor loading of 0.3 or greater were chosen for interpretation of factors.

Results

Patient Screening

Over a period of 16 months at Toronto Western Hospital and Mount Sinai Hospital preoperative clinics, a total of 2,974 patients were willing to complete the questionnaire. Of these, 2,721 patients (91.5%) answered all of the items on the questionnaire completely and had complete documentation of gender, age, and BMI. Only these patients were included in the analysis.

Factor analysis was based on the response of 254 patients who answered all 14 questions from the STOP and Berlin questionnaires. After the STOP questionnaire was developed, it was administered to 2,467 patients. The STOP questionnaire classified 27.5% of patients (679 of 2,467) as being at high risk of having OSA. Of all patients who were invited to undergo the overnight monitored polysomnographic testing, 416 of 2,467 patients (17%) gave consent to participate. In total, 211 patients underwent polysomnography, whereas 205 did not show up at the laboratory (fig. 1). Of 211 patients who underwent polysomnography, the first 34 patients were included in a pilot study and the

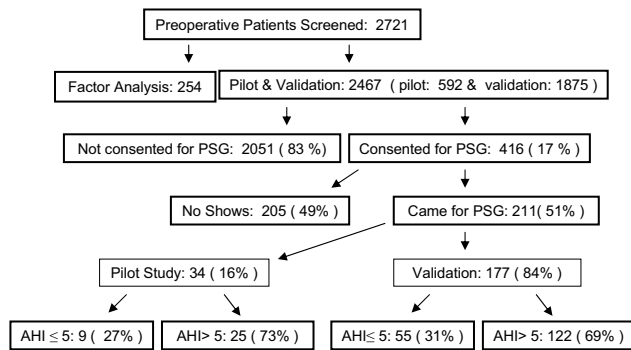


Fig. 1. Screening flow chart of surgical patients in preoperative clinic. The number in the figure shows the number and percentage of patients in the different groups. AHI = apnea-hypopnea index; PSG = polysomnography.

following 177 patients were for the validation of the STOP questionnaire.

Age, gender, and BMI of the different patient groups are shown in table 1. The patients who gave consent but did not actually undergo polysomnographic testing were younger than the group of patients who underwent the polysomnographic study. The BMI of patients who gave consent for polysomnography was significantly greater than that of the patients who did not give consent for polysomnography, regardless of whether the patient underwent polysomnographic testing. Compared with the patients who underwent polysomnographic testing, there was a higher rate of smoking in patients who gave consent but did not show up for the polysomnographic testing (26.8% vs. 14.7%; $P = 0.002$).

Development of the STOP Questionnaire

Four yes/no questions related to snoring, tiredness during the daytime, observed apnea during sleep, and hypertension were combined with 10 questions from the Berlin questionnaire and administered to 278 patients; 254 patients answered all of the questions. Demographic data are shown in table 1. Factor analysis demonstrated that four underlying factors were reflected by the 14 questions. These factors accounted for more than 95% of the total eigenvalue. Based on the category of questions with a loading factor greater than 0.3, four factors were identified: snoring, tiredness during daytime, cessation of breathing during sleep, and high blood pressure. The factor loading value for each question in

the corresponding category is shown in table 2. The factor loading value of two questions related to falling asleep while driving (Berlin Q8 and Q9) is very low for all four factors.

Among the five questions related to snoring, although question 1 (STOP Q1) did not have the highest factor loading value, it still demonstrated a significant association with snoring. Because we wanted to develop a simple and easy-to-use questionnaire with questions in yes/no format, we chose question 1 to reflect snoring in our questionnaire. Using a similar rationale, question 6 was incorporated to reflect daytime tiredness. Regarding the cessation of breathing during sleep and high blood pressure, two questions in each category had similar factor loading values, so questions 12 and 14 were acceptable choices to reflect breathing cessation during sleep and high blood pressure in the STOP questionnaire. The final STOP questionnaire consisted of four yes/no questions: 1, 6, 12, and 14 (appendix 1).

Pilot Study

As a pilot study, the STOP questionnaire was administered to 592 preoperative clinic patients, and all patients were invited to undergo polysomnography. Thirty-four of these patients underwent the polysomnography study. The other patients either declined to give consent or gave consent but did not show up. Of 34 patients, 24 (70.5%) had an AHI greater than 5. According to the analysis of data from these 34 patients, using answering yes to two or more questions as the cutoff for the STOP questionnaire to classify the patients as high or low risk of having OSA demonstrated the best combination of sensitivity and specificity. The sensitivity of the STOP questionnaire was 72% (CI, 54.4–89.6), the specificity was 33.3% (CI, 2.5–64.1), the positive predictive value (PPV) was 75.0% (CI, 57.7–92.3), and the negative predictive value (NPV) was 30% (CI, 6.7–65.3).

To check the test-retest agreement, 55 patients answered the STOP questionnaire twice at a time interval of 1–27 days (median: 8 days); 53 (96.4%) patients were found to have the same score upon retesting with a κ coefficient of 0.923 (CI, 0.82–1.00).

Table 1. Characteristics of Screened Patients

	Total (n = 2,721)	Factor Analysis (n = 254)	No Consent (n = 2,051)	Consented, No Polysomnography (n = 205)	Polysomnography Done (n = 211)
Gender, M/F	1,305/1,416	126/128	967/1,084	106/99	106/105
Age, yr	57 ± 16	56 ± 17	57 ± 16	54 ± 13*	56 ± 13
BMI, kg/m ²	28 ± 6	28 ± 6	28 ± 6	30 ± 8*	30 ± 7*

Continuous data are presented as mean ± SD.

* $P < 0.05$ compared with Factor Analysis and No Consent groups.

BMI = body mass index.

Table 2. Summary of the Principal Components Analysis, Varimax Rotation

	Factor Loadings*
Snoring	
1. (STOP Q1). Do you snore loudly (louder than talking or loud enough to be heard through closed doors)?	0.596
a. Yes	
b. No	
2. (Berlin Q1). Do you snore?	0.747
a. Yes	
b. No	
c. Don't know	
3. (Berlin Q2). Your snoring is:	0.825
a. Slightly louder than breathing	
b. As loud as talking	
c. Louder than talking	
d. Very loud—can be heard in adjacent rooms	
4. (Berlin Q3). How often do you snore?	0.795
a. Nearly every day	
b. 3–4 times a week	
c. 1–2 times a week	
d. 1–2 times a month	
e. Never or nearly never	
5. (Berlin Q4). Has your snoring ever bothered other people?	0.404
a. Yes	
b. No	
c. Don't know	
Tiredness during daytime	
6. (STOP Q2). Do you often feel tired, fatigued, or sleepy during daytime?	0.674
a. Yes	
b. No	
7. (Berlin Q6). How often do you feel tired or fatigued after your sleep?	0.805
a. Nearly every day	
b. 3–4 times a week	
c. 1–2 times a week	
d. 1–2 times a month	
e. Never or nearly never	
8. (Berlin Q7). During your waking time, do you feel tired, fatigued, or not up to par?	0.743
a. Nearly every day	
b. 3–4 times a week	
c. 1–2 times a week	
d. 1–2 times a month	
e. Never or nearly never	
Stop breathing during sleep	
11. (Berlin Q5). Has anyone noticed that you quit breathing during your sleep?	0.644
a. Nearly every day	
b. 3–4 times a week	
c. 1–2 times a week	
d. 1–2 times a month	
e. Never or nearly never	
12. (STOP Q4). Has anyone observed you stop breathing during your sleep?	0.606
a. Yes	
b. No	
High blood pressure	
13. (Berlin Q10). Do you have high blood pressure?	0.947
a. Yes	
b. No	
c. Don't know	
14. (STOP Q3). Do you have or are you being treated for high blood pressure?	0.945
a. Yes	
b. No	
Questions with low factor loading for all four factors	
9. (Berlin Q8). Have you ever nodded off or fallen asleep while driving a vehicle?	
a. Yes	
b. No	
10. (Berlin Q9). How often does nodding off or falling asleep while driving a vehicle occurs?	
a. Nearly every day	
b. 3–4 times a week	
c. 1–2 times a week	
d. 1–2 times a month	
e. Never or nearly never	

* Factor loadings are correlations between the original questions and their factors. Factor loadings greater than 0.30 in absolute value are considered to be significant.

Table 3. Demographic Data of Patients for Validation of STOP Questionnaire

	Total (n = 177)	STOP	
		Low Risk (n = 75)	High Risk (n = 102)
Gender (M/F)	88/89	38/37	50/52
Age, yr	55 ± 13	54 ± 15	56 ± 12
BMI, kg/m ²	30 ± 6	28 ± 6	31 ± 6*
BMI >35 kg/m ² , n	34	10	24
Neck circumference, cm	39 ± 6	38 ± 5	40 ± 7*
ASA physical status, n (%)			
I	11 (6)	8 (11)	3 (3)
II	101 (57)	47 (63)	54 (53)
III	63 (36)	18 (24)	45 (44)
IV	2 (1)	2 (3)	0
Average score	2.3 ± 0.6	2.2 ± 0.7	2.4 ± 0.6*
AHI	20 ± 6	12 ± 14	25 ± 27*
Minimum SaO ₂	82 ± 11	84 ± 9	80 ± 10*
Existing conditions, n (%)			
Hypertension	72 (41)	22 (29)	50 (49)*
GERD	56 (32)	13 (17)	43 (42)*
Diabetes	32 (18)	9 (12)	23 (23)
Asthma	24 (14)	9 (12)	15 (15)
Depression	11 (6)	5 (7)	6 (6)

Categorical data are presented as frequency (percentage), and continuous data are presented as mean ± SD.

* $P < 0.05$, high risk vs. low risk.

AHI = apnea-hypopnea index; ASA = American Society of Anesthesiologists; BMI = body mass index; GERD = gastroesophageal reflux disease; SaO₂ = arterial oxygen saturation.

Validation of the STOP Questionnaire

Demographic Data and Sleep Study. Table 3 shows the demographic data of the patients who participated in the validation study, *i.e.*, they completed the questionnaires and underwent polysomnography. The patients classified by the STOP questionnaire as being at high risk of having OSA had a significantly higher frequency of hypertension and gastroesophageal reflux disease. They

Table 4. Characteristics of Patients Grouped by AHI >5 Cutoff

	AHI ≤5 (n = 55)	AHI >5 (n = 122)
Gender, M/F	18/37	70/52*
Age, yr	49 ± 14	58 ± 12*
BMI, kg/m ²	27 ± 6	31 ± 6*
Blood pressure		
Systolic	129 ± 21	142 ± 18*
Diastolic	78 ± 12	83 ± 14*
ASA physical status, n (%)		
I	7 (13)	4 (3)
II	39 (71)	62 (51)
III	9 (16)	54 (44)
IV	0	2 (2)
Average score	2.0 ± 0.5	2.4 ± 0.6*
Neck circumference	36 ± 4	40 ± 6*
AHI	3 ± 2	27 ± 24*

Categorical data are presented as frequency (percentage), and continuous data are presented as mean ± SD.

* $P < 0.01$, apnea hypopnea index (AHI) >5 vs. AHI ≤5.

ASA = American Society of Anesthesiologists; BMI = body mass index.

Table 5. Sleep Parameters of Patients for Validation

	Total (n = 177)	AHI ≤5 (n = 55)	AHI >5 (n = 122)
Total sleep time, min	351 ± 73	356 ± 76	348 ± 71
Sleep efficiency, %	78 ± 14	80 ± 12	77 ± 15
Wake percent, %	18 ± 13	16 ± 12	19 ± 13†
REM latency, min	124 ± 78	111 ± 67	131 ± 82
REM percent, %	14 ± 8	15 ± 6	13 ± 8†
Sleep stage 1, %	9 ± 9	8 ± 8	10 ± 9
Sleep stage 2, %	49 ± 13	47 ± 13	49 ± 13
Slow wave sleep, %	10 ± 7	13 ± 8	9 ± 7*
AHI	19.5 ± 22.9	2.5 ± 1.5	27.2 ± 23.8*
REM AHI	27.6 ± 23.1	8.6 ± 8.5	35.9 ± 22.6*
Arousal index	29.4 ± 18.3	23.2 ± 16.7	32.1 ± 18.4*
Minimum SaO ₂	82 ± 11	87 ± 7	80 ± 11*

* $P < 0.05$, apnea-hypopnea index (AHI) >5 vs. AHI ≤5. † $P < 0.1$ vs. AHI ≤5.

REM = rapid eye movement; SaO₂ = arterial oxygen saturation.

also had significantly higher average ASA physical status, larger BMI, larger neck circumference, and higher AHI.

Using an AHI greater than 5 as the cutoff for diagnosis of OSA, 122 patients (68.9%) were found to have OSA, 52 (29.4%) mild, 31 (17.5%) moderate, and 39 (22.0%) severe. As shown in table 4, there were clear differences between patients with an AHI of 5 or less and patients with an AHI greater than 5. There was a higher percentage of male patients with an AHI greater than 5 (57% male vs. 43% female; $P < 0.01$). The patients with an AHI greater than 5 were almost more than 10 yr older than patients with an AHI of 5 or less. They also had significantly higher average ASA physical status and blood pressure, greater BMI, and larger neck size.

Table 5 summarizes the sleep parameters in validation patients. Compared with the patients with an AHI of 5 or less, the patients with an AHI greater than 5 demonstrated a significantly increased arousal index, significantly lower minimum arterial oxygen saturation, and significantly decreased slow wave sleep, which is consistent with the sleep features of the patients with OSA.

STOP Questionnaire Effectively Identified the Patients with OSA. Predictive parameters for the STOP questionnaire at AHI greater than 5, greater than 15, and greater than 30 cutoff values are presented in table 6. Using AHI greater than 5 as a cutoff value to evaluate the STOP questionnaire, the sensitivity was 65.6%, the specificity was 60.0%, the PPV was 78.4%, and the NPV was 44.0%. The sensitivity and NPV were 74.3% and 76.0% at AHI greater than 15. They were 79.5% and 89.3% with AHI greater than 30 as the cutoff. This indicates that the STOP questionnaire was more sensitive in detecting the patients with moderate to severe OSA.

Further examination of the predictive parameters of the STOP questionnaire in the different patient groups demonstrates that the PPV with AHI greater than 5 as the cutoff was greatly increased in patients with a certain demographics: BMI greater than 35 kg/m², age older

Table 6. Predictive Parameters for STOP Questionnaire (n = 177)

AHI >5	
Sensitivity, %	65.6 (56.4–73.9)
Specificity, %	60.0 (45.9–73.0)
PPV, %	78.4 (69.2–86.0)
NPV, %	44.0 (32.6–56.0)
Likelihood ratio	1.639 (1.172–2.385)
Odds ratio	2.857 (1.482–5.507)
Area under ROC curve	0.703
AHI >15	
Sensitivity, %	74.3 (62.4–84.0)
Specificity, %	53.3 (43.4–63.0)
PPV, %	51.0 (41.3–60.7)
NPV, %	76.0 (64.8–85.1)
Likelihood ratio	1.590 (1.280–2.057)
Odds ratio	3.293 (1.707–6.352)
Area under ROC curve	0.722
AHI >30	
Sensitivity, %	79.5 (63.5–90.7)
Specificity, %	48.6 (40.0–63.0)
PPV, %	30.4 (21.7–40.3)
NPV, %	89.3 (80.1–95.3)
Likelihood ratio	1.545 (1.261–2.010)
Odds ratio	3.656 (1.636–9.054)
Area under ROC curve	0.769

Data are presented as average (95% confidence interval).

AHI = apnea-hypopnea index; NPV = negative predictive value; PPV = positive predictive value; ROC = receiver operating characteristic.

than 50 yr, male gender, and neck circumference greater than 40 cm (fig. 2). The PPV of those ranked by the STOP questionnaire as being at high risk of having OSA was 84% in the patients with a BMI greater than 35 kg/m², 86.9% in patients older than 50 yr, 87.5% in male patients, 89.7% in male patients older than 50 yr, 94.3% in patients with neck circumference greater than 40 cm, and 100% in male patients older than 50 yr and with a BMI greater than 35 kg/m².

STOP-Bang, an Alternative Scoring Model Combining BMI, Age, Neck Circumference, and Gender with the STOP Questionnaire. To further improve the sensitivity of the STOP questionnaire to detect most

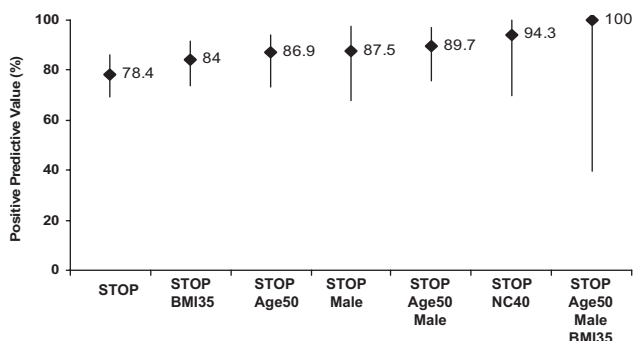


Fig. 2. Positive predictive value for STOP questionnaire in patients with the different clinical characteristics. The *y*-axis represents positive predictive value with 95% confidence interval, and the *x*-axis shows high risk of obstructive sleep apnea ranked by STOP questionnaire in patients with different clinical characteristics. BMI = body mass index; NC = neck circumference.

Table 7. Predictive Parameters for STOP-Bang (n = 177)

AHI >5	
Sensitivity, %	83.6 (75.8–89.7)
Specificity, %	56.4 (42.3–69.7)
PPV, %	81.0 (73.0–87.4)
NPV, %	60.8 (46.1–74.2)
Likelihood ratio	1.9160 (1.416–2.666)
Odds ratio	6.587 (3.217–13.489)
Area under ROC curve	0.806
AHI >15	
Sensitivity, %	92.9 (84.1–97.6)
Specificity, %	43.0 (33.5–52.9)
PPV, %	51.6 (42.5–60.6)
NPV, %	90.2 (78.6–96.7)
Likelihood ratio	1.629 (1.401–1.966)
Odds ratio	9.803 (3.654–26.300)
Area under ROC curve	0.782
AHI >30	
Sensitivity, %	100 (91.0–100.0)
Specificity, %	37.0 (28.9–45.6)
PPV, %	31.0 (23.0–39.8)
NPV, %	100 (93.0–100.0)
Likelihood ratio	1.586 (1.426–1.838)
Odds ratio	>999.999
Area under ROC curve	0.822

Data are presented as average (95% confidence interval).

AHI = apnea-hypopnea index; NPV = negative predictive value; PPV = positive predictive value; ROC = receiver operating characteristic.

patients with OSA, especially moderate and severe OSA, we developed an alternative scoring model, the STOP-Bang (appendix 2), which incorporated BMI, age, neck circumference, and gender into the scoring model of the STOP questionnaire. As shown in table 7, sensitivity and NPV are significantly increased by using the STOP-Bang. The sensitivities at AHI cutoffs of greater than 5, greater than 15, and greater than 30 were 83.6, 92.9, and 100%, respectively; the corresponding NPVs were 60.8, 90.2, and 100%.

Discussion

In this study, the STOP questionnaire was developed and validated as an OSA screening tool for surgical patients. The STOP questionnaire is a self-administered screening tool that includes four yes/no questions (appendix 1). The STOP questionnaire was found to have a moderately high sensitivity and PPV at AHI greater than 5, greater than 15, and greater than 30 as cutoffs. In patients with certain clinical characteristics, such as male gender, age older than 50 yr, BMI greater than 35 kg/m², and neck circumference greater than 40 cm, the PPV was greatly increased. When incorporating BMI, age, neck circumference, and gender into the STOP scoring (STOP-Bang), the sensitivity and NPV significantly increased. They were both more than 90% for the patients with moderate and severe OSA.

Obstructive sleep apnea is known to diminish quality of life³⁵ and is associated with many common comorbid

conditions. Studies have documented an increased incidence of coronary artery diseases, hypertension, cerebrovascular accidents, gastroesophageal reflux disease, congestive heart failure, and myocardial infarction in OSA patients.^{36,37} It is estimated that the average life span of an untreated OSA patient is 58 yr, which is 20 yr shorter than the average life span of the general population (men, 79 yr; women, 83 yr).³⁸ OSA is also associated with an increased incidence of postoperative adverse events.¹¹⁻¹³ Undiagnosed OSA in surgical patients have a serious impact on the postoperative outcome. Identifying patients with a high risk of OSA is the first step for the prevention of adverse health events, adverse perioperative outcomes, and its treatment. Screening tools work as a filter to separate the patients with a high risk of OSA from the patients with a low risk of OSA. A good screening tool should be validated in the target population against an accepted standard. It should be easy to use and have a high sensitivity and acceptable specificity.

Most screening tools for OSA so far have been validated in patients referred to sleep clinics or sleep laboratories. Seven predictive models, based on the different combinations of witnessed apneas, snoring, gasping, BMI, age, gender, and hypertension were developed and validated in the patients from sleep centers.^{16,18,19,21,23,24,39,40} The Sleep Disorders Questionnaire,⁴¹ Apnea Score,²⁵ and Global Sleep Assessment Questionnaire were all tested in patients mainly from sleep centers.⁴² Patients referred to sleep centers are suspected of having sleep-related disorders, especially OSA. They are preselected patients. Screening tools for OSA developed and validated in the sleep center patient population cannot be applied to other patient populations without validation in the target patient population.

The Berlin questionnaire is one of the few questionnaires that have been validated in primary care patients.³⁰ However, instead of monitored polysomnography in a sleep laboratory, home portable sleep monitoring was used for the validation of the Berlin questionnaire. Home portable sleep monitoring has not been accepted as the standard for the diagnosis of OSA. The STOP questionnaire is currently the only questionnaire developed and validated in surgical patients. Although there was some self-selection from the patients' perspective, our study was designed to include all surgical patients in our preoperative clinics regardless of their score of the STOP questionnaire to avoid selection biases.

In most previous studies, reports from monitored polysomnography were used to validate models or questionnaires. However, the staff performing the polysomnography and scoring the polysomnography were usually not blinded to the score on the questionnaire.^{18,19,23-25,39,40,42} This may have introduced bias into the scoring of polysomnography. In our study, in-laboratory polysomnography was used to evaluate the accuracy of the STOP

questionnaire. The staff performing and scoring the polysomnography was blinded to the score on the STOP questionnaire. This practice avoided bias during polysomnographic scoring.

Ease of use is also very important for a screening tool in busy clinical settings. Prediction models need calculation and computer assistance. Most widely used questionnaires have a long question list with a complicated scoring system. Although the questions are similar, the number of questions among the different OSA screening tools varies. For example, there are 11 multiple-choice questions organized into three categories on the Berlin questionnaire,³⁰ and 14 items under three categories on the OSA checklist, which is recommended by the ASA.⁴³ Study has shown that the response rate among patients decreases with increasing length of the questionnaire.⁴⁴ Four questions on the STOP questionnaire were designed in yes/no format, and it takes less than 1 min to finish. As a result, it had a high completion rate (91.5%) and test-retest agreement (96.4%). The STOP questionnaire is based on questions referring to snoring, tiredness/sleepiness, observed stop of breathing during sleep, and blood pressure. The alternative scoring model, the STOP-Bang, is based on eight items including four questions in STOP questionnaire, BMI, age, neck circumference, and gender. This creates the easy mnemonics STOP and STOP-Bang, which may serve as useful reminders for clinicians to use these instruments during the preoperative screening process.

To screen patients for a disease with an important health impact, a high sensitivity with an acceptable specificity is a basic requirement for a screening tool. The sensitivity and specificity of OSA screening tools have demonstrated considerable variability depending on the screening tool, the patient population, and the definition of OSA. For the prediction models tested in sleep center patients, the sensitivity varied from 76% to 96%, and the specificity ranged from 13% to 54%.^{18,19,40} For the questionnaire tested in sleep center patients, the sensitivity varied from 70% to 93%.^{25,41,42} The Berlin questionnaire is the most widely tested screening tool for OSA. The predictive parameters of the Berlin questionnaire largely varied in different patient populations. The sensitivity was 86% in primary care patients,³⁰ 62.5% in patients undergoing pulmonary rehabilitation, and 57-68% in sleep laboratory patients.⁴⁵ The wide variation in the sensitivity of the Berlin questionnaire also indicates the risk of transferring a screening tool between different patient populations without validation in the target patient population.

In terms of the predictive parameters, the STOP questionnaire itself demonstrated a moderately high level of sensitivity and specificity in surgical patients, and it was more sensitive to detect the patients with moderate to severe OSA. In the patients with certain clinical characteristics, such as male gender, age older than 50 yr, BMI greater than 35 kg/m², and neck circumference greater

than 40 cm, the high risk of OSA ranked by the STOP questionnaire could have a very high PPV for OSA (fig. 2). On the other hand, when incorporating BMI, age, neck circumference, and gender (Bang) into the STOP model (STOP-Bang); we could reach a very high level of sensitivity and NPV, especially for the moderate and severe OSA patients (table 6). Therefore, if a patient is ranked as low risk of OSA by the STOP-Bang scoring model, we would have a high confidence to exclude the possibility that the patient would have moderate to severe OSA.

This study has several limitations. In our study, the refusal rate for polysomnography was high. Self-selection from patients may exist because patients who had sleep symptoms might have selectively consented to the overnight polysomnography. The high refusal rate and dropout rate (49% of patients did not show up for their scheduled polysomnographic testing) also indicate the difficulty that the study faced. This may be due to the anxiety about surgery and the need to stay one night in the sleep laboratory. Other factors also played a role in patient refusal and dropout, e.g., smokers and younger patients tended not to show up for their scheduled overnight polysomnography. The high prevalence of OSA in the group of patients who underwent polysomnography may reflect this self-selection. Currently, this tool has only been tested in surgical (noncancer) patients. It needs to be validated in the other settings.

In conclusion, the STOP questionnaire is a concise and easy-to-use screening tool to identify patients with a high risk of OSA. It has been validated in surgical patients at preoperative clinics as a screening tool. The STOP-Bang scoring model, which incorporates BMI, age, neck size, and gender with the STOP questionnaire, has demonstrated a higher sensitivity and NPV, especially for patients with moderate to severe OSA.

The authors thank all of the anesthesiologists at Toronto Western Hospital, Toronto General Hospital, and Mount Sinai Hospital (Toronto, Ontario, Canada).

References

- Kryger MH: Diagnosis and management of sleep apnea syndrome. *Clin Cornerstone* 2000; 2:39-47
- Young T, Hutton R, Finn L, Badr S, Palta M: The gender bias in sleep apnea diagnosis: Are women missed because they have different symptoms? *Arch Intern Med* 1996; 156:2445-51
- Turkington PM, Sircar M, Allgar V, Elliott MW: Relationship between obstructive sleep apnoea, driving simulator performance, and risk of road traffic accidents. *Thorax* 2001; 56:800-5
- Shahar E, Whitney CW, Redline S, Lee ET, Newman AB, Javler F, George N, O'Connor T, Boland LL, Schwartz JE, Samet JM: Sleep-disordered breathing and cardiovascular disease: Cross-sectional results of the sleep heart health study. *Am J Respir Crit Care Med* 2001; 163:19-25
- Frey WC, Pilcher J: Obstructive sleep-related breathing disorders in patients evaluated for bariatric surgery. *Obes Surg* 2003; 13:676-83
- O'Keefe T, Patterson EJ: Evidence supporting routine polysomnography before bariatric surgery. *Obes Surg* 2004; 14:23-6
- Fidan H, Fidan F, Unlu M, Ela Y, Ibis A, Tetik L: Prevalence of sleep apnea in patients undergoing operation. *Sleep Breath* 2006; 10:161-5
- Chung F, Ward B, Ho J, Yuan H, Kayumov L, Shapiro C: Preoperative identification of sleep apnea risk in elective surgical patients, using the Berlin questionnaire. *J Clin Anesth* 2007; 19:130-4
- 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; 20:705-6
- Lofsky A: Sleep apnea and narcotic postoperative pain medication: A morbidity and mortality risk. *Anesth Patient Safety Found Newsletter* 2002; 17:24-5
- Gupta RM, Parvizi J, Hanssen AD, Gay PC: Postoperative complications in patients with obstructive sleep apnea syndrome undergoing hip or knee replacement: A case-control study. *Mayo Clin Proc* 2001; 76:897-905
- Gentil B, Delarminat JM, Boucherez C, Lienhart A: Difficult intubation and obstructive sleep-apnea syndrome. *Br J Anaesth* 1994; 72:368
- Liao P, Yegneswaran B, Vairavanathan S, Zaki A, Chung F: Respiratory complications among obstructive sleep apnea (OSA) patients who underwent surgery (abstract). *Sleep* 2007; 30 (suppl):0582
- Practice parameters for the indications for polysomnography and related procedures. Polysomnography Task Force, American Sleep Disorders Association Standards of Practice Committee. *Sleep* 1997; 20:406-22
- Flemons WW, Douglas NJ, Kuna ST, Rodenstein DO, Wheatley J: Access to diagnosis and treatment of patients with suspected sleep apnea. *Am J Respir Crit Care Med* 2004; 169:668-72
- Deegan PC, McNicholas WT: Predictive value of clinical features for the obstructive sleep apnoea syndrome. *Eur Respir J* 1996; 9:117-24
- Ali NJ, Davies RJO, Fleetham JA, Stradling JR: Periodic movements of the legs during sleep associated with rises in systemic blood pressure. *Sleep* 1991; 14:163-5
- Crocker BD, Olson LG, Saunders NA, Hensley MJ, McKeon JL, Allen KM, Gyulay SG: Estimation of the probability of disturbed breathing during sleep before a sleep study. *Am Rev Respir Dis* 1990; 142:14-8
- Viner S, Szalai JP, Hoffstein V: Are history and physical examination a good screening test for sleep apnea? *Ann Intern Med* 1991; 115:356-9
- Bliwise DL, Nekich JC, Dement WC: Relative validity of self-reported snoring as a symptom of sleep apnea in a sleep clinic population. *Chest* 1991; 99:600-8
- Hoffstein V, Szalai JP: Predictive value of clinical-features in diagnosing obstructive sleep-apnea. *Sleep* 1993; 16:118-22
- Kump K, Whalen C, Tishler PV, Browner I, Ferrette V, Strohl KP, Rosenberg C, Redline S: Assessment of the validity and utility of a sleep-symptom questionnaire. *Am J Respir Crit Care Med* 1994; 150:735-41
- Flemons WW, Whitelaw WA, Brant R, Remmers JE: Likelihood ratios for a sleep apnea clinical prediction rule. *Am J Respir Crit Care Med* 1994; 150:1279-85
- Dealberto MJ, Ferber C, Garma L, Lemoine P, Alperovitch A: Factors related to sleep apnea syndrome in sleep clinic patients. *Chest* 1994; 105:1753-8
- Kapuniai LE, Andrew DJ, Crowell DH, Pearce JW: Identifying sleep apnea from self-reports. *Sleep* 1988; 11:430-6
- Davies RJ, Ali NJ, Stradling JR: Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnoea syndrome. *Thorax* 1992; 47:101-5
- Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. *Sleep* 1999; 22:667-89
- Meoli AL, Casey KR, Clark RW, Coleman JA Jr, Fayle RW, Troell RJ, Iber C: Hypopnea in sleep-disordered breathing in adults. *Sleep* 2001; 24:469-70
- Obuchowski NA: Sample size calculations in studies of test accuracy. *Stat Methods Med Res* 1998; 7:371-92
- 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; 131:485-91
- Strauss RS, Browner WS: Risk for obstructive sleep apnea. *Ann Intern Med* 2000; 132:758-9
- Gami AS, Pressman G, Caples SM, Kanagala R, Gard JJ, Davison DE, Malouf JF, Ammass NM, Friedman PA, Somers VK: Association of atrial fibrillation and obstructive sleep apnea. *Circulation* 2004; 110:364-7
- Netzer NC, Hoegel JJ, Loube D, Netzer CM, Hay B, Alvarez-Sala R, Strohl KP: Prevalence of symptoms and risk of sleep apnea in primary care. *Chest* 2003; 124:1406-14
- Kushida CA, Nichols DA, Simon RD, Young T, Grauke JH, Britzmann JB, Hyde PR, Dement WC: Symptom-based prevalence of sleep disorders in an adult primary care population. *Sleep Breath* 2000; 4:9-14
- Bennett LS, Barbour C, Langford B, Stradling JR, Davies RJO: Health status in obstructive sleep apnea: Relationship with sleep fragmentation and daytime sleepiness, and effects of continuous positive airway pressure treatment. *Am J Respir Crit Care Med* 1999; 159:1884-90
- Dincer HE, O'Neill W: Deleterious effects of sleep-disordered breathing on the heart and vascular system. *Respiration* 2006; 73:124-30
- Ing AJ, Ngu MC, Breslin ABX: Obstructive sleep apnea and gastroesophageal reflux. *Am J Med* 2000; 108:120-5
- Young T, Finn L: Epidemiological insights into the public health burden of sleep disordered breathing: Sex differences in survival among sleep clinic patients. *Thorax* 1998; 53 (suppl 3):S16-19
- Maislin G, Pack AI, Kribbs NB, Smith PL, Schwartz AR, Kline LR, Schwab RJ, Dinges DF: A survey screen for prediction of apnea. *Sleep* 1995; 18:158-66

40. Rowley JA, Aboussouan LS, Badr MS: The use of clinical prediction formulas in the evaluation of obstructive sleep apnea. *Sleep* 2000; 23:929-38
41. Douglass AB, Bornstein R, Nino-Murcia G, Keenan S, Miles L, Zarccone VP Jr, Guilleminault C, Dement WC: The Sleep Disorders Questionnaire: I. Creation and multivariate structure of SDQ. *Sleep* 1994; 17:160-7
42. Roth T, Zammit G, Kushida C, Doghramji K, Mathias SD, Wong JM, Buysse DJ: A new questionnaire to detect sleep disorders. *Sleep Med* 2002; 3:99-108
43. Gross JB, Bachenberg KL, Benumof JL, Caplan RA, Connis RT, Cote CJ, Nickinovich DG, Prachand V, Ward DS, Weaver EM, Ydens L, Yu S: Practice guidelines for the perioperative management of patients with obstructive sleep apnea: A report by the American Society of Anesthesiologists Task Force on Perioperative Management of Patients with Obstructive Sleep Apnea. *ANESTHESIOLOGY* 2006; 104:1081-93
44. Iglesias C, Torgerson D: Does length of questionnaire matter? A randomised trial of response rates to a mailed questionnaire. *J Health Serv Res Policy* 2000; 5:219-21
45. 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

Appendix 1: STOP Questionnaire

Height ____ inches/cm Weight ____ lb/kg
 Age ____ Male/Female BMI ____
 Collar size of shirt: S, M, L, XL, or ____ inches/cm
 Neck circumference* ____ cm

1. Snoring
Do you snore loudly (louder than talking or loud enough to be heard through closed doors)?
Yes No
2. Tired
Do you often feel tired, fatigued, or sleepy during daytime?
Yes No
3. Observed
Has anyone observed you stop breathing during your sleep?
Yes No
4. Blood pressure
Do you have or are you being treated for high blood pressure?
Yes No

* Neck circumference is measured by staff.

High risk of OSA: answering yes to two or more questions

Low risk of OSA: answering yes to less than two questions

Appendix 2: STOP-Bang Scoring Model

1. Snoring
Do you snore loudly (louder than talking or loud enough to be heard through closed doors)?
Yes No
2. Tired
Do you often feel tired, fatigued, or sleepy during daytime?
Yes No
3. Observed
Has anyone observed you stop breathing during your sleep?
Yes No
4. Blood pressure
Do you have or are you being treated for high blood pressure?
Yes No
5. BMI
BMI more than 35 kg/m²?
Yes No
6. Age
Age over 50 yr old?
Yes No
7. Neck circumference
Neck circumference greater than 40 cm?
Yes No
8. Gender
Gender male?
Yes No

High risk of OSA: answering yes to three or more items

Low risk of OSA: answering yes to less than three items