A systematic review of screening questionnaires for obstructive sleep apnea

Une revue méthodique des questionnaires de dépistage de l'apnée obstructive du sommeil

Amir Abrishami, MD · Ali Khajehdehi, MD ·
Frances Chung, MD

Abstract

Purpose *Obstructive sleep apnea (OSA) may lead to life-threatening problems if it is left undiagnosed. Polysomnography is the "gold standard" for OSA diagnosis; however, it is expensive and not widely available. The objective of this systematic review is to identify and evaluate the available questionnaires for screening OSA.*

Source *We carried out a literature search through MEDLINE, EMBASE, and CINAHL to identify eligible studies. The methodological validity of each study was assessed using the Cochrane Methods Group's guideline.*

Principal findings *Ten studies (n = 1,484 patients) met the inclusion criteria. The Berlin questionnaire was the most common questionnaire (four studies) followed by the Wisconsin sleep questionnaire (two studies). Four studies were conducted exclusively on "sleep-disorder patients", and six studies were conducted on "patients without history of sleep disorders". For the first group, pooled sensitivity was 72.0% (95% confidence interval [CI]: 66.0-78.0%; $I^2 = 23.0\%$) and pooled specificity was 61.0% (95% CI: 55.0-67.0%; $I^2 = 43.8\%$). For the second group, pooled sensitivity was 77.0% (95% CI: 73.0-80.0%; $I^2 = 78.1\%$) and pooled specificity was 53.0% (95% CI: 50-57%; $I^2 = 88.8\%$). The risk of verification bias could not be eliminated in eight studies due to insufficient reporting. Studies on snoring, tiredness, observed apnea, and high blood pressure (STOP) and STOP including body mass index, age, neck circumference, gender (Bang) questionnaires had the highest methodological quality.*

Conclusion *The existing evidence regarding the accuracy of OSA questionnaires is associated with promising but inconsistent results. This inconsistency could be due to studies with heterogeneous design (population, questionnaire type, validity). STOP and STOP-Bang questionnaires for screening of OSA in the surgical population are suggested due to their higher methodological quality and easy-to-use features.*

Résumé

Objectif *L'apnée obstructive du sommeil (AOS) peut provoquer des problèmes de santé fatals si elle n'est pas diagnostiquée. La polysomnographie est «l'étalon or» du diagnostic de l'AOS; cependant, cette méthode est onéreuse et n'est pas disponible partout. L'objectif de cette revue méthodique était d'identifier et d'évaluer les questionnaires de dépistage de l'AOS existants.*

Source *Nous avons réalisé une recherche de la littérature dans les bases de données MEDLINE, EMBASE et CINAHL afin d'extraire les études admissibles. La validité méthodologique de chaque étude a été évaluée sur la base de la directive du Groupe de méthode de Cochrane.*

Constatations principales *Dix études (n = 1484 patients) satisfaisaient aux critères d'inclusion. Le questionnaire de Berlin était le questionnaire le plus utilisé (quatre études), suivi par le questionnaire sur le sommeil de Wisconsin (deux études). Quatre études ont été menées exclusivement auprès de «patients avec troubles du sommeil», et six auprès de «patients sans antécédents de troubles du sommeil». Dans le premier groupe, la sensibilité pondérée était de 72,0% (intervalle de confiance [IC] 95%: 66,0-78,0%; $I^2 = 23,0\%$) et la spécificité pondérée de 61,0% (IC95%: 55,0-67,0%; $I^2 = 43,8\%$). Dans le deuxième groupe, la sensibilité pondérée était de 77,0 % ([IC] 95%: 73,0-80,0%; $I^2 = 78,1\%$) et la*

A. Abrishami, MD · A. Khajehdehi, MD · F. Chung, MD (✉)
Department of Anesthesia, Toronto Western Hospital, University
Health Network, University of Toronto, 2-405 McL Wing,
399 Bathurst Street, Toronto, ON M5T 2S8, Canada
e-mail: frances.chung@uhn.on.ca

spécificité pondérée de 53,0% (IC95%: 50-57%; I^2 = 88,8%). Le risque de biais de vérification n'a pas pu être éliminé dans huit des études en raison de présentation insuffisante des données. Les études sur le ronflement, la fatigue, l'apnée observée et une hypertension artérielle (études dites STOP pour l'acronyme anglais) et les questionnaires STOP incluant l'indice de masse corporelle, l'âge, la circonférence du cou, et le sexe (études dites Bang pour l'acronyme anglais) ont démontré la meilleure qualité méthodologique.

Conclusion *Les données probantes existantes concernant l'exactitude des questionnaires sur l'AOS sont associées à des résultats prometteurs mais peu constants. Ce manque de constance pourrait être lié à la conception hétéroclite des études (population, type de questionnaire, validité). Les questionnaires STOP et STOP-Bang sont suggérés pour dépister l'AOS chez les patients chirurgicaux en raison de leur qualité méthodologique supérieure et de leur facilité d'emploi.*

Obstructive sleep apnea (OSA) is a significant medical problem affecting at least 2-26% of the general population.[1] It is estimated that up to 93% of women and 82% of men with moderate to severe OSA remain undiagnosed.[2] Obstructive sleep apnea is independently associated with an increased likelihood of hypertension, cardiovascular disease, and diminished quality of life.[3-7]

The "gold standard" for diagnosis of OSA is laboratory polysomnography (PSG); however, the occurrence of OSA is far more prevalent than can be handled by the available sleep laboratories. Therefore, a screening tool is necessary to stratify patients based on their clinical symptoms, their physical examinations, and their risk factors, in order to ascertain patients at high risk and in urgent need of PSG and/or further treatment and patients at low risk who may not need PSG.

Previous investigators have developed different diagnostic models for the clinical prediction of OSA. Rowley et al.[8] prospectively studied the utility of four clinical prediction models (Crocker,[9] Viner,[10] Flemons,[11] and Maislin[12]) and concluded that they are not sufficiently accurate to discriminate between patients with or without OSA. In addition, some of these clinical models require the assistance of a computer and sophisticated mathematical calculations.

In contrast to clinical diagnostic models, OSA questionnaires do not require complicated calculations to identify high-risk patients, and they are potentially easier for routine clinical applications. This systematic review aims to evaluate and compare the accuracy of existing questionnaires as screening instruments for OSA in adults.

Methods

This systematic review was carried out using the recommended methods established by the Cochrane Methods Group on Screening and Diagnostic Tests[A] and by other authors.[13,14]

Literature search

In order to include all available evidence, a systematic search of the literature was carried out through the Cochrane Library, MEDLINE (from 1950 to April 2009), EMBASE (from 1980 to April 2009), and CINAHL (from 1990 to 2009) using the search strategy that was designed for each database. The search strategy was developed and executed by an expert librarian and included the following free-text and index terms: "obstructive sleep apnoea or apnea", "hypopnea or hypopnoea", "OSA or SHS or OSAHS", "sleep related respiratory disorder", "sleep disordered breathing", "Sleep Apnea Syndromes", "Risk Assessment", "Mass Screening", "validation studies", "questionnaire", "sensitivity", "specificity", "screen", "risk", "score or scale", and "mass screening" (Appendix). The search was extended to checking the reference lists of the included papers.

The search results were evaluated by two independent reviewers (A.A., A.K.) to find the eligible articles for inclusion. First, obviously irrelevant items were excluded by reviewing the title and/or abstract of the records. Next, the full-text articles of the remaining papers were retrieved and carefully evaluated to determine if they met the following eligibility criteria: 1) The study used a patient-based questionnaire as a screening tool for OSA in adult subjects ($\geq$18 yr); 2) The questionnaire's accuracy was evaluated by comparing its results with the results of a PSG as the "gold standard"[15] for diagnosing OSA; 3) OSA was clearly defined as apnea/hypopnea index (AHI), apnea index (AI), or respiratory disturbance index (RDI) $\geq$ 5; 4) Information was adequately presented to allow the construction of a 2 by 2 contingency table; 5) The questionnaire and full text paper were written in English. The studies that were found ineligible and excluded from our study are listed in Table 2.

Assessment of methodological quality

The methodological quality of each paper was assessed independently by the authors (A.A., A.K.), and

---

[A] Cochrane Methods Group on Screening and Diagnostic Tests, 2007. http//www.cochrane.org/cochrane/sadt.html.

disagreements were resolved by arbitration of the senior author (F.C.). Validity criteria assessing internal and external validity were explicitly described and coded according to Cochrane Methods Group on Screening and Diagnostic Tests.[A] Internal validity included the following factors: study design, definition of the disease, blind execution of the index test (questionnaire) and the reference test (polysomnography), valid reference test, avoidance of verification bias, independent interpretation of test results. External validity consisted of the following items: disease spectrum, clinical setting, demographic information, previous screening or referral filter, explicit cut-offs, percentage of missing patients, missing data management, subject selection for polysomnography.

Data extraction and analysis

Data were extracted by two reviewers (A.A., A.K.) independently using standard data collection forms. In each study, the true positive, false positive, true negative, and false negative values were extracted for each AI, AHI, or RDI cut-off, and 2 by 2 contingency tables were constructed accordingly. The AI/AHI or RDI ≥ 5 were considered as diagnosis cut-offs for the existence of OSA. The AI/AHI or RDI ≥ 15 and 30 were considered as diagnosis cut-offs for moderate and severe OSA, respectively. Using the 2 by 2 contingency tables, we recalculated the following predictive parameters in each study: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and prevalence of OSA in each study. Results were not extracted that would require extrapolations from equations, graphs, or derivations from figures or tables. Studies were excluded from the review if there was inadequate information to draw the 2 by 2 contingency tables (seven papers).[11,16-20]

A validated computer program for meta-analysis of test accuracy data (Meta-DiSc,[21] version 1.4, Hospital Ramony Cajal, Madrid, Spain) was used to describe the overall accuracy of the questionnaires and to assess inconsistencies in accuracy parameters (sensitivity and specificity) across studies (heterogeneity). Accuracy parameters with a similar target population were analyzed together (sleep-disorder patients *vs* patients without history of sleep disorders). Inconsistency (I^2) > 50% was considered as significant. Subgroup analysis was carried out on studies using the same questionnaire to explore the reasons of heterogeneity. Meta-analysis was not carried out on other predictive parameters (PPV and NPV), as they are not related to the intrinsic quality of the questionnaires. However, all parameters were presented in the review for descriptive analysis.

Results

Our extensive search strategy yielded 4,105 citations (Figure 1). After screening, most studies were eliminated based on the eligibility criteria, and only ten papers were considered for final inclusion in the review. The characteristics of the included studies are summarized in Table 1. The excluded studies[11,16-20,22-24] and the reasons for their exclusion are listed in Table 2.

Subject characteristics

A total of 1,484 participants were in the included studies, and 350 of them were "patients with sleep-related disorders", i.e., patients in the sleep clinics or those with habitual snoring. Weatherwax 2003, who studied the validity of a questionnaire on epileptic patients,[25] was considered in this category, as up to one-third of epilepsy patients could have coexisting OSA.[25] Studies on "patients without history of sleep disorders" included a total of 1,134 subjects (Table 1). In this group, Sharma 2006a enrolled patients using a pre-screening questionnaire that included questions related to OSA risk factors. The sample size for all studies ranged from 42 to 602 patients. The mean age of the subjects in the studies varied from 42 to 55 yr. The male ratio and body mass index (BMI) ranged from 45.8-79.3% and from 24.6-30.2, respectively. Four studies were
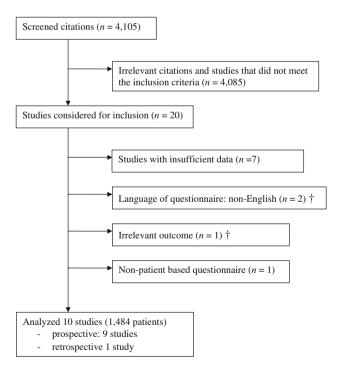

Fig. 1 Flow chart of screened, excluded, and analyzed papers. †The total of the excluded studies is not the overall sum, because one study was excluded due to an irrelevant outcome and a questionnaire that was not written in English

Table 1 Characteristics of the included studies

Study ID	Questionnaire	No. of patients	Validation process	Type of patients	Age (Mean ± SD)	Male (%)	BMI (Kg·m^{-2})	Validation tool	OSA definition
Ahmadi 2008	Berlin questionnaire	130	Retrospective	Sleep clinic patients	42‡	54.0%	27.9‡	Two-overnight sleep lab PSG	RDI ≥ 5, 10
Chung 2008a	STOP questionnaire STOP-Bang questionnaire	177†	Prospective	Surgical patients	55 ± 13	49.8%	30 ± 6	One-overnight sleep lab PSG	AHI ≥ 5
Chung 2008b	Berlin questionnaire ASA checklist STOP questionnaire	211†	Prospective	Surgical patients	55 ± 13	49.8%	30 ± 6	One-overnight sleep lab PSG	AHI ≥ 5
Sharma 2006a	Berlin questionnaire	104	Prospective	Pre-screened patients in outpatient tertiary centre	40 ± 9	80.0%	28.0 ± 6.2	One-overnight sleep lab PSG	AHI ≥ 5
Sharma 2006b	Wisconsin questionnaire	151	Prospective	General population (electoral database)	42 ± 9	59.0%	24.6 ± 5.2	One-overnight sleep lab PSG	AHI ≥ 5
Weatherwax 2003	Sleep Apnea of Sleep Disorders Questionnaire SA-SDQ	125	Prospective	Epileptic patients	–	56.0%	–	One-overnight sleep lab PSG	AHI ≥ 5
Netzer 1999	Berlin questionnaire	100	Prospective	Primary care setting	49 ± 17	45.8%	29.0 ± 7.2	Portable unattended overnight PSG	RDI ≥ 5, 10
Young 1993	Wisconsin questionnaire	602	Prospective	General population (state employees)	–	58.5%	–	One-overnight sleep lab PSG	AHI ≥ 5
Haraldsson 1992	No-name	42	Prospective	Habitual snorers	–	–	–	One-overnight sleep lab PSG	AI ≥ 5
Kapuniani 1988	Apnea Score (developed from Hawaii Sleep Questionnaire)	53	Prospective	Sleep clinic patients	46 ± 14	79.3%	–	One-overnight sleep lab PSG	AHI ≥ 5

BMI = Body mass index; ASA = American Society of Anesthesiologists; STOP = snoring, tiredness, observed apnea, and high blood pressure; STOP-Bang = STOP questionnaire including BMI, age, neck circumference, gender; SA-SDQ = Sleep Apnea scale of the Sleep Disorders Questionnaire; PSG = polysomnography; RDI = Respiratory Disturbance Index; AHI = Apnea/Hypopnea Index; lab = laboratory; † = In these two studies, the same patients were used to validate the STOP (or STOP-Bang) questionnaire. ‡The standard deviations for age and BMI were not reported in the paper

Table 2 Excluded studies

Study ID	Summary	Reason of exclusion
Izci 2008	This study is a Turkish version of ESS that was validated for prediction of OSA	This study was excluded because English was not the language of the questionnaire.
Onen 2008	This study measures the ability of the Observation-based Nocturnal Sleep Inventory (ONSI) to detect the presence or absence of sleep apnea syndrome (SAS) in older adults.	The questionnaire was designed to be answered by the patient's caregiver, e.g., a nurse, and it is not a patient-based questionnaire
Rosenthal 2008	This study evaluates the accuracy of the Epworth Sleepiness Scale in identification of OSA	- The study provided specificity and sensitivity for different ESS values; however; the overall scoring was not developed in the study. The overall accuracy of the questionnaire could not be evaluated in this study.
Mazeika 2005	The objective of this study is to develop and validate the G.A.S.P. questionnaire, which consists of questions regarding snoring, witnessed apnea, fatigue or sleepiness, hypertension or reflux, and being overweight.	- The study was published as an abstract and meeting presentation only. The data were insufficient to draw a two-by-two contingency table (true positives, true negatives, false positives, and false negatives) from the accuracy results.
Teculescu 2003	This study evaluates the reliability of a French version of the Wisconsin Sleep Questionnaire designed to investigate snoring, obstructive apnea, and sleeping problems. The assessment of reliability included the study of internal consistency and the three months repeatability of the questionnaire.	This study was excluded because English was not the language of the questionnaire. The outcome of the study is only the reliability of the questionnaire, and the study does not provide the accuracy results.
Gurubhagavatula 2001	This study evaluates the results of an algorithm including the results of the Multivariable Apnea Prediction (MAP) questionnaire along with nocturnal pulse oximetry to predict OSA. The MAP is a questionnaire containing different questions regarding snoring, sleep breathing problems, and daytime sleepiness.	- The results of the questionnaire were incorporated into a predictive model (a formula) to predict the presence of OSA, so the independent accuracy of the questionnaire could not be evaluated. - Predictive models and formulas (with mathematical calculations) are not in the scope of our review.
Pouliot 1997	This study combines the results of the Epworth Sleepiness Scale (ESS), BMI, and whether apnea had been observed by the patient's partner in predicting OSA.	- The relative accuracy of each factor or the combination of the factors was provided in the paper; however; the overall scoring was not developed in the study. The overall accuracy of the questionnaire could not be evaluated in this study.
Flemons 1994	This study includes different clinical factors, such as BMI and neck circumference, along with the results of questionnaire for predicting OSA. The questionnaire items are as follows: collar size, habitual snoring, disruptive snoring, nocturnal gasping, blood pressure, and weight gain.	- The relative likelihood ratio for different scores was provided in the paper; however, the cut-off score (for predicting OSA) and its respective accuracy results (sensitivity or specificity) were not provided in the paper. The overall accuracy of the questionnaire could not be evaluated.
Douglass 1994	The objective of this study is to develop and validate the Sleep Disorders Questions (SDQ). The questionnaire included items such as age, BMI, weight, number of years as a smoker, loud snoring, aggravating factor of snoring, stop breathing during sleep, sweating at night, high blood pressure.	- There were not sufficient data in the paper to draw a two-by-two contingency table (true positives, true negatives, false positives, and false negatives) from the accuracy results.
Bilwise 1991	The study valuates the relative accuracy of each of the following questions in predicting sleep apnea. 1. How often during the night do you snore in any way? 2. How often during the night do you snore loudly and disruptively? 3. How often during the night do you hold your breath or stop breathing during sleep?	- The overall accuracy of all questions was not evaluated. - The data were insufficient to draw a two-by-two contingency table (true positives, true negatives, false positives, and false negatives) from the accuracy results.

OSA = obstructive sleep apnea; BMI = body mass index

performed in the USA,[25,27-29] three in Canada,[30-32] two in India,[33,34] and one in Europe.[35] Nine studies were prospective cohort, while one study[30] was a retrospective chart review of patients who had already undergone PSG in a sleep laboratory (Table 1).

The questionnaire characteristics

There were eight questionnaires developed and/or validated in the included studies. The Berlin questionnaire was the most common questionnaire among the studies (four
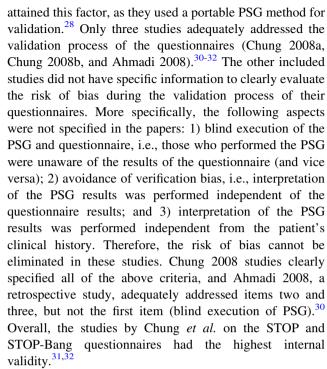
studies), followed by the Wisconsin sleep questionnaire (two studies). Apnea score (AS), Haraldsson's questionnaire, and the Sleep Apnea scale of the Sleep Disorders Questionnaire (SA-SDQ) were validated in sleep-disorder patients only. The latter was a modified version of SA-SDQ, which was validated on epileptic patients. The original SA-SDQ questionnaire was developed by Douglas et al.;[17] however, their study could not be included in this review due to insufficient data regarding the accuracy parameters (Table 2). The checklist of the American Society of Anesthesiologists (ASA), the STOP questionnaire (snoring, tiredness, observed apnea, and high blood pressure), and the STOP-Bang questionnaire (STOP including BMI, age, neck circumference, gender) were validated on surgical patients. The Wisconsin questionnaire was validated on patients selected from the general population, and the Berlin questionnaire was validated in differing populations, i.e., sleep clinic patients, general population, and surgical patients.

The details of each questionnaire and its scoring method are shown in Table 3. All eight questionnaires used loud snoring (or snoring) and stop breathing during sleep as two of the components of their questionnaires. In five questionnaires, daytime sleepiness was one of the questions. Three questionnaires noted body mass index and high blood pressure. A history of adenoidectomy or anatomical problems of airways was considered in three questionnaires, and measurement of neck circumference was used in two questionnaires. The summary of items used in the OSA questionnaires is shown in Table 4. Three questionnaires (ASA checklist, STOP, and STOP-Bang) used a yes/no format; however, others had Likert-type (frequency) questions. The number of questions in each questionnaire ranged from three to 12 items.

The validation tool was one overnight sleep laboratory PSG in eight studies, except for Netzer's study of the Berlin questionnaire. In Netzer's study, the Berlin questionnaire was validated by the patient taking the ambulatory PSG recorder home. The patient was given instructions on how to use the recording device and was told to turn on the device at bedtime and to turn it off upon arising. One retrospective chart review on the Berlin questionnaire used the result of two consecutive nights of PSG as the "gold standard". In seven studies, the AHI $\geq$ 5 was used as the PSG cut-off score for OSA diagnosis; AI $\geq$ 5 was used by one study, and RDI $\geq$ 5 or 10 was used by two studies.

Methodological quality of the included studies

In terms of the internal validity, all of the included studies used a valid reference test to verify the accuracy of the questionnaires (Table 5). Netzer 1999, however, partially attained this factor, as they used a portable PSG method for validation.[28] Only three studies adequately addressed the validation process of the questionnaires (Chung 2008a, Chung 2008b, and Ahmadi 2008).[30-32] The other included studies did not have specific information to clearly evaluate the risk of bias during the validation process of their questionnaires. More specifically, the following aspects were not specified in the papers: 1) blind execution of the PSG and questionnaire, i.e., those who performed the PSG were unaware of the results of the questionnaire (and vice versa); 2) avoidance of verification bias, i.e., interpretation of the PSG results was performed independent of the questionnaire results; and 3) interpretation of the PSG results was performed independent from the patient's clinical history. Therefore, the risk of bias cannot be eliminated in these studies. Chung 2008 studies clearly specified all of the above criteria, and Ahmadi 2008, a retrospective study, adequately addressed items two and three, but not the first item (blind execution of PSG).[30] Overall, the studies by Chung et al. on the STOP and STOP-Bang questionnaires had the highest internal validity.[31,32]

In terms of the external validity (generalizability), most of the studies met the appraisal items adequately (Table 5). In Sharma 2006a, a pre-screening set of questions was used to select the subjects out of the general population;[33] therefore, this study is missing one of the important components of the external validity and is at risk of screening bias, i.e., those who were selected for the validation of the screening questionnaire already had some type of sleep-related problems, and they do not necessarily represent the target population of the study. Another important aspect of the external validity was the management of the missing data, which was only carried out in three studies (Chung 2008a, Chung 2008b and Young 1993).[29,31,32] Chung et al. and Young et al. compared the basic characteristics (age, sex, BMI) of those patients who agreed to undergo PSG vs those who refused. Chung et al. showed a significant difference between the two groups, i.e., the PSG group had a higher BMI (30 $\pm$ 7 vs 28 $\pm$ 6; $P < 0.05$). Other characteristics in both studies were similar between the two groups.

Results of accuracy outcomes and other predictive parameters

In studies on "sleep-disorder patients", the prevalence of OSA (AHI/AI or RDI $\geq$ 5) ranged from 42-76%. The sensitivity of different questionnaires in predicting OSA ranged from 59-81%, with Haraldsson's questionnaire showing the highest sensitivity. The pooled sensitivity was 72.0% (95% CI: 66.0-78.0%; $I^2 = 23.0\%$). In this category, the specificity ranged from 46-80%, with

Table 3 Characteristics of the included questionnaires

Questionnaire -study(s)	Questions	Answers	Scoring
Berlin Questionnaire -Ahmadi 2008 -Chung 2008b -Sharma 2006a	**Category 1** 1- Do you snore? 2- If yes, how much loud is it? 3- How often do you snore? 4- Has your snoring ever bothered other people? 5- Has anyone noticed that you quit breathing during your sleep? **Category 2** 6- How often do you feel tired or fatigued after your sleep? 7- During your waking time, do you feel tired, fatigued, or not up to par? 8- Have you ever nodded off or fallen asleep while driving a vehicle? 9- If yes, How often does this occur? **Category 3** 10- Do you have high blood pressure?	**Items 1, 4, 8, 10 :** yes (1), no (0), don't know (0) **Item 2 :** slightly louder than breathing (0) as loud as talking (0) louder than talking (1) very loud -can be heard in adjacent rooms (1) **Item 5, 6, 7, 9 :** nearly every day (1) 3-4 times a week (1) 1-2 times a week (0) 1-2 times a month (0) Never or nearly never (0)	*Category 1*is positive if the total score is ≤2 points *Category 2* is positive if the total score is ≤2 points *Category 3* is positive if the answer to item 10 is yes or if the BMI of the patient is greater than 30 kg/m2. *High risk of OSA:* two or more categories scored as positive *Low risk of OSA:* only one or no category scored as positive
STOP questionnaire -Chung 2008a	1- Do you snore loudly (louder than talking or loud enough to be heard through closed doors)? 2- Do you often feel *tired*, fatigued, or sleepy during daytime? 3- Has anyone *observed* you stop breathing during your sleep? 4- Do you have or are you being treated for high blood pressure?	- all items: yes /no	*High risk of OSA:* answering yes to two or more questions *Low risk of OSA:* answering yes to less than two questions
STOP Bang questionnaire -Chung 2008a	1- Do you snore loudly (louder than talking or loud enough to be heard through closed doors)? 2- Do you often feel *tired*, fatigued, or sleepy during daytime? 3- Has anyone *observed* you stop breathing during your sleep? 4- Do you have or are you being treated for high blood pressure? 5- BMI ≥ 35 kg/m2? 6- Age ≥ 50 yr old? 7- Neck circumference ≥ 40 cm? 8. Gender male?	- all items: yes /no	*High risk of OSA:* answering yes to three or more items *Low risk of OSA:* answering yes to less than three items

Table 3 continued

Questionnaire -study(s)	Questions	Answers	Scoring
ASA checklist -Chung 2008b	*Category 1: Predisposing Physical Characteristics*	- all items: yes /no	If $\lesssim$-items in category 1 are positive, category 1 is positive.
	1 - BMI $\lesssim$ 35 kg/m2		If $\lesssim$-items in category 2 are positive, category 2 is positive.
	2 - Neck circumference > 43 cm/17 inches (men) or 40 cm/16 inches (women)		If $\lesssim$+items in category 3 are positive, category 3 is positive.
	3 - Craniofacial abnormalities affecting the airway		*High risk of OSA:* two or more categories scored as positive
	4 - Anatomical nasal obstruction		*Low risk of OSA:* only one or no category scored as positive
	5 - Tonsils nearly touching or touching the midline		
	Category 2: History of Apparent Airway		
	4 - Snoring (loud enough to be heard through closed door)		
	5 - Frequent snoring		
	6 - Observed pauses in breathing during sleep		
	7 - Awakens from sleep with choking sensation		
	8 - Frequent arousals from sleep		
	Category 3: Somnolence		
	9 - Frequent somnolence or fatigue despite adequate "sleep"		
	10 - Falls asleep easily in a nonstimulating environment (*e.g.*, watching		
	11- [Parent or teacher comments that child appears sleepy during the day, is easily distracted, is overly aggressive, or has difficulty concentrating]*		
	12 - [Child often difficult to arouse at usual awakening time]*		
	* Items in brackets refer to pediatric patients		
Wisconsin questionnaire† - Sharma 2006b - Young 1993	1 - habitual snoring †	- N/A †	- the questionnaire identifies habitual snorer vs. not-habitual snorer †
	2-extremely loud snoring †		
	3 - breathing pauses during the sleep †		

Table 3 continued

Questionnaire -study(s)	Questions	Answers	Scoring
SA-SDQ - Weatherwax 2003	1- 'I am told I snore loudly or bother others.' 2- 'I am told I stop breathing at night.' 3- 'I awake suddenly gasping for breath, unable to breathe.' 4- 'I sweat a great deal at night.' 5- 'I have high blood pressure (or once had it)' 6- 'I have a problem with my nose blocking up when I want to sleep' 7- 'My snoring/breathing problem is much worse if I sleep on my back' 8- 'My snoring/breathing problem is much worse if I'm asleep right after drinking alcohol' 9- Current weight in pounds 10- How many years as a smoker? 11- Current age 12- Body mass index	**Items 1-8 :** - never (1), rarely (2), sometimes (3), usually (4), always (5) **Item 9:** $\leq$134 pounds (1), $\leq$159 pounds (2),$\leq$183 pounds (3) $\leq$209 pounds (4), $\geq$210 pounds (5) **Item 10:** 0 years (1), 1 year (2), $\leq$12 years (3), $\leq$25 years (4), $\geq$26 years (5) **Item 11:** $\leq$25 years (1), $\leq$35 years (2), $\leq$44 years (3), $\leq$50 years (4), $\geq$51 years (5) **Item 12:** $\leq$22.1 (1), $\leq$25.6 (2), $\leq$27.4 (3), $\leq$30.9 (4), $\leq$31 (5)	OSA if total score $\geq$ 29 (male) or 26 (female)
Haraldsson's questionnaire -Haraldsson1992	1- Do you suffer from snoring? 2- Have you been told of repeated breath cessations during sleep ? 3- Do you suffer from mid-sleep awakening? 4- Do have difficulty initially falling sleep, or falling sleep after mid-sleep awakening ? 5- Are you so tired that you suffer from involuntary day-time sleep spells?	- all items: Don't know (0), hardly ever or never (1), occasionally (2), often (3), incessantly (4)	OSA if score $\geq$ 3 in each question
Apnea Score -Kapuniani1988	1- Frequency of stop breathing during sleep 2- Frequency of loud snoring 3- History of adenoidectomy	**Items 1-2 :** never (0), rarely (0), sometimes (0), often (1), always (1) **Item 3:** yes (1)/no (0)	OSA is *probable* if total score $\geq \times$

BMI = Body mass index, OSA: obstructive sleep apnea, ASA = American Society of Anesthesiologists, STOP: snoring, tiredness, observed apnea and high blood pressure, STOP-Bang = STOP questionnaire including BMI, age, neck circumference, gender, SA-SDQ: Sleep Apnea scale of the Sleep Disorders Questionnaire. †: the details regarding the questions and scoring was not provided in the papers

Table 4 Items used in different OSA questionnaires

Questionnaire	Age	BMI	Male	↑BP	Neck size	Snoring	loud snoring	Apnea during sleep	Somnolence/tiredness during day	Anatomical problems in airways	Others
Berlin				✔		✔	✔	✔	✔		
STOP				✔			✔	✔	✔		
STOP-Bang	✔	✔	✔	✔	✔		✔	✔	✔		
ASA checklist		✔			✔		✔	✔	✔	✔	
Wisconsin						✔	✔	✔			
SA-SDQ	✔	✔						✔	✔		Sweating during sleep, History of smoking
Haraldsson's						✔		✔	✔	✔	
Apnea Score							✔	✔			History of adenoidectomy

OSA = obstructive sleep apnea; BMI = body mass index; ASA = American Society of Anesthesiologists; STOP = snoring, tiredness, observed apnea, and high blood pressure; STOP-Bang = STOP questionnaire including BMI, age, neck circumference, gender; SA-SDQ = Sleep Apnea scale of the Sleep Disorders Questionnaire; ↑BP = hypertension

Haraldsson's questionnaire showing the highest specificity. The pooled specificity was 61.0% (95% CI: 55.0-67.0%; $I^2 = 43.8\%$). The PPV and NPV values ranged from 48-92% and from 57-72%, respectively (Table 6).

In studies on "patients without history of sleep disorders", the prevalence of OSA ranged from 21-69%. The sensitivity of different questionnaires in predicting OSA (AHI/AI or RDI $\geq$ 5) ranged from 66-95% with the Wisconsin Sleep questionnaire showing the highest sensitivity followed by the Berlin and the STOP-Bang questionnaires (Table 7). The pooled sensitivity was 77.0% (95% CI: 73.0-80.0%; $I^2 = 78.1\%$). To explore the reasons of heterogeneity, i.e., $I^2 > 50\%$, we performed subgroup analysis on studies using the same questionnaire, i.e., the Berlin or the Wisconsin sleep questionnaire, but this did not yield consistent results. The pooled sensitivity for the Berlin questionnaire was 77.0% (95% CI: 71.0-82.0%; $I^2 = 79.4\%$) and for the Wisconsin questionnaire was 83.0% (95% CI: 76.0-88.0%; $I^2 = 84.0\%$). As it is shown in Table 7, the average sensitivity of the questionnaires varied from 66-86% in all studies except Wisconsin (Sharma 2006b), which was shown to have an unexpectedly high value (95%) of sensitivity. This study was carried out by the same author of another study where the authors used a pre-screening set of questions to select the subjects out of the general population. After excluding this study as an outlier, the overall sensitivity was calculated as 76.0% (95% CI: 72.0-79.0%), and the index of heterogeneity of the results was reduced to $I^2 = 73\%$.

The specificity of different questionnaires in studies on "patients without history of sleep disorders" ranged from 38-95% with the Berlin questionnaire showing the highest specificity. The pooled specificity was 53.0% (95% CI: 50-57%; $I^2 = 88.8\%$). Subgroup analyses on studies using the same questionnaire, i.e., the Berlin or the Wisconsin sleep questionnaire, delivered inconsistent results. The pooled specificity for the Berlin questionnaire was 74.0% (95% CI: 65.0-81.0%; $I^2 = 90.7\%$) and the Wisconsin questionnaire was 50.0% (95% CI: 46.0-52.0%; $I^2 = 90.9\%$). As it is shown in Table 7, the average specificity of the questionnaires ranged from 38-76% in all studies except the Berlin questionnaire (Sharma 2006a), which was shown to have an unexpectedly high value (97%) of specificity. In that study, a pre-screening set of questions was used to select the subjects out of the general population. After excluding this study as an outlier, the overall specificity was calculated as 51.0% (95% CI: 48.0-55.0%), and the index of heterogeneity of the results was reduced to $I^2 = 74.7\%$. The PPV and NPV values ranged from 28-96% and from 38-97%, respectively (Table 7).

The sensitivity of different questionnaires in predicting moderate OSA (AHI/AI $\geq$ 15) ranged from 54-93%, with STOP-Bang questionnaires showing the highest sensitivity (Table 8). The prevalence of moderate OSA was 8-70% among the included studies. The pooled sensitivity was 77.0% (95% CI: 73.0-81.0%; $I^2 = 85.6\%$). The specificity of the questionnaires in predicting moderate OSA varied from 37-97% with the Berlin questionnaires showing the highest specificity (Table 8). The pooled specificity was 44% (95% CI: 41.0-47.0%; $I^2 = 84.0\%$). The PPV and NPV ranged from 11-97% and from 48-97%, respectively.

With regard to predicting severe OSA (AHI/AI $\geq$ 30), the sensitivity was very variable in the studies ranging from 17-100%, with the STOP-Bang questionnaire having the

Table 5 Appraisal of the included studies based on validity criteria

Items	Definition	Ahmadi 2008 (Berlin)	Chung 2008a (STOP)	Chung 2008b (STOP, ASA, Berlin)	Sharma 2006a (Berlin)	Sharma 2006b (Wisconsin)	Weatherwax 2003 (SA-SDQ)	Netzer 1999 (Berlin)	Young 1993 (Wisconsin)	Haraldsson 1992	Kapuniani 1988 (Apnea score)
Criteria of internal validity											
Valid reference standard	Overnight (F) or portable PSG (P)	F	F	F	F	F	F	P	F	F	F
Definition of the disease based on reference standard	OSA was diagnosed based on the PSG results (F)	F	F	F	F	F	F	F	F	F	F
Blind execution of index test and reference test	Those who performed PSG were not aware of the questionnaire results and vice versa (F)	U	F	F	U	U	U	U	U	U	U
Avoidance of verification bias	Interpretation of the PSG results independent from the questionnaire results (F)	F	F	F	U	U	U	U	U	U	U
Index test interpreted independently of all clinical information	Explicitly mentioned in the study that the questionnaire interpreted independently of all other patient's clinical information (F)	F	F	F	U	U	U	U	U	U	U
Study design	Prospective (F), retrospective (N)	N	F	F	F	F	F	F	F	F	F
Criteria of external validity											
Spectrum of disease	Inclusion and/or exclusion criteria mentioned (F)	F	F	F	F	F	F	F	F	F	F
Setting	Enough information to identify setting (F)	F	F	F	F	F	F	F	F	F	F
Previous screening / referral filter	The subject population is not screened before application of the questionnaire (F)	F	F	F	N	F	F	F	F	F	F
Demographic information	Age, gender, BMI data provided (F)	F	F	F	F	F	F	F	F	U	F
Explication of cut-off point of index test	Results presented for AHI or AI $\geq$ 5, 10 or more (F)	F	F	F	F	F	N	F	F	F	F
Percentage missing	Percentage missing mentioned (F)	N/A	F	F	F	F	F	F	F	U	U
Missing data management	Analysis of missing data for basic characteristics (F)	N/A	F	F	U	U	N/A	U	F	U	U
Subject selection for reference test	PSG done on all the subjects (F), all the subjects are invited or randomly selected to do PSG (P)	F	P	P	N	F	F	F	F	F	F

F = full meeting criteria; P = partially meeting criteria; U = unsure if meeting criteria in subgroups; N = not meeting criteria in subgroups; N/A = not applicable; AHI = apnea-hypopnea index; AI = apnea index; PSG = polysomnography; ASA = American Society of Anesthesiologists' checklist; STOP = STOP questionnaire; SA-SDQ = the Sleep Apnea scale of the Sleep Disorders Questionnaire

Table 6 Predictive parameters of questionnaires for screening of OSA (AHI ≥ 5) in "sleep-disorder patients"

Questionnaire/Study ID	Prevalence	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)
AS- Kapuniani 1988	0.51	0.59 (0.39, 0.78)	0.69 (0.48, 0.86)	0.66 (0.47, 0.85)	0.62 (0.44, 0.79)
AS w/o adenectomy	0.42	0.70 (0.50, 0.86)	0.65 (0.44, 0.83)	0.67 (0.50, 0.85)	0.68 (0.49, 0.86)
Berlin- Ahmadi 2008	0.43	0.68 (0.54, 0.80)	0.46 (0.34, 0.58)	0.48 (0.37, 0.59)	0.65 (0.52, 0.78)
Haraldsson 1991	0.76	**0.81 (0.64, 0.93)**	**0.80 (0.44, 0.97)**	**0.92 (0.83, 1.00)**	0.57 (0.31, 0.83)
SA-SDQ- Weatherwax 2003	0.55	0.80 (0.68, 0.88)	0.66 (0.52, 0.78)	0.74 (0.64, 0.84)	**0.72 (0.60, 0.84)**
Pooled estimates:	—	0.72 (0.66, 0.78) $I^2 = 23.0\%$	0.61 (0.55, 0.67) $I^2 = 43.8\%$	—	▬

PPV = positive predictive value; NPV = negative predictive value; AS = apnea score; OSA = obstructive sleep apnea; AHI = apnea-hypopnea index; SA-SDQ = Sleep Apnea scale of the Sleep Disorders Questionnaire; CI = confidence interval; I^2 = heterogeneity index. The highest values in each column are highlighted in bold fonts

Table 7 Predictive parameters of questionnaires for screening of OSA (AHI ≥ 5) in "patients without history of sleep disorders"

Questionnaire/Study ID	Prevalence	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)
ASA- Chung 2008	0.69	0.72 (0.63, 0.80)	0.38 (0.25, 0.52)	0.72 (0.64, 0.80)	0.38 (0.25, 0.51)
Berlin- Chung 2008	0.69	0.69 (0.60, 0.77)	0.56 (0.42, 0.70)	0.77 (0.69, 0.85)	0.44 (0.33, 0.56)
Berlin- Netzer 1999	0.68	0.86 (0.75, 0.93)	0.77 (0.59, 0.90)	0.88 (0.79, 0.98)	0.72 (0.54, 0.91)
Berlin- Sharma 2006a	0.60	0.85 (0.74, 0.93)	**0.95 (0.84, 0.99)**	**0.96 (0.91, 1.00)**	0.81 (0.70, 0.92)
STOP- Chung 2008	0.69	0.66 (0.56, 0.74)	0.60 (0.46, 0.73)	0.78 (0.69, 0.86)	0.44 (0.32, 0.56)
STOP-Bang- Chung 2008	0.69	0.84 (0.76, 0.90)	0.56 (0.42, 0.70)	0.81 (0.73, 0.87)	0.61 (0.46, 0.74)
Wisconsin- Sharma 2006b	0.25	**0.95 (0.82, 0.99)**	0.64 (0.54, 0.73)	0.46 (0.35, 0.57)	**0.97 (0.93, 1.00)**
Wisconsin- Young 1993	0.21	0.79 (0.71, 0.86)	0.46 (0.42, 0.51)	0.28 (0.23, 0.33)	0.89 (0.85, 0.92)
Pooled estimates:	—	0.77 (0.73, 0.80) $I^2 = 78.1\%$	0.53 (0.50, 0.57) $I^2 = 88.8\%$	—	—

PPV = positive predictive value; NPV = negative predictive value; OSA = obstructive sleep apnea; AHI: apnea-hypopnea index; ASA = American Society of Anesthesiologists' checklist; STOP = snoring, tiredness, observed apnea, and high blood pressure; STOP-Bang = STOP questionnaire including BMI, age, neck circumference, gender; CI = confidence interval; I^2 = heterogeneity index. The highest values in each column are highlighted in bold fonts

Table 8 Predictive parameters of questionnaires for screening of moderate OSA (AHI ≥ 15) in "patients without history of sleep disorders"

Questionnaire/Study ID	Prevalence	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)
ASA- Chung 2008	0.40	0.79 (0.67, 0.87)	0.37 (0.28, 0.47)	0.42 (0.36, 0.53)	0.72 (0.60, 0.84)
Berlin- Chung 2008	0.40	0.79 (0.67, 0.87)	0.50 (0.41, 0.60)	0.50 (0.41, 0.60)	0.78 (0.68, 0.87)
Berlin- Netzer 1999	0.70	0.54 (0.41, 0.66)	**0.97 (0.83, 1.00)**	**0.97 (0.92, 1.00)**	0.48 (0.35, 0.60)
STOP- Chung 2008	0.40	0.74 (0.62, 0.84)	0.53 (0.43, 0.63)	0.51 (0.41, 0.60)	0.76 (0.64, 0.85)
STOP-Bang- Chung 2008	0.40	**0.93 (0.84, 0.98)**	0.43 (0.33, 0.53)	0.51 (0.42, 0.60)	0.90 (0.78, 0.96)
Wisconsin- Young 1993	0.08	0.87 (0.76, 0.95)	0.40 (0.36, 0.44)	0.11 (0.08, 0.15)	**0.97 (0.95, 0.99)**
Pooled estimates:	—	0.77 (0.73, 0.81) $I^2 = 85.6\%$	0.44 (0.41, 0.47) $I^2 = 84.0\%$	—	▬

PPV = positive predictive value; NPV = negative predictive value; OSA = obstructive sleep apnea; AHI = apnea-hypopnea index; ASA = American Society of Anesthesiologists' checklist; STOP = snoring, tiredness, observed apnea, and high blood pressure; STOP-Bang = STOP questionnaire including BMI, age, neck circumference, gender; CI = confidence interval; I^2 = heterogeneity index. The highest values in each column are highlighted in bold fonts

highest value. The specificity varied from 36-97%, with the Berlin questionnaire showing the highest specificity (Table 9) among the included studies. The pooled sensitivity for The prevalence of severe OSA was 22-69% among the included studies. The pooled sensitivity for predicting severe OSA was 67% (95% CI: 60.0-73.0%; $I^2 = 96.8\%$). The respective value for pooled specificity was 45% (95% CI: 41.0-49.0%; $I^2 = 91.9\%$). The PPV and NPV ranged from 31-92% and from 34-100%, respectively.

Table 9 Predictive parameters of questionnaires for screening of severe OSA (AHI ≥ 30) in "patients without history of sleep disorders"

Questionnaire/Study ID	Prevalence	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)
ASA- Chung 2008	0.22	0.87 (0.73, 0.96)	0.36 (0.28, 0.45)	0.27 (0.19, 0.35)	0.90 (0.83, 0.98)
Berlin- Chung 2008	0.22	0.87 (0.73, 0.96)	0.46 (0.38, 0.55)	0.31 (0.22, 0.40)	0.92 (0.86, 0.98)
Berlin- Netzer 1999	0.69	0.17 (0.09, 0.28)	**0.97 (0.83, 1.00)**	**0.92 (0.77, 1.00)**	0.34 (0.24, 0.44)
STOP- Chung 2008	0.22	0.79 (0.64, 0.91)	0.49 (0.40, 0.57)	0.30 (0.21, 0.40)	0.89 (0.80, 0.95)
STOP-Bang- Chung 2008	0.22	**1.00 (0.91, 1.00)**	0.37 (0.29, 0.46)	0.31 (0.23, 0.39)	**1.00 (0.93, 1.00)**
Pooled estimates:	—	0.67 (0.60, 0.73) $I^2 = 96.8\%$	0.45 (0.41, 0.49) $I^2 = 91.9\%$	—	—

PPV = positive predictive value; NPV = negative predictive value; OSA = obstructive sleep apnea; AHI = apnea-hypopnea index; ASA = American Society of Anesthesiologists' checklist; STOP = snoring, tiredness, observed apnea, and high blood pressure; STOP-Bang = STOP questionnaire including BMI, age, neck circumference, gender; CI = confidence interval; I^2 = heterogeneity index. The highest values in each column are highlighted in bold fonts

Discussion

This systematic review identified and evaluated eight available patient-based questionnaires for screening of OSA. Among the questionnaires validated on "sleep-disorder patients", we found Haraldsson with the highest sensitivity and specificity. The accuracy of results are significantly heterogeneous in the studies on "patients without history of sleep disorders", even in the studies on the same questionnaire. In terms of predicting the existence of OSA (AHI/AI ≥ 5), the Wisconsin and the Berlin questionnaires were shown to have the highest sensitivity and specificity, respectively. However, the validity of these studies is unclear due to the potential effects of pre-screening and the risk of verification bias. In terms of predicting moderate or severe OSA, the STOP-Bang and the Berlin questionnaires were found to have the highest sensitivity and specificity, respectively. The STOP and STOP-Bang questionnaires were found to have the highest methodological validity, reasonable accuracy, and easy-to-use features.

Due to the relatively high prevalence of undiagnosed OSA and its short- and long-term complications, a reliable screening tool is required for a quick prediction of OSA. A quick and reliable screening test would enable clinicians to detect the possibility of OSA during initial clinical visits and then determine those patients at high risk and either in need of further assessment or in need of immediate therapeutic treatment. Questionnaires can be appropriate tools for quick prediction of obstructive sleep apnea as they can be applied and scored easily as part of routine daily practice. This approach is tremendously important to anesthesia practitioners and to surgical patients, as there is insufficient time in the short preoperative period to complete an assessment of every patient with the standard diagnostic approach, i.e., sleep lab PSG.

An ideal screening questionnaire should have three important characteristics,[36] namely, 1) *Feasibility*: Patients and healthcare providers should find the questionnaire user-friendly; 2) *Accuracy*: There should be a clear validation process that leads to high accuracy parameters; 3) *Generalizability:* Valid results should be realized when the questionnaire is used on different target populations, i.e., the questionnaire has been validated in different study populations.

The response rate is generally considered as an index factor of the *feasibility* of a questionnaire.[36] However, only three of the ten studies that we evaluated reported their response rate.[29,31,34] All three studies were on the general population, and the response rate ranged from 82-91%. We assumed that most patients who were referred to sleep clinics for their sleep-related complaints were more willing to respond to the questions. If this assumption is true, the response rate should be higher in this group of patients than in the general population. Thus, an easier and more straightforward questionnaire is needed to be used in the general population to achieve a higher response rate. Some authors have taken this need into consideration. While the older type of questionnaire, such as the SA-SDQ,[25] included a greater number of questions that were more complicated, the newer questionnaires, such as STOP and STOP-Bang, have fewer and more straightforward questions.[31,32]

The imprecision we detected in the accuracy of these questionnaires as screening tools could have arisen from several factors. One factor could be the methodological quality of the papers,[37] which was poorly reported in the majority of the studies we evaluated. Methodological quality was clearly reported in the studies by Chung et al. and Ahmadi 2008.[30-32] Compared with the studies by Ahmadi 2088,[8,30] the studies by Chung et al.[31,32] are considered as being even higher quality due to their prospective design and the blind execution of PSG, i.e., the persons responsible for carrying out the standard PSG for the patients were not aware of the results of the

questionnaires. Other studies did not address the validation process of the questionnaires, fully evaluate the validity of their results, or, more specifically, eliminate the risk of verification bias. This bias occurs when the disease status, i.e., OSA, is not determined in all subjects who are tested (or screened by the questionnaire) and when the probability of verification depends on the questionnaire results, other clinical variables, or both, rather than those of the "gold standard", i.e., PSG.[38]

When verification of disease status is anticipated among high-risk patients, a bias is introduced that can markedly increase the apparent sensitivity of the questionnaire and reduce its specificity. Sharma 2006a studied the Berlin questionnaire in the general population and yielded high predictive parameter values (sensitivity 85%).[33] This study has selection bias, as a primary "pre-screening" was performed before the subjects were selected for the questionnaire being studied. This primary screening was performed using a questionnaire covering four main variables of OSA (snoring, tiredness, obesity, and hypertension). Subjects answering yes to these questions proceeded to the next step to be selected as the subjects for the validation analysis of the study questionnaire. Consequently, the study subjects could not represent the general population. This approach resulted in heterogeneity when a meta-analysis was attempted.

Another factor in the inaccuracy of these questionnaires was the variety of the target populations among the different studies. In an effort to unify the subject populations, all studies were divided into two major groups: studies on patients "with sleep disorder problems"[25,27,30,35] and studies on "patients without known sleep problems".[28,29,31-34] Due to the high prevalence of OSA in patients with sleep disorders, we could not use the first group of studies as a reference for the strength of the questionnaires designed to screen OSA in the general population. Recent evidence shows that a test performance varies in different populations because of the severity of the disease. For example, a patient population with a higher disease prevalence may include more severely diseased patients; therefore, the test would perform better in this population.[39] It is also important to emphasize that an ideal diagnostic test in a general population should have a relatively high specificity to minimize false positives, nevertheless, it should have sufficient sensitivity. Conversely, an ideal diagnostic test in a population with a high pre-test probability of disease should have higher sensitivity while maintaining high specificity.[40]

Lack of a standard definition for some factors involved in OSA questionnaires could also result in the heterogeneity of the data among the studies. For example, the cut-off numbers used for BMI ranged from 25-35, and different scales were used for snoring in the questionnaires. The Epworth Sleepiness Scale (ESS) is a standard questionnaire to measure daytime sleepiness;[41] however, Osman et al. showed that this questionnaire had no value in distinguishing simple snorers from patients with OSA.[42] Standard definition of factors is a prerequisite for standardizing questionnaire components and evaluating their values in different combinations.

It is important to screen OSA patients in the perioperative setting to identify those who require further management. An ideal questionnaire should be sensitive enough to detect any patients with OSA, but more importantly, it should recognize patients with severe forms of OSA, as they may likely have perioperative complications. In this regard, there is not enough evidence available in the literature to indicate the AHI/Al cut-off that places the patient at significant increased risk of postoperative complications. However, the correlation between AHI scores and lifetime complications, such as motor vehicle accidents and the risk of atrial fibrillation, has been suggested in non-surgical populations.[43] Therefore, we can hypothesize that the more apnea episodes patients experience, the more vulnerable they are to the effect of anesthesia and surgery. In this regard, the Wisconsin sleep questionnaire was not validated for detection of severe OSA. The Berlin questionnaire, which was shown to have high sensitivity for detecting OSA (69-86%), was found to be relatively less sensitive in detecting moderate and severe cases, i.e., sensitivity: 54-79% and 17-87%, respectively. The STOP-Bang questionnaire was shown to have consistently high sensitivity for detecting OSA in different AHI cut-offs and severity levels (AHI $\geq$ 5: 84%, AHI $\geq$ 15: 93%, AHI $\geq$ 30: 100%). This was achieved in exchange of losing the specificity of the questionnaire, which is not the major concern in the preoperative setting, as screened patients will always be advised to confirm their diagnosis with postoperative PSG at their leisure.

A meta-analysis of clinical screening tests for obstructive sleep apnea was conducted by Ramachandran et al. in 2008.[44] This study included eight papers on questionnaires and 18 articles on clinical prediction tests, including clinical scales, algorithms, and prediction models. We can distinguish our systematic review from Ramachandran's review in at least three different areas, i.e., the main focus, the presentation of the results, and the conclusion. First, our systematic review is focused only on questionnaires, whereas Ramachandran's review involves other types of clinical screening tools. Since questionnaires do not need sophisticated mathematical calculations, we consider them as being more practical and, therefore, more convenient to be used in the daily clinical practice. Second, although our method of presenting the details of our data is similar to Ramachandran's method, we have included descriptive features, such as tables of the included and excluded

studies and tables of the included questionnaires and quality assessments. While the included studies are extremely diverse in their quality, design, and patient population, the summary tables allow for individual interpretation of the available literature on the OSA questionnaires. Finally, because of the inconsistent results, even between studies regarding the same questionnaire, we did not make a definite conclusion regarding the most accurate questionnaire. However, we did recommend the STOP (or STOP-Bang) questionnaire in consideration of its high-quality methodology. Although Ramachandran *et al.* evaluated the quality of the papers, their review did not provide the details of the assessment, and this factor was not taken into consideration in the synthesis of their conclusion.

We concluded that questionnaires have the potential to screen patients for high risk of having OSA, and this approach can raise the awareness of anesthesiologists and surgeons to the possibility of OSA in surgical patients. Also, this approach may facilitate early detection of patients who need further assessment and who would benefit from perioperative precautions for OSA patients. Identification of patients at risk of OSA could potentially reduce the rate of OSA-related postoperative complications.[45] Due to the inconsistent literature, it is difficult to draw a definite conclusion regarding the most accurate OSA questionnaire available. However, we suggest the STOP questionnaire and its extended version, STOP-Bang, for OSA screening in surgical patients.

Despite the STOP questionnaire being developed in our institute by the corresponding author of this review, we used a systematic approach to review the literature objectively following the standard guidelines. The STOP and STOP-Bang questionnaires have high-quality methodological and reasonably accurate results. The scoring method is straightforward, and the acronym, *STOP*, (*snoring, tiredness, observed apnea, and high blood pressure*) makes it easy for clinicians to remember. For anesthesiologists facing a large number of patients in the preoperative clinic, this questionnaire could serve as a quick tool to screen OSA in surgical patients and could increase the awareness of OSA precautions in perioperative management.

Appendix Search Strategy for Ovid Medline

1.	(obstructive adj3 sleep* adj5 (apnoea* or apnea*)).mp.
2.	(hypopnea* or hypopnoea*).mp.
3.	(OSA? or SHS or OSAHS).mp.
4.	(sleep adj3 related adj3 respiratory adj3 disorder*).mp.
5.	(sleep adj3 disordered adj3 breathing).mp.
6.	exp Sleep Apnea Syndromes/
7.	1 or 2 or 3 or 4 or 5 or 6
8.	exp Risk Assessment/cl, mt (Classification, Methods)
9.	exp Mass Screening/mt (Methods)
10.	exp "validation studies (publication type)"/
11.	questionnaire*.mp.
12.	(sensitivity* or specificity*).mp.
13.	(screen* adj5 risk).mp.
14.	(score* or scale*).mp.
15.	8 or 9 or 10 or 11 or 12 or 13 or 14
16.	7 and 15

References

1. *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.
2. *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.
3. *Shahar E, Whitney CW, Redline S, et al.* 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.
4. *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.
5. *Ing AJ, Ngu MC, Breslin AB.* Obstructive sleep apnea and gastroesophageal reflux. Am J Med 2000; 108(Suppl 4): 120S-5S.
6. *Idris I, Hall AP, O'Reilly J, et al.* Obstructive sleep apnoea in patients with type 2 diabetes: aetiology and implications for clinical care. Diabetes Obes Metab 2009; 11: 733-41.
7. *Abrams B.* Long-term sleep apnea as a pathogenic factor for cell-mediated autoimmune disease. Med Hypotheses 2005; 65: 1024-7.
8. *Rowley JA, Aboussouan LS, Badr MS.* The use of clinical prediction formulas in the evaluation of obstructive sleep apnea. Sleep 2000; 23: 929-38.
9. *Crocker BD, Olson LG, Saunders NA, et al.* Estimation of the probability of disturbed breathing during sleep before a sleep study. Am Rev Respir Dis 1990; 142: 14-8.
10. *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.
11. *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.
12. *Maislin G, Pack AI, Kribbs NB, et al.* A survey screen for prediction of apnea. Sleep 1995; 18: 158-66.
13. *Deville WL, Buntinx F, Bouter LM, et al.* Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol 2002; 2: 9.

14. *Irwig L, Tosteson AN, Gatsonis C, et al.* Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994; 120: 667-76.

15. *Walter SD, Irwig L, Glasziou PP.* Meta-analysis of diagnostic tests with imperfect reference standards. J Clin Epidemiol 1999; 52: 943-51.

16. *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.

17. *Douglass AB, Bornstein R, Nino-Murcia G, et al.* The Sleep Disorders Questionnaire I: creation and multivariate structure of SDQ. Sleep 1994; 17: 160-7.

18. *Gurubhagavatula I, Maislin G, Pack AI.* An algorithm to stratify sleep apnea risk in a sleep disorders clinic population. Am J Respir Crit Care Med 2001; 164: 1904-9.

19. *Pouliot Z, Peters M, Neufeld H, Kryger MH.* Using self-reported questionnaire data to prioritize OSA patients for polysomnography. Sleep 1997; 20: 232-6.

20. *Rosenthal LD, Dolan DC.* The Epworth sleepiness scale in the identification of obstructive sleep apnea. J Nerv Ment Dis 2008; 196: 429-31.

21. *Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A.* Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006; 6: 31.

22. *Izci B, Ardic S, Firat H, Sahin A, Altinors M, Karacan I.* Reli-ability and validity studies of the Turkish version of the Epworth Sleepiness Scale. Sleep Breath 2008; 12: 161-8.

23. *Onen SH, Dubray C, Decullier E, Moreau T, Chapuis F, Onen F.* Observation-based Nocturnal Sleep Inventory: screening tool for sleep apnea in elderly people. J Am Geriatr Soc 2008; 56: 1920-5.

24. *Teculescu D, Guillemin F, Virion JM, et al.* Reliability of the Wisconsin Sleep Questionnaire: a French contribution to inter-national validation. J Clin Epidemiol 2003; 56: 436-40.

25. *Weatherwax KJ, Lin X, Marzec ML, Malow BA.* Obstructive sleep apnea in epilepsy patients: the Sleep Apnea scale of the Sleep Disorders Questionnaire (SA-SDQ) is a useful screening instru-ment for obstructive sleep apnea in a disease-specific population. Sleep Med 2003; 4: 517-21.

26. *Malow BA, Levy K, Maturen K, Bowes R.* Obstructive sleep apnea is common in medically refractory epilepsy patients. Neurology 2000; 55: 1002-7.

27. *Kapuniai LE, Andrew DJ, Crowell DH, Pearce JW.* Identifying sleep apnea from self-reports. Sleep 1988; 11: 430-6.

28. *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.

29. *Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S.* The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 1993; 328: 1230-5.

30. *Ahmadi N, Chung SA, Gibbs A, Shapiro CM.* The Berlin ques-tionnaire for sleep apnea in a sleep clinic population: relationship

31. *Chung F, Yegneswaran B, Liao P, et al.* STOP questionnaire: a tool to screen patients for obstructive sleep apnea. Anesthesiol-ogy 2008; 108: 812-21.

32. *Chung F, Yegneswaran B, Liao P, 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: 822-30.

33. *Sharma SK, Vasudev C, Sinha S, Banga A, Pandey RM, Handa KK.* Validation of the modified Berlin questionnaire to identify patients at risk for the obstructive sleep apnoea syndrome. Indian J Med Res 2006; 124: 281-90.

34. *Sharma SK, Kumpawat S, Banga A, Goel A.* Prevalence and risk factors of obstructive sleep apnea syndrome in a population of Delhi, India. Chest 2006; 130: 149-56.

35. *Haraldsson PO, Carenfelt C, Knutsson E, Persson HE, Rinder J.* Preliminary report: validity of symptom analysis and daytime pol-ysomnography in diagnosis of sleep apnea. Sleep 1992; 15: 261-3.

36. *Fitzpatrick R, Davey C, Buxton MJ, Jones DR* (1998) Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 2: i-iv, 1-74.

37. *Reid MC, Lachs MS, Feinstein AR.* Use of methodological stan-dards in diagnostic test research. Getting better but still not good. JAMA 1995; 274: 645-51.

38. *Punglia RS, D'Amico AV, Catalona WJ, Roehl KA, Kuntz KM.* Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. N Engl J Med 2003; 349: 335-42.

39. *Leeflang MM, Bossuyt PM, Irwig L.* Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. J Clin Epidemiol 2009; 62: 5-12.

40. *Ross SD, Sheinhait IA, Harrison KJ, et al.* Systematic review and meta-analysis of the literature regarding the diagnosis of sleep apnea. Sleep 2000; 23: 519-32.

41. *Johns MW.* A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991; 14: 540-5.

42. *Osman EZ, Osborne J, Hill PD, Lee BW.* The Epworth Sleepiness Scale: can it be used for sleep apnoea screening among snorers? Clin Otolaryngol Allied Sci 1999; 24: 239-41.

43. *Shafazand S.* Perioperative management of obstructive sleep apnea: ready for prime time? Cleve Clin J Med 2009; 76(Suppl 4): S98-103.

44. *Ramachandran SK, Josephs LA.* A meta-analysis of clinical screening tests for obstructive sleep apnea. Anesthesiology 2009; 110: 928-39.

45. *Liao P, Yegneswaran B, Vairavanathan S, Zilberman P, Chung F.* Postoperative complications in patients with obstructive sleep apnea: a retrospective matched cohort study. Can J Anesth 2009; 56: 819-28.