SLEEP BREATHING PHYSIOLOGY AND DISORDERS · ORIGINAL ARTICLE

Diagnostic accuracy of screening questionnaires for obstructive sleep apnoea in adults in different clinical cohorts: a systematic review and meta-analysis

Lizelle Bernhardt[1,2] · Emer M. Brady[1] · Suzanne C. Freeman[3] · Helena Polmann[4] · Jéssica Conti Réus[4] · Carlos Flores-Mir[5] · Graziela De Luca Canto[4] · Noelle Robertson[6] · Iain B. Squire[1]

Abstract

Purpose The majority of individuals with clinically significant obstructive sleep apnoea (OSA) are undiagnosed and untreated. A simple screening tool may support risk stratification, identification, and appropriate management of at-risk patients. Therefore, this systematic review and meta-analysis evaluated and compared the accuracy and clinical utility of existing screening questionnaires for identifying OSA in different clinical cohorts.

Methods We conducted a systematic review and meta-analysis of observational studies assessing the diagnostic value of OSA screening questionnaires. We identified prospective studies, validated against polysomnography, and published to December 2020 from online databases. To pool the results, we used random effects bivariate binomial meta-analysis.

Results We included 38 studies across three clinical cohorts in the meta-analysis. In the sleep clinic cohort, the Berlin questionnaire's pooled sensitivity for apnoea-hypopnoea index (AHI) ≥ 5, ≥ 15, and ≥ 30 was 85%, 84%, and 89%, and pooled specificity was 43%, 30%, and 33%, respectively. The STOP questionnaire's pooled sensitivity for AHI ≥ 5, ≥ 15, and ≥ 30 was 90%, 90%, and 95%, and pooled specificity was 31%, 29%, and 21%. The pooled sensitivity of the STOP-Bang questionnaire for AHI ≥ 5, ≥ 15, and ≥ 30 was 92%, 95%, and 96%, and pooled specificity was 35%, 27%, and 28%. In the surgical cohort (AHI ≥ 15), the Berlin and STOP-Bang questionnaires' pooled sensitivity were 76% and 90% and pooled specificity 47% and 27%.

Conclusion Among the identified questionnaires, the STOP-Bang questionnaire had the highest sensitivity to detect OSA but lacked specificity. Subgroup analysis considering other at-risk populations was not possible. Our observations are limited by the low certainty level in available data.

Keywords Obstructive sleep apnoea · Diagnostic accuracy · Screening questionnaires · Systematic review · Meta-analysis

✉ Lizelle Bernhardt
lb382@le.ac.uk

1 Department of Cardiovascular Sciences and NIHR Biomedical Research Centre, University of Leicester, Glenfield Hospital, Groby Road, Leicester LE3 9QP, UK

2 Community Health Services, Leicestershire Partnership NHS Trust, Leicester, UK

3 Department of Health Sciences, University of Leicester, Leicester, UK

4 Brazilian Centre for Evidence-Based Research, Federal University of Santa Catarina, Florianopolis, Brazil

5 Department of Dentistry, University of Alberta, Edmonton, Canada

6 Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester, UK

Abbreviations

AHI	Apnoea-hypopnoea index
CI	Confidence interval
DOR	Diagnostic odds ratio
FN	False negative
FP	False positive
NPV	Negative predictive values
OSA	Obstructive sleep apnoea
PPV	Positive predictive values
PRISMA	Reporting Items for Systematic Reviews and Meta-analysis
PSG	Polysomnography
QUADAS	Quality Assessment for Diagnostic Accuracy Studies
RDI	Respiratory disturbance index
ROC	Receiver operating characteristic
SROC	Summary receiver operating characteristic
TN	True negative
TP	True positive

Introduction

With an estimated 425 million individuals affected worldwide, clinically important obstructive sleep apnoea (OSA) poses a global public health problem [1]. Characterised by upper airway collapse, exaggerated negative intrathoracic pressure, oxidative stress, and systemic inflammation, OSA is associated with significant cardiovascular and metabolic complications, including hypertension, stroke, heart failure, and diabetes [2–7].

Despite the high prevalence and associated sequelae, most individuals with OSA remain undiagnosed, posing a significant risk to the individual patient and health care systems as complications develop [1, 8–10]. Barriers to the diagnosis and treatment of OSA are multifaceted and include geographical variation and inequity in the availability of sleep services and access to polysomnography (PSG), often limited by cost and long waiting times [11].

To support risk stratification and appropriate referrals in individuals at-risk, a simple and reliable screening tool may help triage patients at risk of OSA, for consideration of referral to specialist services for appropriate management [12–14]. Clinical prediction formulae have been developed but are limited by complexity and the requirement for a computer or mathematical calculations [15]. In contrast, OSA screening questionnaires are less complicated and may be a viable alternative to clinical prediction formulae in specific settings.

To date, there have been four systematic reviews exploring the accuracy of OSA screening tools in adults [12, 16–18]. One of the first systematic reviews and meta-analyses to explore the accuracy of screening tools for OSA identified four screening questionnaires; however, due to heterogeneity pertaining to the questionnaire, OSA definition, and threshold, these were not meta-analysed [16]. Ramachandran [17] reported that clinical prediction models performed better than the eight questionnaires studied to predict OSA in pre-operative cohorts. Abrishami [12] focused on a 'sleep disorder' cohort and a cohort 'without a history of sleep disorders'. It was concluded that questionnaires were useful for early detection of OSA, especially in the surgical population. Despite finding it difficult to draw a definite conclusion about questionnaire accuracy, the STOP and STOP-Bang questionnaires were recommended for screening in a surgical population [12]. Recently, Chui [18] compared the diagnostic accuracy of the Berlin, STOP-Bang, STOP, and Epworth Sleepiness Scale. In line with Abrishami [12], they reported the STOP-Bang to have the highest sensitivity in both the sleep clinic and surgical populations.

Since the publication of these systematic reviews, new OSA screening questionnaires have emerged, further validation studies conducted, and different clinical settings and patient cohorts considered. As test performance often varies across clinical cohorts, it is recommended that tools are evaluated in clinically relevant cohorts [19]. Hence, the objective of this systematic review and meta-analysis was to evaluate the accuracy and clinical utility of existing questionnaires, when used alone, as screening tools for the identification of OSA in adults in different clinical cohorts.

Methods

The protocol was registered at the International Prospective Register of Systematic Reviews (PROSPERO) (CRD42018104018) and conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines [20].

Types of studies

We included observational studies that met the following eligibility criteria:

Inclusion criteria: (1) prospective studies measuring the diagnostic value of screening questionnaires for OSA; (2) studies in adults (> 18 years of age); (3) studies in which the accuracy of the questionnaire was validated by level one or two PSG; (4) OSA was defined as apnoea-hypopnoea index (AHI) or Respiratory Disturbance Index (RDI) > 5; (5) data allowed for construction of 2×2 contingency tables; (6) publication in English, Spanish, or Portuguese.

Exclusion criteria: (1) studies measuring the diagnostic value of clinical scales, scores, and prediction equations

as screening tools for OSA; (2) conference proceedings, reviews, or case reports; (3) insufficient data for analysis after several attempts to contact the author; (4) studies in children (< 18 years of age); (5) level three and four portable studies were used as the reference standard; (6) studies conducted in in-patient settings; (7) publication language is other than English, Spanish, or Portuguese.

Index test: the test under evaluation was only OSA screening questionnaires (self-reported or clinician completed).

Reference standard: the reference standard was a level one or two PSG.

Target conditions: the target condition was OSA, defined as AHI or RDI.

- AHI/RDI ≥ 5—diagnostic cut-off for OSA
- AHI/RDI ≥ 15—diagnostic cut-off for moderate to severe OSA
- AHI/RDI ≥ 30—diagnostic cut-off for severe OSA

Search methods for identification of studies

Comprehensive literature searches in CINAHL PLUS, Scopus, PubMed, Web of Science, and the Latin American and Caribbean Health Sciences Literature (LILACS) database were conducted from inception to 18 December 2020. Detailed individual search strategies (Online Resource 1 & 2), with appropriate truncation and word combinations, were developed for each database. Additional records were identified from grey literature sources comprising ETHos, OpenGrey, Google Scholar, ProQuest, and New York Grey Literature Report. The reference lists from the final articles for analysis and related review articles were manually searched for references that could have been omitted during the electronic database searches.

Data collection and analysis

Study selection

Two reviewers (LB, EB) screened the titles and abstracts of the electronic search results independently to identify studies eligible for inclusion in the review. Records classified as 'excluded' by both reviewers were excluded. The full text of any study about which there was disagreement or uncertainty was assessed independently against the selection criteria and resolved through discussion and consultation with a third reviewer (IS or NR). Duplicates were identified and excluded before recording the selection process in sufficient detail to complete the PRISMA flow diagram and tables describing the characteristics of the excluded studies (Online Resource 3) [20].

Data extraction and management

Two reviewers (LB, EB) independently conducted data extraction on all studies included and extracted the data required to reconstruct the 2 × 2 contingency tables, including true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values. Where these values were not documented, we extrapolated the values from equations when data allowed. A data collection form tailored to the research question and fulfilling the data entry requirements of MetaDTA (Diagnostic Test Accuracy Meta-Analysis v1.43) was utilised [21].

HP and JR extracted the study characteristics and demographic data for all included studies, and LB and EB entered the data into Review Manager 5.3 [22].

No studies with inconclusive results were identified.

Assessment of methodological quality

The quality of studies included was appraised independently by the reviewers (LB, EB) utilising the Quality Assessment for Diagnostic Accuracy Studies tool (QUADAS-2) with disagreements resolved through consultation with a third reviewer (IS or NR) [23].

Statistical analysis and data synthesis

Statistical analysis was performed according to "Chapter 10" of the *Cochrane Handbook for Systematic Review of Diagnostic Test Accuracy* [24].

Questionnaire screening was considered positive for OSA if the questionnaire score was above the defined threshold specified in the primary study and negative if the questionnaire score was below the defined threshold. The TP, FP, TN, and FN results were produced by cross-classifying the questionnaire results with those of the PSG results. These were based on the ability of screening questionnaires to classify and detect OSA correctly.

The sensitivity and specificity of individual studies were calculated using 2 × 2 contingency tables and presented as forest plots. The meta-analysis was conducted using MetaDTA version 1.43, which models sensitivity and specificity by fitting the random effects bivariate binomial model of Chu and Cole [25, 26]. The summary receiver operating characteristic (SROC) plot was drawn using the hierarchical SROC parameters, which are estimated from the bivariate model parameters using the equivalence equations of Harbord [27]. Following guidance from the *Cochrane Handbook for Systematic Review of Diagnostic Test Accuracy*, we did not pool the positive and negative

predictive values due to the prevalence of OSA varying across studies [24].

As per the Cochrane DTA handbook, we investigated heterogeneity by plotting the observed study results and SROC curve in the ROC space alongside the 95% confidence region [24].

We conducted a meta-regression to investigate differences in sensitivity and specificity between questionnaires, including the type of questionnaire as a covariate. Meta-regression was conducted in R version 4.0.1 using the lme4 package [28].

To assess the robustness of the meta-analysis, sensitivity analyses were conducted by excluding studies based on their QUADAS-2 assessment score [23]. Those identified as high risk in any QUADAS-2 domain or as unclear in four domains were excluded. Different AASM (American Academy of Sleep Medicine) scoring criteria and desaturation (and arousal) thresholds were applied to the included studies. We conducted additional sensitivity analyses by analysing studies that applied the $\geq 3\%$ desaturation scoring criteria together and those that applied the $\geq 4\%$ desaturation scoring criteria (summarised in Table 1).

We neither explored reporting bias, nor assessed publication bias due to the uncertainty about the determinants of publication bias for diagnostic accuracy studies, and the inadequacy of tests for detecting funnel plot asymmetry [74].

Results

Search results and study characteristics

Search results are summarised in Fig. 1.

Of 45 studies, 29 were included for meta-analysis in the sleep clinic population ($n = 10,951$), 7 were included for meta-analysis in the surgical population ($n = 2275$), and 2 were included in the resistant hypertension population ($n = 541$). The remaining 7 studies were excluded from the meta-analysis due to heterogeneity of included populations. Study characteristics and demographic data of the included studies are summarised in Tables 1 and 2 [29–73]. Overall, 10 clinical settings were identified, of which the sleep clinic, surgical, and resistant hypertension cohorts had sufficient studies for inclusion in the meta-analysis.

OSA obstructive sleep apnoea, *AHI* apnoea-hypopnoea index, *RDI* respiratory disturbance index, *Lab* laboratory,

PSG polysomnography, *AASM* American Academy of Sleep Medicine.

SD standard deviation, *kg* kilogramme, *m* metre, *cm* centimetre, *NC* neck circumference, *WC* waist circumference, *AHI* apnoea-hypopnoea index, *n/a* not applicable.

Methodological quality of included studies

Results of the QUADAS-2 assessment are summarised in Fig. 2 and Online Resource 4 [29–73].

In the patient selection domain, 3 studies were rated as high risk of bias due to the case–control study design. For both the index test and reference standard domains, 18 studies were rated as unclear risk of bias due to inadequate information related to blinding; it was unclear if the index test and reference standard findings were interpreted without the knowledge of the other. Thirty-four studies were rated as unclear risk of bias in the flow and timing domain due to lack of reporting on the time interval between the index test and the reference standard. Applicability was rated as low risk in all 45 studies.

Sleep clinic population

In the sleep clinic population ($N = 10,951$) (Fig. 3), the Berlin (score cut-off ≥ 2) (Online Resource 5), STOP (score cut-off ≥ 2), and STOP-Bang (score cut-off ≥ 3) (Online Resource 6) questionnaires were included in the meta-analysis [58, 75]. The ASA checklist, SA-SDQ, and STOP-Bang (cut-off ≥ 5) questionnaires were excluded due to insufficient studies.

Predictive parameters of the Berlin questionnaire (score cut-off ≥ 2)

The prevalence of AHI ≥ 5 (all OSA), AHI ≥ 15 (moderate to severe), and AHI ≥ 30 (severe) OSA was 84%, 64%, and 50% respectively. The pooled sensitivity of the Berlin questionnaire to predict all OSA, moderate–severe, and severe OSA was 85% (95% confidence interval (CI): 79%, 89%), 84% (95% CI: 79%, 89%), and 89% (95% CI: 80%, 94%) respectively. Pooled sensitivity remained consistent across OSA severity. Pooled specificity was 43% (95% CI: 30%, 58%), 30% (95% CI: 20%, 41%), and 33% (95% CI: 21%, 46%) respectively. The corresponding diagnostic odds ratio (DOR) were 4.3 (95% CI: 0.7, 7.8), 2.3 (95% CI: 1.3, 3.3), and 3.9 (95% CI: 2.1, 5.7) (Fig. 4, Table 3).

Table 1 Study characteristics

Study	Country	Questionnaire	N	Validation tool	OSA definition	Apnoea definition	Hypopnoea definition	% desaturation	Scoring criteria
Sleep clinic population									
Abdullah 2018 [29]	Malaysia	STOP-Bang	134	Level 1—Lab PSG	AHI ≥ 5	Complete cessation of airflow for ≥ 10 s	A reduction of airflow of ≥ 50% for ≥ 10 s associated with 3% desaturation	3%	AASM 2012
Alhouqani 2015 [30]	UAE	STOP-Bang	193	Level 1—Lab PSG	AHI ≥ 5	A cessation of airflow for > 10 s	≥ 30% decrease in airflow with 3% oxygen desaturation and/or arousal	3%	AASM 2012
Amra 2013 [31]	Iran	Berlin	157	Level 1—Lab PSG	AHI > 5	Complete cessation of airflow for ≥ 10 s	A reduction of airflow of ≥ 50% for ≥ 10 s associated with ≥ 3% desaturation	≥ 3%	AASM 1999
Amra 2018 [32]	China	Berlin, STOP-Bang	400	Level 1—Lab PSG	AHI ≥ 5	≥ 90% drop in airflow from baseline for ≥ 10 s	≥ 30% decrease in airflow with 3% oxygen desaturation and/or arousal	3%	AASM 2012
Oktay Arslan 2020 [33]	Turkey	Berlin	1003	Level 1—Lab PSG	AHI ≥ 5	≥ 90% drop in airflow from baseline for ≥ 10 s	≥ 30% decrease in airflow with 3% oxygen desaturation and/or arousal	3%	AASM 2012
Avincsal 2017 [34]	Turkey	STOP-Bang	162	Level 1—Lab PSG	AHI > 5	≥ 90% drop in airflow from baseline, which lasted ≥ 10 s	≥ 30% reduction in airflow that lasted ≥ 10 s and was associated with ≥ 4% desaturation	≥ 4%	AASM 2007
BaHammam 2015 [35]	Saudi Arabia	STOP-Bang	100	Level 1—Lab PSG	AHI ≥ 5	≥ 90% drop in airflow from baseline, which lasted ≥ 10 s	≥ 30% decrease in airflow with 3% oxygen desaturation and/or arousal	≥ 3%	AASM 2012
Boynton 2013 [36]	USA	STOP-Bang	219	Level 1—Lab PSG	AHI > 5	Complete absence of airflow for ≥ 10 s	≥ 50% decrease in airflow followed by an arousal, awakening, or ≥ 3% desaturation	≥ 3%	AASM 2007
Deflandre 2018 [37]	Belgium	STOP-Bang, OSA50	159	Level 1—Lab PSG	AHI > 30	≥ 90% drop in airflow from baseline, which lasted ≥ 10 s	≥ 30% reduction in airflow that lasted ≥ 10 s and associated with ≥ 4% desaturation	≥ 4%	AASM 2012
Delgado-Vargas 2020 [38]	Spain	STOP-Bang	193	Level 1—Lab PSG	AHI ≥ 5	≥ 90% drop in airflow from baseline, which lasted ≥ 10 s	≥ 30% reduction in airflow that lasted ≥ 10 s and associated with ≥ 4% desaturation	≥ 4%	AASM 2012
Duarte 2017 [39]	Brazil	STOP-Bang	456	Level 1—Lab PSG	AHI ≥ 5	≥ 90% drop in airflow from baseline, which lasted ≥ 10 s	≥ 30% decrease in airflow with 3% oxygen desaturation and/or arousal	≥ 3%	AASM 2012

Table 1 (continued)

Study	Country	Questionnaire	N	Validation tool	OSA definition	Apnoea definition	Hypopnoea definition	% desaturation	Scoring criteria
Duarte 2020 [40]	Brazil	STOP-Bang	3606	Level 1—Lab PSG	AHI≥5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% decrease in airflow with 3% oxygen desaturation and/or arousal	≥3%	AASM 2012
El Sayed 2012 [41]	Egypt	Berlin, STOP, STOP-Bang	234	Level 1—Lab PSG	AHI≥5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% reduction in airflow that lasted≥10 s and associated with≥4% desaturation	≥4%	AASM 2007
Ha 2014 [42]	China	Berlin, STOP, STOP-Bang, ASA checklist	141	Level 1—Lab PSG	RDI>5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% reduction in airflow that lasted≥10 s and associated with≥4% desaturation	≥4%	AASM 2007
Hu 2019 [43]	China	STOP-Bang	196	Level 1—Lab PSG	AHI≥5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% decrease in airflow with 3% oxygen desaturation and/or arousal	3%	AASM 2012
Kashaninasab 2017 [44]	Iran	Berlin, STOP, STOP-Bang	250	Level 1—Lab PSG	AHI≥5	A decrease in airflow that lasted≥10 s	A reduction in airflow by≥50% or a 3% desaturation	≥3%	AASM 2005
Khaledi-Paveh 2016 [45]	Iran	Berlin	100	Level 1—Lab PSG	AHI≥5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% reduction in airflow that lasted≥10 s and associated with≥4% desaturation	≥4%	AASM 2007
Kim 2015 [46]	Korea	Berlin, STOP-Bang, SA-SDQ	592	Level 1—Lab PSG	AHI>5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% reduction in airflow that lasted≥10 s and was associated with≥4% desaturation	≥4%	AASM 2007
Ong 2010 [47]	Singapore	STOP-Bang	314	Level 1—Lab PSG	AHI>5	Cessation of airflow for >10 s	≥30% decrease in airflow with 4% oxygen desaturation or with 3% oxygen desaturation and arousal	Unclear	Rechtschaffen & Kales 1968
Pataka 2016 [48]	Greece	STOP, STOP-Bang, Berlin	204	Level 1—Lab PSG	AHI≥5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% reduction in airflow that lasted≥10 s and was associated with≥4% desaturation	≥4%	AASM 2007
Pecotic 2012 [49]	Croatia	STOP	425	Level 1—Lab PSG	AHI>5	Complete cessation of airflow for≥10 s	A decrease in airflow >50% from baseline for≥10 and≥3% desaturation	≥3%	Not referenced

Table 1 (continued)

Study	Country	Questionnaire	N	Validation tool	OSA definition	Apnoea definition	Hypopnoea definition	% desaturation	Scoring criteria
Pereira 2013 [50]	Canada	Berlin, STOP-Bang	128	Level 1—Lab PSG	AHI≥5	A cessation of airflow ≥50% for≥10 s	A decrease in airflow of>50% for≥10 s followed by ≥3% oxygen desaturation	≥3%	AASM 2007
Perumalsamy 2017 [51]	Chennai	STOP-Bang, Berlin	62	Level 1—Lab PSG	AHI≥5	Not described	Not described	Not described	AASM Manual—not referenced
Sadeghniiat-Haghighi 2015 [52]	Iran	STOP, STOP-Bang	603	Level 1—Lab PSG	AHI>5	Total cessation of airflow for≥10 s	Reduction of airflow for>50% for≥10 s with 3% desaturation or with arousal	3%	AASM 2007
Saleh 2011 [53]	Egypt	Berlin	100	Level 1—Lab PSG	AHI>5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% reduction in airflow that lasted≥10 s and was associated with≥4% desaturation	≥4%	AASM 2007
Sangkum 2017 [54]	USA	STOP, STOP-Bang	208	Level 1—Lab PSG	AHI>5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% decrease in airflow with 3% oxygen desaturation and/or arousal	≥3%	AASM 2012
Suksakorn 2014 [55]	Thailand	Berlin	132	Level 1—Lab PSG	AHI>5	Not reported or referenced	Not reported or referenced	Not reported or referenced	Not reported or referenced
Vana 2013 [56]	USA	STOP-Bang	47	Level 1—Lab PSG	AHI>5	Complete cessation of airflow>10 s In paper	≥30% reduction in airflow that lasted≥10 s and associated with 4% desaturation or arousal	4%	AASM 2007
Yüceege 2015 [57]	Turkey	Berlin	433	Level 1—Lab PSG	AHI>5	≥90% drop in airflow from baseline, which lasted≥10 s	A≥50% drop in flow for≥10 associated with≥3% oxygen desaturation	≥3%	AASM 2007
Surgical population									
Chung 2008a [58]	Canada	Berlin, STOP, ASA checklist	177	Level 1—Lab PSG	AHI>5	Complete cessation of breathing	A decrease of>50% of airflow for≥10 s and>3% desaturation or an arousal	>3%	AASM1999
Chung 2008b [59]	Canada	STOP, STOP-Bang	177	Level 1—Lab PSG	AHI≥5	Complete cessation of breathing	A decrease of>50% of airflow for≥10 s and>3% desaturation or an arousal	≥3%	AASM 1999
Chung 2013 [60]	Canada	STOP-Bang	384	Level 2—portable PSG	AHI>5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% reduction in airflow that lasts≥10 s and≥4% desaturation	≥4%	AASM 2007

Table 1 (continued)

Study	Country	Questionnaire	N	Validation tool	OSA definition	Apnoea definition	Hypopnoea definition	% desaturation	Scoring criteria
Chung 2014 [61]	Canada	STOP-Bang	516	Level 2—portable PSG	AHI>5	≥90% drop in airflow from baseline, which lasted ≥10 s	≥50% reduction in airflow that lasted ≥10 s followed by ≥3% oxygen desaturation	≥3%	AASM 2007
Deflandre 2017 [62]	Belgium	STOP-Bang, OSA50	150	Level 1—Lab PSG	AHI>5	≥90% drop in airflow from baseline, which lasted ≥10 s	≥30% decrease in airflow for ≥10 s with ≥3% oxygen desaturation and/or arousal	≥3%	AASM 2012
Nunes 2015 [63]	Brazil	STOP-Bang, Berlin	81	Level 1—Lab PSG	AHI≥15	Apnoea was defined as complete cessation of airflow for ≥10 s	≥50% reduction in airflow for 10 s associated with desaturation of >3% or an arousal	>3%	AASM 2007
Xia 2018 [64]	China	STOP-Bang	790	Level 1—Lab PSG	AHI≥5	≥90% drop in airflow from baseline, which lasted ≥10 s	≥30% reduction in air flow and ≥4% desaturation	≥4%	AASM 2012
Resistant hypertension									
Giampá 2018 [65]	Brazil	Berlin	119	Level 1—Lab PSG	AHI>5	≥90% drop in airflow from baseline, which lasted ≥10 s	At least 30% decrease in airflow with 3% desaturation and/or arousal	3%	AASM 2007
Margello 2014 [66]	Brazil	Berlin	422	Level 1—Lab PSG	AHI≥5	≥90% drop in airflow from baseline, which lasted ≥10 s	At least 30% decrease of airflow for ≥10 s accompanied by desaturation≥4%	≥4%	AASM 2007
Asthma population									
Lu 2017 [67]	China	Berlin, STOP-Bang	123	Level 1—Lab PSG	AHI≥5	Apnoea was defined as the absence of oronasal airflow for ≥10 s	≥50% reduction in airflow accompanied by oxygen desaturation ≥4%	≥4%	AASM 2007
Community clinic									
Gantner 2010 [68]	China	Berlin	143	Level 1—Lab PSG	AHI≥15	Not considered necessary to distinguish apnoeas from hypopnoeas	>50% reduction in airflow for ≥10 s associated with>3% desaturation or arousal	>3%	AASM 1999
Highway bus drivers									
Firat 2012 [69]	Turkey	Berlin, STOP, STOP-Bang, OSA50	85	Level 1—Lab PSG (daytime)	AHI>15	≥90% decrease in airflow persisting for ≥10 s	≥50% decrease in the airflow with 3% desaturation or arousal, persisting for at least 10 s	≥3%	AASM 2007

Table 1 (continued)

Study	Country	Questionnaire	N	Validation tool	OSA definition	Apnoea definition	Hypnoea definition	% desaturation	Scoring criteria
Neurology population									
ElKholy 2017 [70]	Egypt	Berlin	30	Level 1—Lab PSG	AHI≥5	Not reported	Not reported	Not reported	Rechtschaffen and Kales (1968)
Primary care									
Bouloukaki 2013 [71]	Crete	Berlin	129	Level 1—Lab PSG	AHI≥5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% reduction in airflow that lasted≥10 s and was associated with≥4% desaturation	≥4%	AASM 2007
Respiratory population									
Yunus 2013 [72]	Malaysia	Berlin	144	Level 1—Lab PSG	AHI≥5	Not reported	Not reported	Not reported	Not reported
Snoring clinic									
Banhiran 2014 [73]	Thailand	STOP, STOP-Bang	303	Level 1—Lab PSG	AHI≥5	≥90% drop in airflow from baseline, which lasted≥10 s	≥30% reduction in airflow that lasted≥10 s and was associated with≥4% desaturation	≥4%	AASM 2007

Predictive parameters of the STOP questionnaire (score cut-off ≥ 2)

The prevalence of AHI ≥ 5 (all OSA), AHI ≥ 15 (moderate to severe), and AHI ≥ 30 (severe) OSA was 67%, 58%, and 46% respectively. The pooled sensitivity of the STOP questionnaire to predict all OSA, moderate–severe, and severe OSA was 90% (95% CI: 82%, 95%), 90% (95% CI: 75%, 97%), and 95% (95% CI: 88%, 98%) respectively. The pooled specificity was 31% (95% CI: 15%, 53%), 29% (95% CI: 10%, 61%), and 21% (95% CI: 10%, 39%) respectively. The corresponding DOR were 4.2 (95% CI: 0.8, 7.6), 3.8 (95% CI: 1.7, 5.9), and 4.7 (95% CI: 2.6, 6.8) respectively (Fig. 5, Table 3). Greater uncertainty and variability in specificity were noted in the CI width and scatter of individual study estimates.

Predictive parameters of the STOP-Bang questionnaire (score cut-off ≥ 3)

The prevalence of AHI ≥ 5 (all OSA), AHI ≥ 15 (moderate to severe), and AHI ≥ 30 (severe) OSA was 80%, 59%, and 39%, respectively. The pooled sensitivity of the STOP-Bang questionnaire to predict all OSA, moderate–severe, and severe OSA was 92% (95% CI: 87%, 95%), 95% (95% CI: 92%, 96%), and 96% (95% CI: 93%, 98%) respectively. The pooled specificity was 35% (95% CI: 25%, 46%), 27% (95% CI: 18%, 34%), and 28% (95% CI: 20%, 38%) respectively. The corresponding DOR were 6.0 (95% CI: 4.4, 7.6), 6.4 (95% CI: 3.3, 9.5), and 9.2 (95% CI: 5.9, 12.4) respectively (Fig. 6, Table 3). Greater uncertainty and variability in specificity were noted in the CI width and scatter of individual trial estimates, particularly for AHI ≥ 5.

SROC plots were used to display the results of individual questionnaires in the ROC space, plotting each questionnaire as a single sensitivity–specificity point [24]. When we plotted the SROC for all three questionnaires on the same axes, the confidence regions of the Berlin, STOP, and STOP-Bang questionnaires, for all OSA (AHI ≥ 5) (Fig. 7) and severe OSA (AHI ≥ 30) (Fig. 9), overlapped, suggesting that there was no statistically significant difference in sensitivity among the 3 questionnaires.

Figure 8 shows no overlap of the confidence regions for the Berlin and STOP-Bang questionnaires, suggesting a possible difference in sensitivity between the two questionnaires. A meta-regression model assuming equal variances for logit sensitivity and logit specificity suggested that the expected sensitivity or specificity differed between the two tests (chi-square = 14.1, 2df, $p = 0.0008$) (Fig. 9).

Records identified through database searching
(n = 4922)

| CINAHL PLUS (n = 352) | Scopus (n = 1443) | PubMed (n = 1480) | Web of Sciences (n = 1612) | LILACS (n = 35) |

Additional records identified from other sources
(n = 2439)

- ETHos (n=100)
- OpenGrey (n=28)
- Google Scholar (n=545)
- Proquest (n=1764)
- New York Grey Literature Report (n=2)

Additional records identified from reference lists
(n = 218)

Records after duplicates removed
(n = 1902)

Records screened from titles
(n = 3257)

Records excluded from titles
(n = 2918)

Records screened from abstracts
(n = 339)

Records excluded from abstracts
(n = 241)

Full-text articles assessed for eligibility
(n = 98)

Full-text articles excluded, with reasons
(n = 53)

- Retrospective (n= 10)
- Studies that did not use PSG as reference standard (n=10)
- Study population (in-patient; not a clinical cohort, other) (n= 7)
- Questionnaire not patient-based or unsuitable for meta-analysis: (n=3)
- Conference abstracts (n=1)
- Studies in which the data presented was not possible to be extrapolated (n=22)

Studies included in the systematic review
(n=45)

Studies included in quantitative synthesis (meta-analysis)
(n = 29)

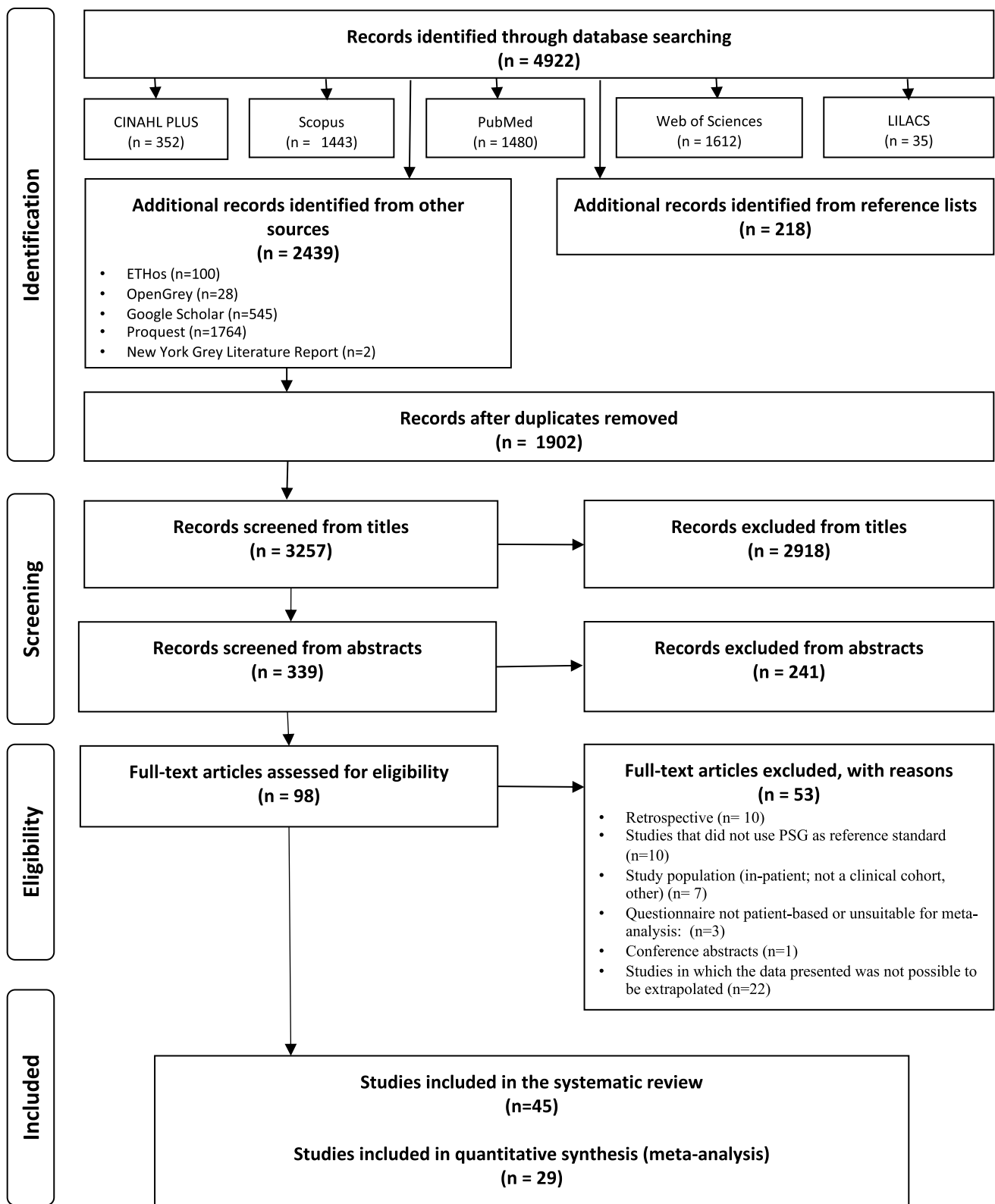Identification | Screening | Eligibility | Included

Fig. 1 Flowchart of search results

Table 2 Demographic data

Study	Age (mean ± SD)	Gender (%), male	BMI (kg/m²)	NC (cm) (mean ± SD)	WC (cm) (mean ± SD)	AHI (mean ± SD)
Sleep clinic population						
Abdullah 2018 [29]	41.22 ± 12.66	63.4	n/a	n/a	n/a	n/a
Alhouqani 2015 [30]	42.87 ± 11.838	77.7	34.9 ± 8.602	39.5 ± 3.463	n/a	34.87 ± 31.273
Amra 2013 [31]	52.3 ± 13.6	55.4	31.5 ± 6	n/a	n/a	37.8 ± 30.8
Amra 2018 [32]	49.29 ± 9.75	58.5	32.4 ± 7.43	40.8 ± 3.13	n/a	n/a
Oktay Arslan 2020 [33]	50.65 ± 11.38	70	32.5 ± 5.9	41.6 ± 3.3	n/a	n/a
Avincsal 2017 [34]	50 ± 0.79	69.8	34.2 ± 0.42	41.3 ± 0.32	n/a	n/a
BaHammam 2015 [35]	46.6 ± 14	61.0	34.4 ± 7.8	38.0 ± 3.81	n/a	50 ± 37
Boynton 2013 [36]	46.3 ± 13.9	44.8	33.4 ± 8.76	39.9 ± 5.4	n/a	n/a
Deflandre 2018 [37]	55.8 ± 14	68.0	31.8 ± 12.07	n/a	n/a	n/a
Delgado-Vargas 2020 [38]	50.42 ± 12.05	73	29.83 ± 6.12	n/a	n/a	n/a
Duarte 2017 [39]	43.7 ± 12.5	63.0	32.1 ± 7.8	40.8 ± 4.3	n/a	24.6 ± 25.2
Duarte 2020 [40]	45.7 ± 14.6	54	32.9 ± 7.7	40.5 ± 4.8	n/a	n/a
El Sayed 2012 [41]	50.38 ± 11.29	85.5	37.8 ± 9.54	42.4 ± 4.26	n/a	45.57 ± 32.74
Ha 2014 [42]	45 ± 11	81.6	26.0 ± 4	n/a	n/a	25 ± 24
Hu 2019 [43]	18–70 (range)	84	27.23 ± 4.03	40.41 ± 3.7	n/a	n/a
Kashaninasab 2017 [44]	48.1 ± 12	76.0	n/a	40.2 ± 3.8	n/a	44.1 ± 31.2
Khaledi-Paveh 2016 [45]	47.8 ± 14.1	60.0	29.5 ± 6.1	n/a	n/a	24 ± 21.5
Kim 2015 [46]	47.6 ± 12.7	83.3	24.7 ± 3.5	39.2 ± 3.5	n/a	25.9 ± 21.8
Ong 2010 [47]	46.8 ± 15	70.5	27.9 ± 6	39.8 ± 4.1	n/a	26.2 ± 26.9
Pataka 2016 [48]	51.8 ± 13.8	77.5	32.8 ± 6.2	41.6 ± 3.9	1 ± 0.4	29.7 ± 24.7
Pecotic 2012 [49]	55	70.0	30.1 ± 4.7	n/a	n/a	31.4 ± 22.6
Pereira 2013 [50]	50 ± 12.3	65.6	31.0 ± 6.6	41.0 ± 4.4	n/a	33.1 ± 28
Perumalsamy 2017 [51]	53	59.7	n/a	n/a	n/a	n/a
Sadeghniiat-Haghighi 2015 [52]	45.8 ± 12.7	74.8	29.2 ± 5.9	39.7 ± 3.6	n/a	n/a
Saleh 2011 [53]	45	51.0	33.1	n/a	n/a	n/a
Sangkum 2017 [54]	52 ± 0.9	36.0	36.9 ± 0.7	40.6 ± 0.4	0.95 ± 0.01	n/a
Suksakorn 2014 [55]	48.15 ± 8.8	68.2	29.2 ± 6.8	n/a	n/a	28 ± 29.7
Vana 2013 [56]	46.4 ± 13.2	34.0	36.3 ± 9.2	n/a	n/a	n/a
Yüceege 2015 [57]	47.5 ± 10.5	65.8	31.1 ± 5.6	39.4 ± 3.9	102.9 ± 12.9	28.27 ± 26.5
Surgical population						
Chung 2008a [58]	55 ± 13	49.7	30.0 ± 6	39.0 ± 6	n/a	20 ± 6
Chung 2008b [59]	55 ± 13	49.7	30.0 ± 6	39.0 ± 6	n/a	20 ± 6
Chung 2013 [60]	60 ± 11	46.0	31.2 ± 7	39.1 ± 4	n/a	n/a
Chung 2014 [61]	59.5 ± 12	54.0	30.6 ± 7	39.0 ± 4	n/a	n/a
Deflandre 2017 [61]	59.66 ± 12.41	70.0	32.4 ± 2.26	42.0 ± 4.64	n/a	n/a
Nunes 2015 [63]	56 ± 07	70.0	29.5 ± 5	n/a	n/a	n/a
Xia 2018 [64]	41.4 ± 10	58.5	28.5 ± 4.7	40.0 ± 4	n/a	13.1 ± 4.4
Resistant hypertension						
Giampá 2018 [65]	52 ± 9	43	52 ± 9	40.0 ± 4	104 ± 14	27 ± 24
Margello 2014 [66]	62.4 ± 9.9	31.3	37.8 ± 3.7	37.8 ± 3.7	101 ± 12.1	n/a
Asthma population						
Lu 2017 [67]	47.56 ± 12.12	57.7	26.4 ± 2.99	36.3 ± 2.97	n/a	15.07 ± 12.87
Community clinic						
Gantner 2010 [68]	62.2 ± 7.6	40.5	26.6 ± 3.7	37.5 ± 4	n/a	n/a
Highway bus drivers						
Firat 2012 [69]	48 ± 5.7	100	29.1 ± 3.8	41.1 ± 2.8	99.6 ± 9.8	21.1 ± 17.4

Table 2 (continued)

Study	Age (mean ± SD)	Gender (%), male	BMI (kg/m^2)	NC (cm) (mean ± SD)	WC (cm) (mean ± SD)	AHI (mean ± SD)
Neurology population						
ElKholy 2017 [70]	50.67 ± 14.94	60	n/a	n/a	n/a	29.11 ± 33.16
Primary care						
Bouloukaki 2013 [71]	47 ± 13	61.9	35.0 ± 25.1	n/a	n/a	41 ± 32
Respiratory population						
Yunus 2013 [72]	44.7 ± 11.5	64	36.3 ± 11	39.3 ± 4.9	94.1 ± 16.9	38.8 ± 31.9
Snoring clinic						
Banhiran 2014 [73]	49.6	61.4	27.5	37.05	n/a	n/a

Surgical population

In the surgical population ($n = 2710$) (Fig. 10), we identified the Berlin, STOP, and STOP-Bang questionnaires for inclusion in the meta-analysis. The ASA checklist and OSA50 questionnaires were excluded from meta-analysis due to an insufficient number of studies. Nunes included two surgical cohorts, abdominal and coronary artery bypass grafting, which were entered as separate cohorts [63].

Predictive parameters of the Berlin questionnaire (score cut-off ≥ 2)

Two studies were included in the meta-analysis of the Berlin Questionnaire for moderate to severe OSA (AHI ≥ 15) (Fig. 11). Due to insufficient data, we were unable to conduct a meta-analysis for all (AHI > 5) and severe OSA (AHI > 30).

The prevalence of moderate to severe OSA or AHI of ≥ 15 was 42%. The pooled sensitivity of the Berlin questionnaire to predict moderate to severe OSA (AHI ≥ 15) was 76% (95% CI: 66%, 84%), and the pooled specificity was 47% (95% CI: 32%, 62%). The DOR was 2.9 (95% CI: 0.2, 5.5) (Table 4).

Predictive parameters of the STOP questionnaire (score cut-off ≥ 2)

Two studies were eligible for inclusion in the STOP questionnaire meta-analysis for moderate to severe OSA (AHI ≥ 15). However, due to insufficient studies and large heterogeneity around the specificity, the STOP questionnaire was excluded from the meta-analysis (Fig. 12).

Predictive parameters of the STOP-Bang questionnaire (score cut-off ≥ 3)

We included 6 studies in the meta-analysis of the STOP-Bang questionnaire for moderate to severe OSA (AHI ≥ 15) (Fig. 13).

The prevalence of AHI ≥ 5 (all OSA), AHI ≥ 15 (moderate to severe), and AHI ≥ 30 (severe) OSA was 72%, 33%, and 21%, respectively. The pooled sensitivity of the STOP-Bang questionnaire to predict all OSA, moderate–severe, and severe OSA was 85% (95% CI: 81%, 88%), 90% (95% CI: 87%, 93%), and 96% (95% CI: 92%, 98%) respectively. The pooled specificity was 40% (95% CI: 30%, 50%), 27% (95% CI: 19%, 37%), and 26% (95% CI: 21%, 46%). The corresponding DOR were 3.6 (95% CI: 2.3, 4.8), 3.4 (95% CI: 1.9, 4.9), and 8.4 (95% CI: 2.7, 14.2), respectively (Table 4). Compared to the Berlin and STOP questionnaires, individual trial estimates of sensitivity appeared to be more homogeneous for the STOP-Bang questionnaire (Figs. 11, 12, and 13).

Predictive performance of STOP-Bang questionnaires at various questionnaire scores

In the surgical population, two of six studies reported data at multiple cut-off points for the STOP-Bang questionnaire for moderate-to-severe OSA (AHI ≥ 15) [62, 63]. Increasing the threshold from 4 to 7 increased specificity from 31% (95% CI: 0.2, 0.4) to 96% (95% CI: 0.89, 0.99) and was greatest at cut-off values ≥ 6 and ≥ 7 (Table 5). However, increase in specificity was at the expense of a reduction in sensitivity.
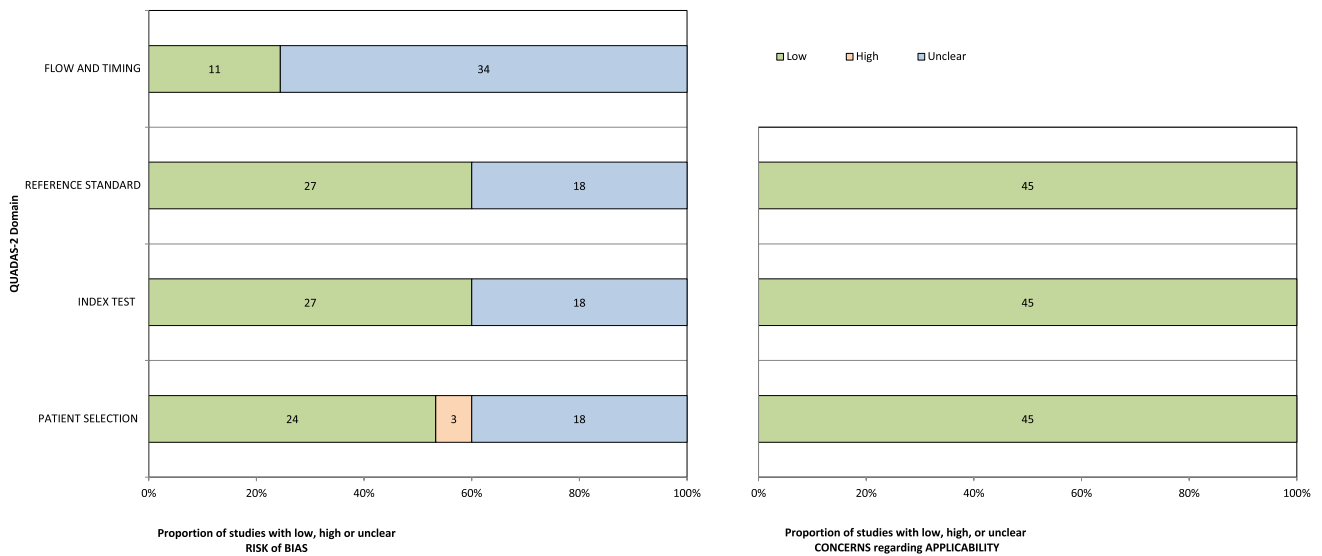
Fig. 2 Risk of bias summary using the QUADAS-2 tool

Resistant hypertension population

We included 2 studies ($n = 517$) in the meta-analysis of the Berlin questionnaire (cut-off ≥ 2) for all OSA (AHI of ≥ 5) [65, 66]. Due to insufficient study data, we were unable to conduct a meta-analysis for moderate–severe (AHI > 15) and severe OSA (AHI > 30).

The prevalence of all OSA or an AHI of ≥ 5 was 80%. The Berlin questionnaire's pooled sensitivity to predict all OSA or AHI of ≥ 5 was 80% (95% CI: 60%, 92%), and the pooled

Fig. 3 Questionnaire studies in sleep clinic population
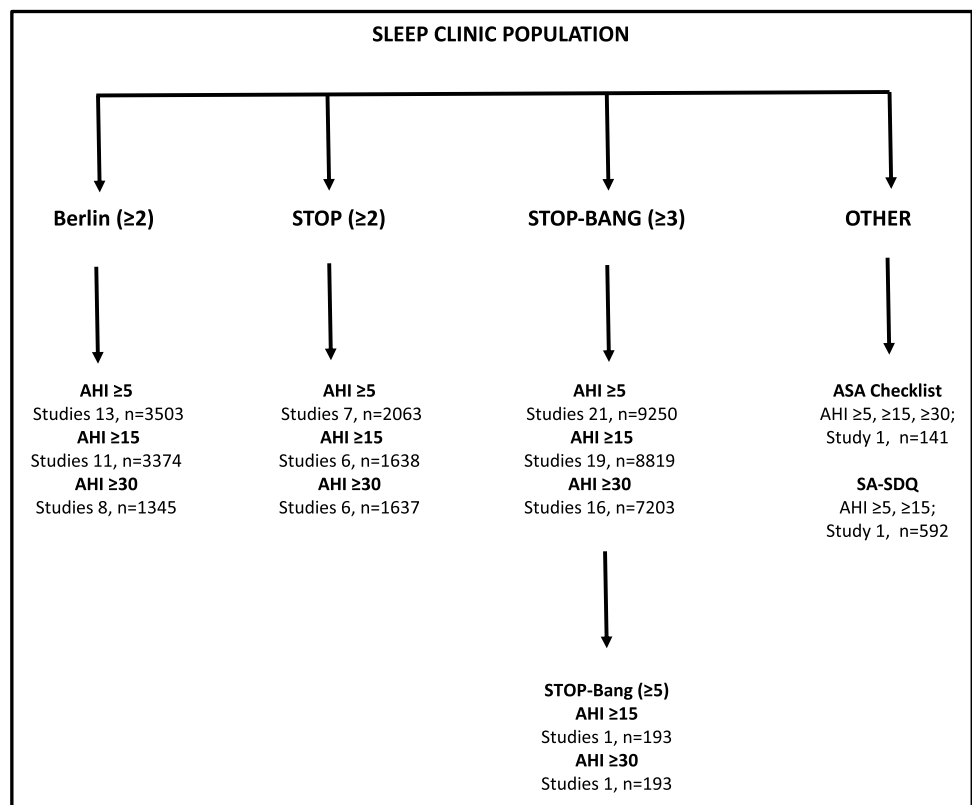
specificity was 36% (95% CI: 21%, 55%). The DOR was 2.2 (95% CI: 0.7, 3.8).

Other cohorts

Asthma, community clinic, highway bus drivers, neurology clinic, primary care, respiratory and snoring clinic cohorts were identified but were excluded from the meta-analysis due to having only one study per cohort (Online Resource 7) [67–73].

Sensitivity analyses

Risk of bias

No studies were evaluated as high risk in the surgical and resistant hypertension populations; therefore, no sensitivity analyses were conducted.

In the sleep clinic population, sensitivity analyses were conducted for the Berlin (Online Resource 8), STOP-Bang (Online Resource 9), and the STOP questionnaires for AHI > 5, AHI ≥ 15, and AHI ≥ 30 (Online Resource 10) excluding studies identified as high risk

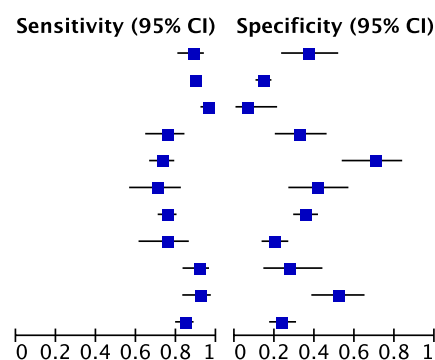**Sleep Clinic: Berlin Questionnaire AHI > 5**

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Amra 2013	331	8	52	9	0.86 [0.83, 0.90]	0.53 [0.28, 0.77]
Amra 2018	121	5	23	8	0.84 [0.77, 0.90]	0.62 [0.32, 0.86]
Arslan 2020	812	72	92	27	0.90 [0.88, 0.92]	0.27 [0.19, 0.37]
El Sayed 2012	203	16	11	5	0.95 [0.91, 0.97]	0.24 [0.08, 0.47]
Ha 2014	84	17	28	11	0.75 [0.66, 0.83]	0.39 [0.22, 0.59]
Kashaninasab 2017	194	2	49	5	0.80 [0.74, 0.85]	0.71 [0.29, 0.96]
Khaledi–Paveh 2016	54	25	16	5	0.77 [0.66, 0.86]	0.17 [0.06, 0.35]
Kim 2015	354	66	141	31	0.72 [0.67, 0.75]	0.32 [0.23, 0.42]
Pataka 2016	19	142	11	32	0.63 [0.44, 0.80]	0.18 [0.13, 0.25]
Pereira 2013	100	9	16	3	0.86 [0.79, 0.92]	0.25 [0.05, 0.57]
Perumalsamy 2017	45	4	9	4	0.83 [0.71, 0.92]	0.50 [0.16, 0.84]
Saleh 2011	67	3	2	28	0.97 [0.90, 1.00]	0.90 [0.74, 0.98]
Suksakorn 2014	87	8	13	24	0.87 [0.79, 0.93]	0.75 [0.57, 0.89]
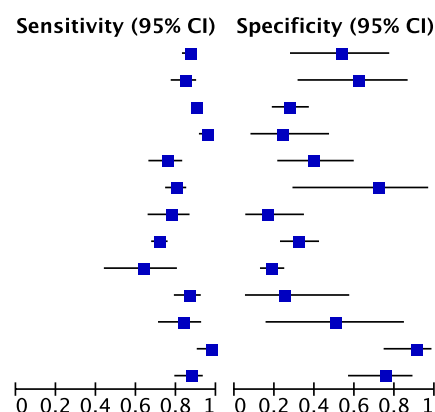
Sleep Clinic: Berlin Questionnaire AHI > 15

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Amra 2013	95	31	13	18	0.88 [0.80, 0.93]	0.37 [0.23, 0.52]
Arslan 2020	603	281	72	47	0.89 [0.87, 0.92]	0.14 [0.11, 0.19]
El Sayed 2012	194	29	9	2	0.96 [0.92, 0.98]	0.06 [0.01, 0.21]
Ha 2014	63	38	21	18	0.75 [0.64, 0.84]	0.32 [0.20, 0.46]
Kashaninasab 2017	153	12	57	28	0.73 [0.66, 0.79]	0.70 [0.53, 0.83]
Khaledi–Paveh 2016	38	27	16	19	0.70 [0.56, 0.82]	0.41 [0.27, 0.57]
Kim 2015	261	159	85	87	0.75 [0.71, 0.80]	0.35 [0.29, 0.42]
Pataka 2016	39	122	13	30	0.75 [0.61, 0.86]	0.20 [0.14, 0.27]
Pereira 2013	80	29	8	11	0.91 [0.83, 0.96]	0.28 [0.15, 0.44]
Suksakorn 2014	66	29	6	31	0.92 [0.83, 0.97]	0.52 [0.38, 0.65]
Yuceege 2015	223	129	42	40	0.84 [0.79, 0.88]	0.24 [0.17, 0.31]

Sleep Clinic: Berlin Questionnaire AHI > 30

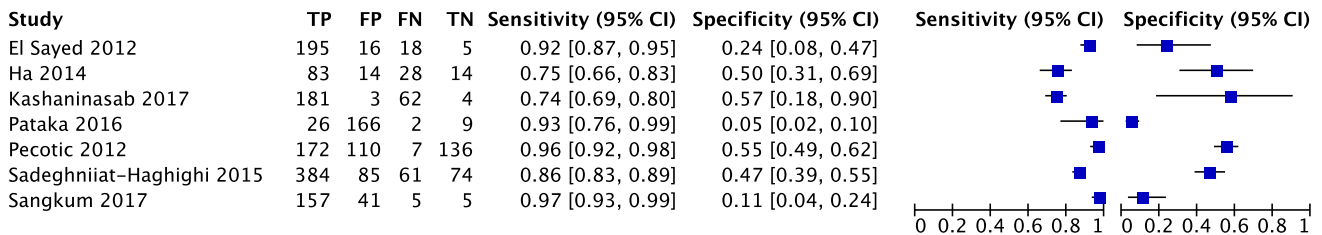| Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Amra 2013	65	61	9	22	0.88 [0.78, 0.94]	0.27 [0.17, 0.37]
El Sayed 2012	165	57	5	7	0.97 [0.93, 0.99]	0.11 [0.05, 0.21]
Ha 2014	41	60	10	29	0.80 [0.67, 0.90]	0.33 [0.23, 0.43]
Kashaninasab 2017	101	21	58	70	0.64 [0.56, 0.71]	0.77 [0.67, 0.85]
Khaledi–Paveh 2016	20	45	7	28	0.74 [0.54, 0.89]	0.38 [0.27, 0.50]
Pataka 2016	80	80	5	39	0.94 [0.87, 0.98]	0.33 [0.24, 0.42]
Pereira 2013	50	59	6	13	0.89 [0.78, 0.96]	0.18 [0.10, 0.29]
Suksakorn 2014	50	45	2	35	0.96 [0.87, 1.00]	0.44 [0.33, 0.55]
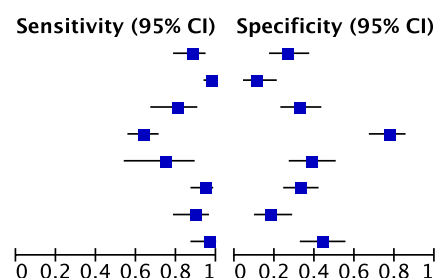
Fig. 4 Forest plots for Berlin questionnaire in sleep clinic population (generated using the software Review Manager 5.3, The Cochrane Collaboration)

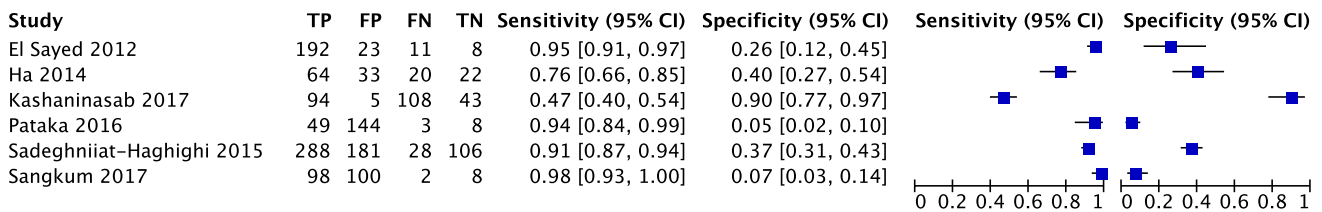**Table 3** Summary statistics for Berlin, STOP, and STOP-Bang questionnaires in the sleep clinic population

Predictive parameters	Berlin AHI ≥ 5	STOP AHI ≥ 5	STOPB AHI ≥ 5	Berlin AHI ≥ 15	STOP AHI ≥ 15	STOPB AHI ≥ 15	Berlin AHI ≥ 30	STOP AHI ≥ 30	STOPB AHI ≥ 30
	13 studies; $n = 3503$	7 studies; $n = 2063$	21 studies; $n = 9250$	11 studies; $n = 3374$	6 studies; $n = 1638$	19 studies; $n = 8819$	8 studies; $n = 1345$	6 studies; $n = 1637$	16 studies; $n = 7203$
Prevalence (% and range)	83.8 (14.71–97.2)	66.94 (13.79–97.20)	79.98 (13.73–97.20)	64 (25.49–86.72)	58.42 (25.49–86.75)	58.78 (25.00–86.75)	50.11 (27.0–72.65)	45.94 (28.37–76.00)	39.25 (28.31–83.0)
Sensitivity (95% CI)	0.848 (0.79, 0.891)	0.904 (0.824, 0.95)	0.919 (0.874, 0.949)	0.843 (0.785, 0.887)	0.903 (0.754, 0.966)	0.945 (0.920, 0.963)	0.886 (0.804, 0.936)	0.945 (0.883, 0.975)	0.959 (0.930, 0.976)
Specificity (95% CI)	0.433 (0.296, 0.582)	0.306 (0.148, 0.528)	0.345 (0.248, 0.457)	0.298 (0.204, 0.413)	0.29 (0.098, 0.606)	0.271 (0.181, 0.384)	0.334 (0.211, 0.458)	0.214 (0.104, 0.391)	0.282 (0.199, 0.384)
False positive rate	0.567 (0.418, 0.704)	0.694 (0.472, 0.852)	0.655 (0.543, 0.752)	0.702 (0.587, 0.796)	0.71 (0.394, 0.902)	0.729 (0.616, 0.819)	0.666 (0.515, 0.789)	0.786 (0.609, 0.896)	0.718 (0.616, 0.801)
Log LR + ve (95% CI)	1.497 (1.066, 1.927)	1.304 (0.970, 1.637)	1.403 (1.232, 1.598)	1.201 (1.049, 1.353)	1.273 (0.904, 1.642)	1.296 (1.125, 1.466)	1.330 (0.110, 1.550)	1.203 (1.027, 1.279)	1.336 (1.184, 1.488)
Log LR − ve (95% CI)	0.350 (0.155–0.546)	0.312 (0.118–0.506)	0.235 (0.183, 0.301)	0.527 (0.361, 0.693)	0.333 (0.194, 0.472)	0.203 (0.123, 0.466)	0.343 (0.210, 0.475)	0.256 (0.154, 0.357)	0.146 (0.095, 0.196)
DOR (95% CI)	4.270 (0.718, 7.822)	4.174 (0.767, 7.581)	5.969 (4.410, 7.529)	2.279 (1.309, 3.249)	3.825 (1.7, 5.949)	6.383 (3.255, 9.511)	3.882 (2.06, 5.704)	4.704 (2.615, 6.794)	9.168 (5.932, 12.405)

AHI apnoea-hypopnoea index, *STOPB* STOP-Bang, *CI* confidence interval, *LR* likelihood ratio, *DOR* diagnostic odds ratio, *LR* likelihood ratio, + positive, − negative.
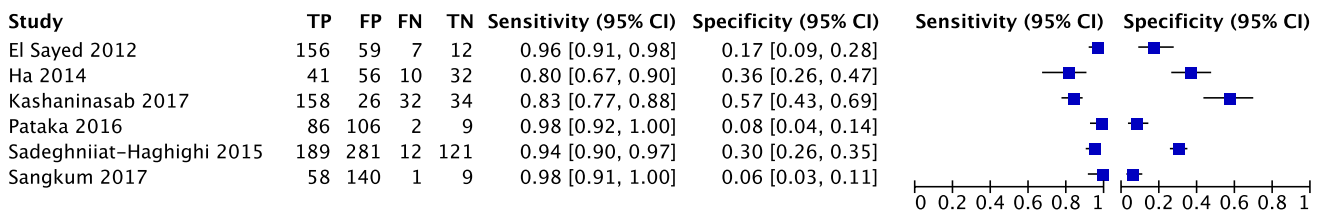
Sleep Clinic: STOP Questionnaire AHI > 5

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
El Sayed 2012	195	16	18	5	0.92 [0.87, 0.95]	0.24 [0.08, 0.47]
Ha 2014	83	14	28	14	0.75 [0.66, 0.83]	0.50 [0.31, 0.69]
Kashaninasab 2017	181	3	62	4	0.74 [0.69, 0.80]	0.57 [0.18, 0.90]
Pataka 2016	26	166	2	9	0.93 [0.76, 0.99]	0.05 [0.02, 0.10]
Pecotic 2012	172	110	7	136	0.96 [0.92, 0.98]	0.55 [0.49, 0.62]
Sadeghniiat–Haghighi 2015	384	85	61	74	0.86 [0.83, 0.89]	0.47 [0.39, 0.55]
Sangkum 2017	157	41	5	5	0.97 [0.93, 0.99]	0.11 [0.04, 0.24]

Sleep Clinic: STOP Questionnaire AHI > 15

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
El Sayed 2012	192	23	11	8	0.95 [0.91, 0.97]	0.26 [0.12, 0.45]
Ha 2014	64	33	20	22	0.76 [0.66, 0.85]	0.40 [0.27, 0.54]
Kashaninasab 2017	94	5	108	43	0.47 [0.40, 0.54]	0.90 [0.77, 0.97]
Pataka 2016	49	144	3	8	0.94 [0.84, 0.99]	0.05 [0.02, 0.10]
Sadeghniiat–Haghighi 2015	288	181	28	106	0.91 [0.87, 0.94]	0.37 [0.31, 0.43]
Sangkum 2017	98	100	2	8	0.98 [0.93, 1.00]	0.07 [0.03, 0.14]

Sleep Clinic: STOP Questioinnaire AHI > 30

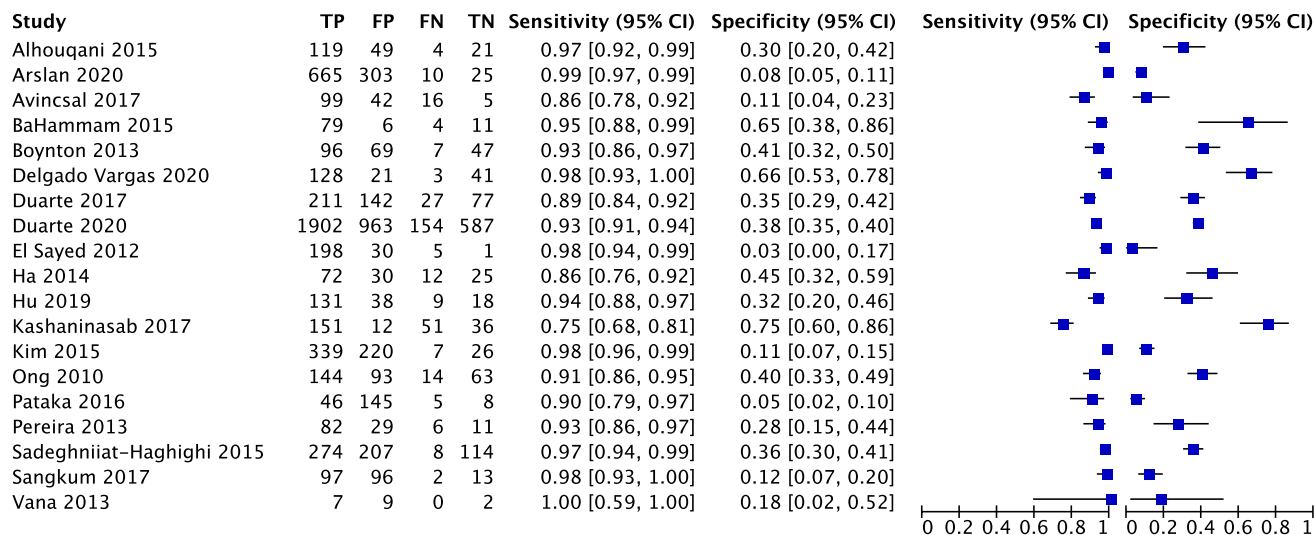| Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
El Sayed 2012	156	59	7	12	0.96 [0.91, 0.98]	0.17 [0.09, 0.28]
Ha 2014	41	56	10	32	0.80 [0.67, 0.90]	0.36 [0.26, 0.47]
Kashaninasab 2017	158	26	32	34	0.83 [0.77, 0.88]	0.57 [0.43, 0.69]
Pataka 2016	86	106	2	9	0.98 [0.92, 1.00]	0.08 [0.04, 0.14]
Sadeghniiat–Haghighi 2015	189	281	12	121	0.94 [0.90, 0.97]	0.30 [0.26, 0.35]
Sangkum 2017	58	140	1	9	0.98 [0.91, 1.00]	0.06 [0.03, 0.11]

Fig. 5 Forest plots for STOP questionnaire in sleep clinic population (generated using the software Review Manager 5.3, The Cochrane Collaboration)
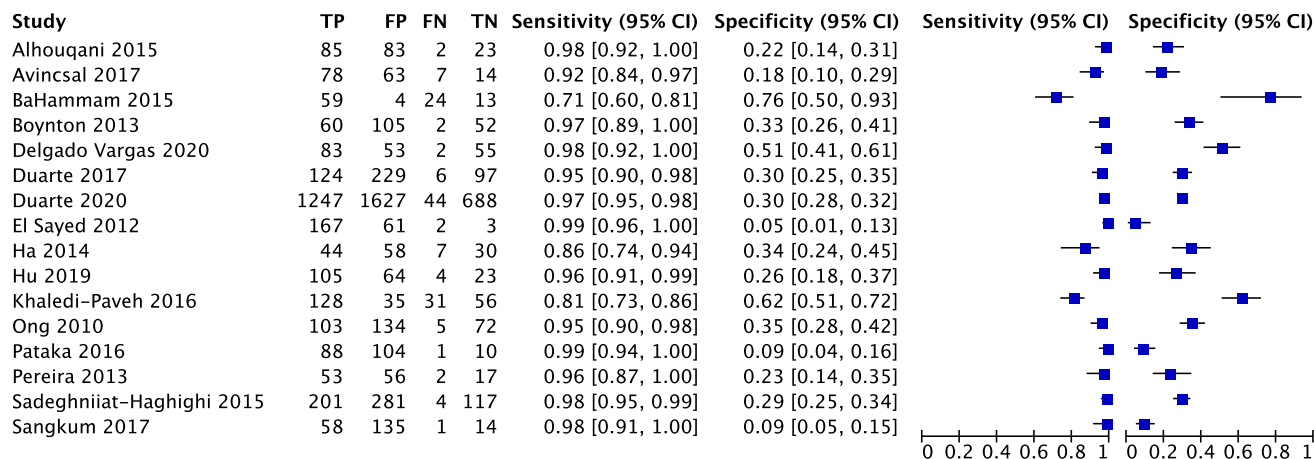
Sleep Clinic: STOP–Bang Questionnaire AHI > 5

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdullah 2017	58	11	38	27	0.60 [0.50, 0.70]	0.71 [0.54, 0.85]
Alhouqani 2015	148	20	16	9	0.90 [0.85, 0.94]	0.31 [0.15, 0.51]
Amra 2018	312	3	71	14	0.81 [0.77, 0.85]	0.82 [0.57, 0.96]
Arslan 2020	665	303	10	25	0.99 [0.97, 0.99]	0.08 [0.05, 0.11]
Avincsal 2017	111	29	18	3	0.86 [0.79, 0.92]	0.09 [0.02, 0.25]
BaHammam 2015	82	13	2	3	0.98 [0.92, 1.00]	0.19 [0.04, 0.46]
Boynton 2013	139	26	30	24	0.82 [0.76, 0.88]	0.48 [0.34, 0.63]
Duarte 2017	298	54	59	45	0.83 [0.79, 0.87]	0.45 [0.35, 0.56]
Duarte 2020	1902	963	154	587	0.93 [0.91, 0.94]	0.38 [0.35, 0.40]
El Sayed 2012	209	15	5	5	0.98 [0.95, 0.99]	0.25 [0.09, 0.49]
Ha 2014	90	12	21	16	0.81 [0.73, 0.88]	0.57 [0.37, 0.76]
Hu 2019	152	17	16	11	0.90 [0.85, 0.94]	0.39 [0.22, 0.59]
Kashaninasab 2017	161	2	82	5	0.66 [0.60, 0.72]	0.71 [0.29, 0.96]
Kim 2015	480	79	15	18	0.97 [0.95, 0.98]	0.19 [0.11, 0.28]
Ong 2010	200	37	36	41	0.85 [0.80, 0.89]	0.53 [0.41, 0.64]
Pataka 2016	27	165	1	11	0.96 [0.82, 1.00]	0.06 [0.03, 0.11]
Pereira 2013	105	7	12	5	0.90 [0.83, 0.95]	0.42 [0.15, 0.72]
Perumalsamy 2017	54	8	0	0	1.00 [0.93, 1.00]	0.00 [0.00, 0.37]
Sadeghniiat-Haghighi 2015	376	105	35	87	0.91 [0.88, 0.94]	0.45 [0.38, 0.53]
Sangkum 2017	156	37	6	9	0.96 [0.92, 0.99]	0.20 [0.09, 0.34]
Vana 2013	30	10	2	5	0.94 [0.79, 0.99]	0.33 [0.12, 0.62]

Sleep Clinic: STOP–Bang Questionnaire AHI > 15

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Alhouqani 2015	119	49	4	21	0.97 [0.92, 0.99]	0.30 [0.20, 0.42]
Arslan 2020	665	303	10	25	0.99 [0.97, 0.99]	0.08 [0.05, 0.11]
Avincsal 2017	99	42	16	5	0.86 [0.78, 0.92]	0.11 [0.04, 0.23]
BaHammam 2015	79	6	4	11	0.95 [0.88, 0.99]	0.65 [0.38, 0.86]
Boynton 2013	96	69	7	47	0.93 [0.86, 0.97]	0.41 [0.32, 0.50]
Delgado Vargas 2020	128	21	3	41	0.98 [0.93, 1.00]	0.66 [0.53, 0.78]
Duarte 2017	211	142	27	77	0.89 [0.84, 0.92]	0.35 [0.29, 0.42]
Duarte 2020	1902	963	154	587	0.93 [0.91, 0.94]	0.38 [0.35, 0.40]
El Sayed 2012	198	30	5	1	0.98 [0.94, 0.99]	0.03 [0.00, 0.17]
Ha 2014	72	30	12	25	0.86 [0.76, 0.92]	0.45 [0.32, 0.59]
Hu 2019	131	38	9	18	0.94 [0.88, 0.97]	0.32 [0.20, 0.46]
Kashaninasab 2017	151	12	51	36	0.75 [0.68, 0.81]	0.75 [0.60, 0.86]
Kim 2015	339	220	7	26	0.98 [0.96, 0.99]	0.11 [0.07, 0.15]
Ong 2010	144	93	14	63	0.91 [0.86, 0.95]	0.40 [0.33, 0.49]
Pataka 2016	46	145	5	8	0.90 [0.79, 0.97]	0.05 [0.02, 0.10]
Pereira 2013	82	29	6	11	0.93 [0.86, 0.97]	0.28 [0.15, 0.44]
Sadeghniiat-Haghighi 2015	274	207	8	114	0.97 [0.94, 0.99]	0.36 [0.30, 0.41]
Sangkum 2017	97	96	2	13	0.98 [0.93, 1.00]	0.12 [0.07, 0.20]
Vana 2013	7	9	0	2	1.00 [0.59, 1.00]	0.18 [0.02, 0.52]

Sleep Clinic: STOP–Bang Questionnaire AHI > 30

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Alhouqani 2015	85	83	2	23	0.98 [0.92, 1.00]	0.22 [0.14, 0.31]
Avincsal 2017	78	63	7	14	0.92 [0.84, 0.97]	0.18 [0.10, 0.29]
BaHammam 2015	59	4	24	13	0.71 [0.60, 0.81]	0.76 [0.50, 0.93]
Boynton 2013	60	105	2	52	0.97 [0.89, 1.00]	0.33 [0.26, 0.41]
Delgado Vargas 2020	83	53	2	55	0.98 [0.92, 1.00]	0.51 [0.41, 0.61]
Duarte 2017	124	229	6	97	0.95 [0.90, 0.98]	0.30 [0.25, 0.35]
Duarte 2020	1247	1627	44	688	0.97 [0.95, 0.98]	0.30 [0.28, 0.32]
El Sayed 2012	167	61	2	3	0.99 [0.96, 1.00]	0.05 [0.01, 0.13]
Ha 2014	44	58	7	30	0.86 [0.74, 0.94]	0.34 [0.24, 0.45]
Hu 2019	105	64	4	23	0.96 [0.91, 0.99]	0.26 [0.18, 0.37]
Khaledi-Paveh 2016	128	35	31	56	0.81 [0.73, 0.86]	0.62 [0.51, 0.72]
Ong 2010	103	134	5	72	0.95 [0.90, 0.98]	0.35 [0.28, 0.42]
Pataka 2016	88	104	1	10	0.99 [0.94, 1.00]	0.09 [0.04, 0.16]
Pereira 2013	53	56	2	17	0.96 [0.87, 1.00]	0.23 [0.14, 0.35]
Sadeghniiat-Haghighi 2015	201	281	4	117	0.98 [0.95, 0.99]	0.29 [0.25, 0.34]
Sangkum 2017	58	135	1	14	0.98 [0.91, 1.00]	0.09 [0.05, 0.15]

Fig. 6 Forest plots for STOP-Bang questionnaire in sleep clinic population (generated using the software Review Manager 5.3, The Cochrane Collaboration)

in any QUADAS-2 domain, unclear in four domains or outliers.

We excluded one study for the STOP questionnaire for AHI > 5 [49], AHI ≥ 15 [44], and AHI ≥ 30 [44]. For the STOP-Bang questionnaire, we excluded five studies for AHI > 5 [29–31, 34, 46] and four studies for AHI ≥ 15 [30, 35, 38, 44] and AHI ≥ 30 [30, 35, 38, 44]. For the Berlin questionnaire AHI > 5 [45, 46, 53, 55] and AHI ≥ 15 [45, 46, 53, 55], we excluded four studies, and for an AHI ≥ 30 [44, 45, 55], we excluded three studies.
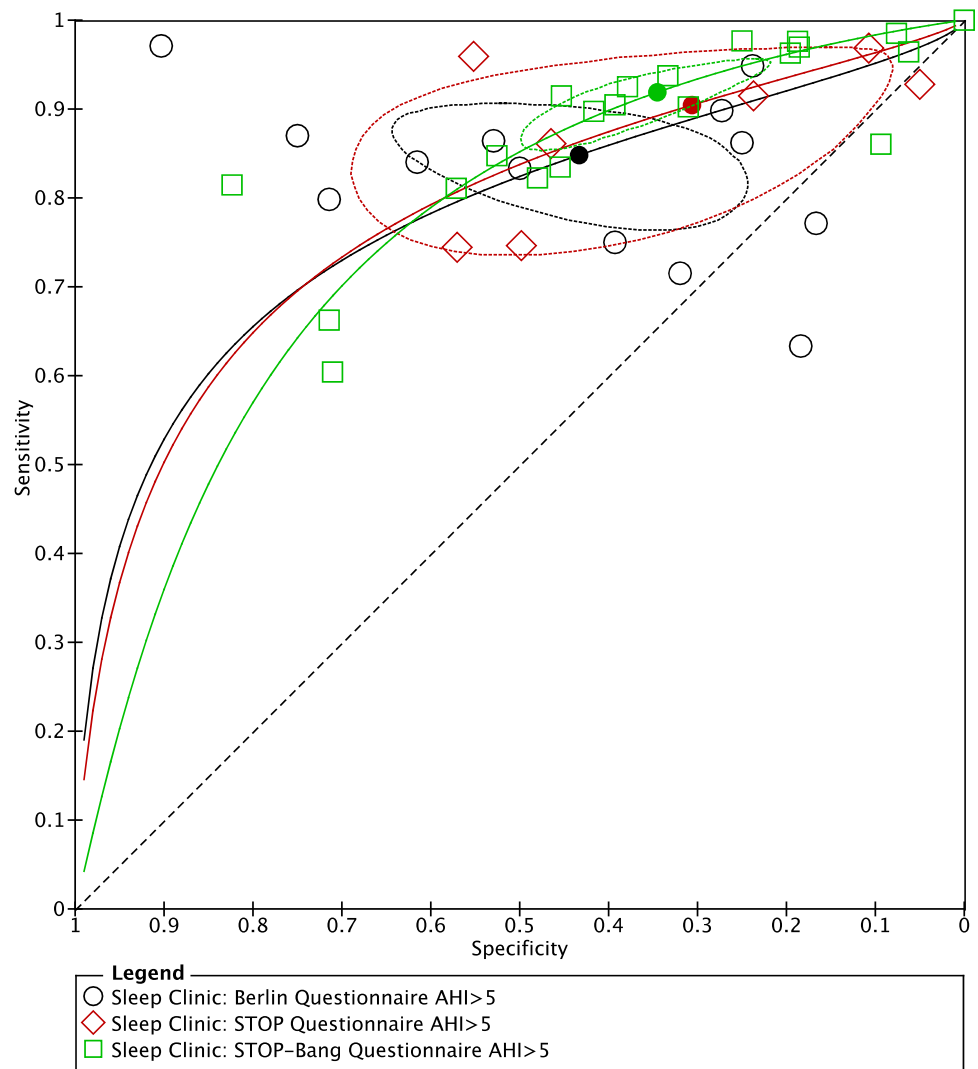
Across all three questionnaires, exclusion of studies was associated with stable or slightly increased sensitivity. In contrast, sensitivity analysis was associated with reduced specificity (Online Resources 8–10). The STOP-Bang questionnaire remained the most effective questionnaire with the highest sensitivity compared to the Berlin and STOP questionnaires. Specificity among all three questionnaires remained low.

Desaturation and arousal criteria

Due to an insufficient number of studies, no sensitivity analysis was conducted in the resistant hypertension population.

In the surgical population, the Berlin and STOP questionnaire studies utilised the ≥ 3% desaturation scoring criteria; therefore, no sensitivity analyses were conducted. For the STOP-Bang questionnaire, studies applied either ≥ 3% or ≥ 4% desaturation criteria. When we applied the ≥ 3% desaturation criteria to the STOP-Bang questionnaire, we excluded one study for AHI > 5 [60], two studies for AHI ≥ 15 [60, 64], and one study for AHI ≥ 30 [60]. In turn, when we applied the ≥ 4% desaturation criteria, we excluded four studies for AHI ≥ 15 [59, 61–63]. Across the three AHI thresholds, sensitivity remained stable, compared to a stable or slightly decreased sensitivity with application of the ≥ 3% desaturation criteria. For AHI ≥ 15, application of the ≥ 4% desaturation criterion was associated with a slight reduction

Fig. 7 Summary ROC for Berlin, STOP, and STOP-Bang questionnaires AHI ≥ 5 (generated using the software Review Manager 5.3, The Cochrane Collaboration)

Legend
○ Sleep Clinic: Berlin Questionnaire AHI > 5
◇ Sleep Clinic: STOP Questionnaire AHI > 5
□ Sleep Clinic: STOP–Bang Questionnaire AHI > 5

in sensitivity and an increase in specificity (Online Resource 11).

We conducted a sensitivity analysis in the sleep clinic population for the Berlin, STOP, and STOP-Bang questionnaires, applying both the $\geq 3\%$ and $\geq 4\%$ desaturation criteria respectively. Studies were excluded on the basis of high risk of bias, scoring criteria not specified, and desaturation criteria ($\geq 3\%$ or $\geq 4\%$) (Online Resource 12).

Across all three questionnaires in the sleep clinic population, exclusion of studies was associated with stable sensitivity and reduced specificity, particularly when applying the $\geq 4\%$ desaturation criterion (Online Resources 13, 14, 15). Overall, the STOP-Bang questionnaire remained the most effective questionnaire with the highest sensitivity compared to the Berlin and STOP questionnaires. Specificity among all three questionnaires remained low.

Discussion

This systematic review and meta-analysis investigated questionnaires' accuracy and clinical utility as screening tools for OSA in adults in different clinical cohorts.

Consistent with previous studies, our findings showed that the STOP-Bang questionnaire (score cut-off ≥ 3) suggested the highest sensitivity to detect OSA and the highest diagnostic odds ratio in both the sleep clinic and surgical populations [12, 18, 76]. However, the STOP-Bang questionnaire was limited by consistently low specificity across all AHI thresholds, resulting in high false positive rates. The Berlin questionnaire (score cut-off ≥ 2) appeared to be the least useful, demonstrating overall low sensitivity and low specificity across all three cohorts [12, 18, 77]. Although there was no comparison with other

Fig. 8 Summary ROC for Berlin, STOP, and STOP-Bang questionnaires AHI ≥ 15 (generated using the software Review Manager 5.3, The Cochrane Collaboration)
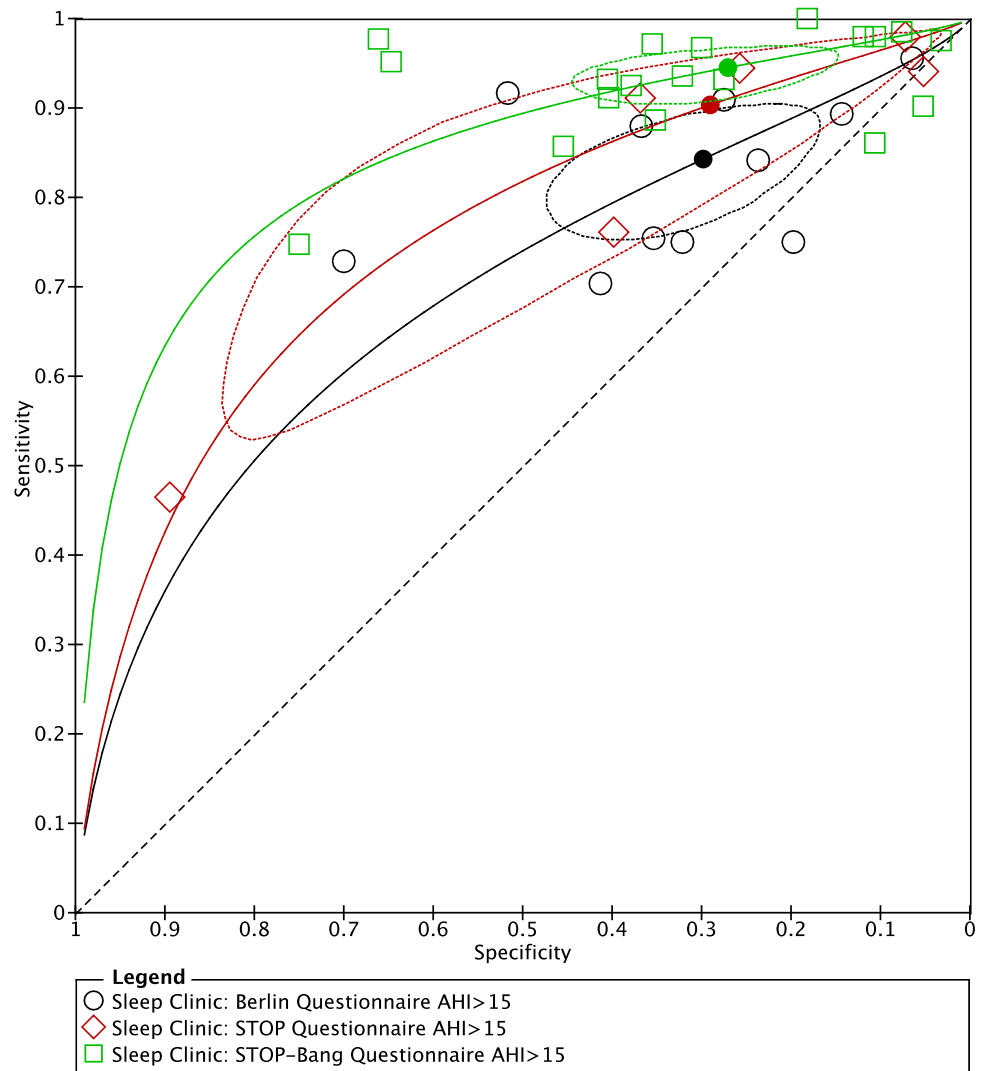
questionnaires in the resistant hypertension cohort, findings were comparable with the sleep clinic and surgical cohorts.

OSA screening questionnaires are intended to provide the information required to identify patients most likely to benefit from downstream management decisions, such as onward referral for objective sleep testing and possible treatment following a positive full diagnostic test. The potential utility of OSA screening questionnaires in risk stratification of patients has been demonstrated in several cohorts. Not only has OSA been associated with risk of peri-operative complications and consequent longer length of hospital stay, but it has also been linked to poor clinical outcomes including higher rates of post CABG atrial fibrillation [78–80]. In the context of the ongoing coronavirus disease 2019 (Covid-19) pandemic, a recent study reported worse clinical outcomes in patients with Covid-19 classified by the Berlin questionnaire as high risk, compared to those at low risk, of

OSA [81]. The study also highlighted the challenges with objective assessment of OSA with PSG during the Covid-19 pandemic, emphasising the need for alternative approaches beyond PSG, such as validated screening questionnaires. In this context, we would encourage the assessment and validation of OSA screening questionnaires, in particular STOP-Bang, as screening tools for risk stratification appropriate clinical settings, with the aim of improving outcomes for patients.

Although sensitivity and specificity provide us with the necessary information to discern between the available screening questionnaires, the clinical value and application of the screening questionnaires are demonstrated by means of the positive and negative predictive values which are dependent on the prevalence of the disease in the given clinical population. Although we were unable to pool the predictive values of individual questionnaires due to variation in prevalence across studies, the point estimates of PPV and

Fig. 9 Summary ROC for Berlin, STOP, and STOP-Bang questionnaires AHI ≥ 30 (generated using the software Review Manager 5.3, The Cochrane Collaboration)
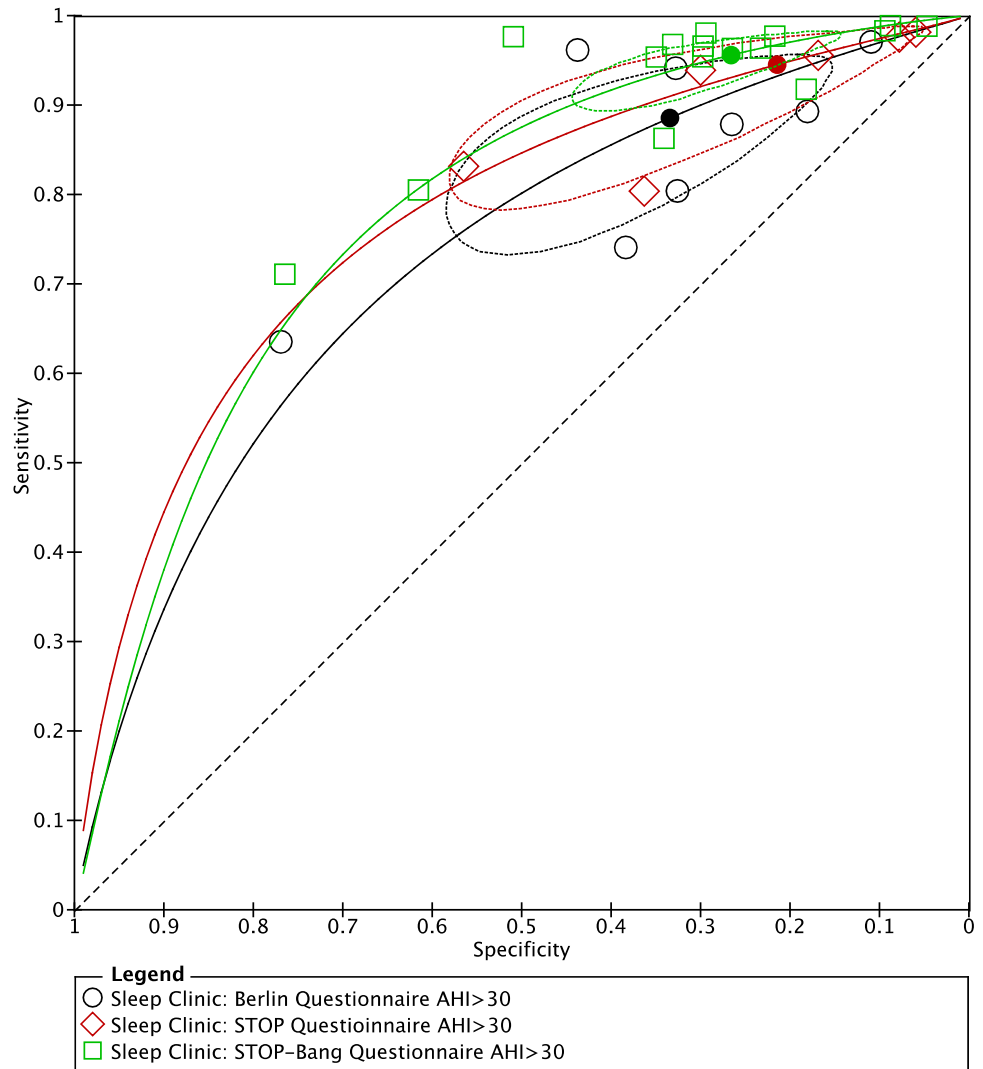
Fig. 10 Questionnaire studies in surgical population
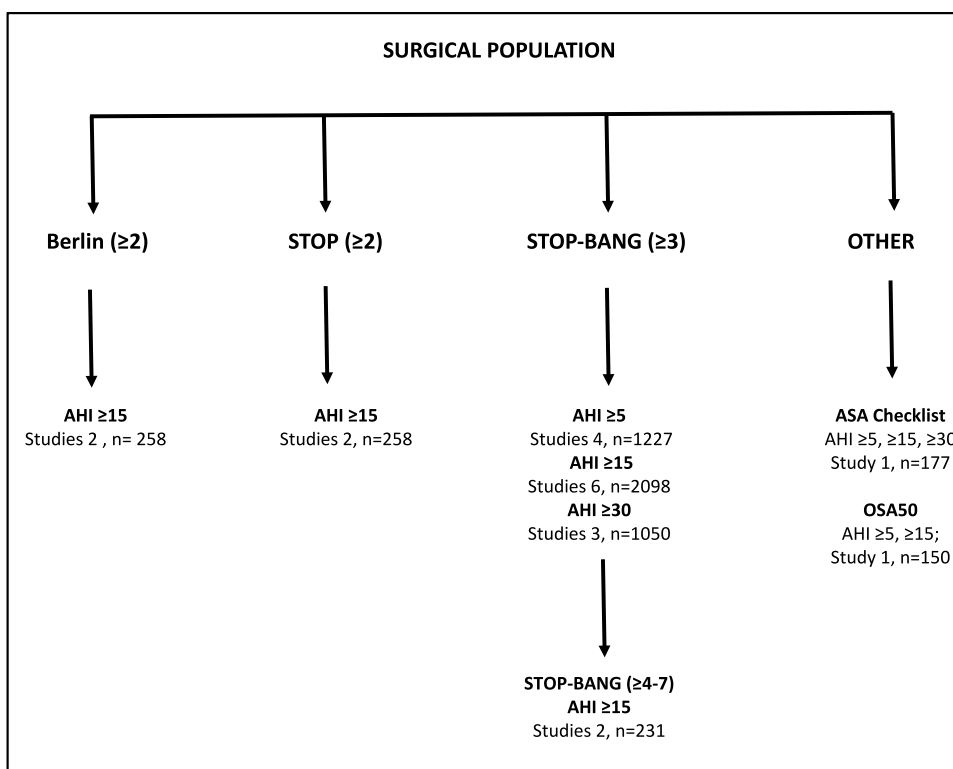
negative predictive value (NPV) for the STOP-Bang questionnaire in both the sleep clinic and surgical population (Online Resource 16) demonstrated an increase in NPV as OSA severity increases. The combination of high sensitivity and NPV of the STOP-Bang questionnaire is therefore useful to help clinicians exclude patients with low risk of clinically significant OSA.

At the same time, the low specificity of the STOP-Bang questionnaire (and therefore its relative inability to correctly identify patients without OSA) leads to a high rate of false positive findings; this may have emotional and cognitive implications for individual patients with added consequences for clinical services, not least cost [80, 82].

This systematic review's main strength lies in our comprehensive literature search with stringent eligibility criteria to identify all relevant studies reporting on the accuracy and clinical utility of existing OSA screening questionnaires that were validated against the gold standard PSG. Our inclusion of the LILACS database expanded our search to include Latin America and the Caribbean studies. Of previous reports, the review by Ramachandran [17] was limited to a search of two databases, English publications only, and omitted any grey literature sources in their search strategy. Additionally, it was unclear if Ross [16] and Abrishami [12] included any grey literature sources in their searches.

Two independent reviewers completed data extraction, and we used the QUADAS-2 tool to assess rigorously all included studies for risk of bias. To evaluate the robustness of the meta-analysis, we conducted sensitivity analyses to investigate the potential influence on our findings from

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Chung 2008a	55	53	15	54	0.79 [0.67, 0.87]	0.50 [0.41, 0.60]		
Nunesa 2015	14	9	3	15	0.82 [0.57, 0.96]	0.63 [0.41, 0.81]		
Nunesc 2015	14	14	7	5	0.67 [0.43, 0.85]	0.26 [0.09, 0.51]		
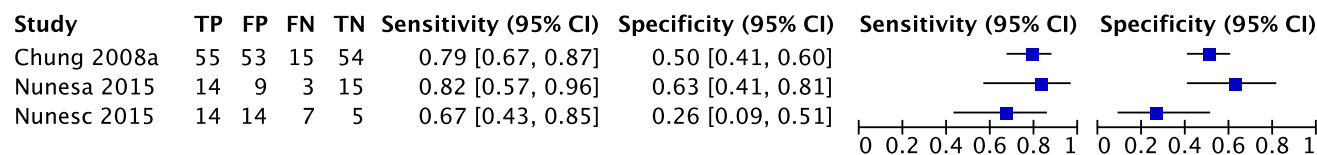
Fig. 11 Forest plot for Berlin questionnaire in surgical population for AHI ≥ 15 (generated using the software Review Manager 5.3, The Cochrane Collaboration)

Table 4 Pooled predictive parameters of Berlin and STOP-Bang questionnaires in surgical population for AHI ≥ 15

Parameter (95% CI)	Berlin, AHI ≥ 15 2 studies, n = 258	STOP-Bang, AHI > 5 4 studies, n = 1227	STOP-Bang, AHI ≥ 15 6 studies, n = 2098	STOP-Bang, AHI ≥ 30 3 studies, n = 1050
Prevalence (% and range)	41.86 (39.55–52.50)	71.56 (67.05–89.33)	32.79 (17.22–62.00)	21.24 (16.67–42.00)
Sensitivity (95% CI)	0.764 (0.661, 0.843)	0.846 (0.811, 0.876)	0.903 (0.871, 0.927)	0.96 (0.924, 0.979)
Specificity (95% CI)	0.468 (0.320, 0.623)	0.394 (0.298, 0.498)	0.269 (0.189, 0.367)	0.261 (0.232, 0.292)
False positive rate (95% CI)	0.532 (0.377, 0.680)	0.606 (0.502, 0.702)	0.731 (0.633, 0.811)	0.739 (0.708, 0.768)
Log LR + ve (95% CI)	1.437 (0.929, 1.945)	1.395 (1.187, 1.603)	1.235 (1.093, 1.377)	1.299 (1.236, 1.362)
Log LR − ve (95% CI)	0.504 (0.206, 0.802)	0.391 (0.304, 0.479)	0.361 (0.235, 0.487)	3.169 (2.502, 3.836)
DOR (95% CI)	2.849 (0.194, 5.505)	3.564 (2.311, 4.817)	3.421 (1.892, 4.949)	8.406 (2.65, 14.162)

CI confidence interval, *LR* likelihood ratio, *DOR* diagnostic odds ratio, + positive, − negative.

studies at high, or unclear, risk of bias. Although our study did not explore source differences from an ethnicity or geographical perspective, we conducted a further sensitivity analysis to evaluate the impact of varying scoring criteria on our study findings. The utilisation of different AASM scoring criteria and desaturation (and arousal) thresholds across studies created a source of variability [83–85]. Although the definition for apnoeas remained stable, there has been much controversy about the definition of hypopnoeas, specific to flow reduction, oxygen desaturation, and the presence or absence of arousal [86]. Varying definitions of hypopnoea not only impacts on prevalence estimates but is likely to underestimate OSA in patients who may benefit from treatment [86]. A study by Guilleminault et al. (2009) showed that by using the 30% flow reduction and 4% desaturation without arousal criteria would have missed 40% of patients who were identified using the criteria with arousal and who were responsive to CPAP therapy with reduction in AHI and symptomatic improvement [87].

On this background, our review is based on a larger number of studies than prior analyses [12, 16, 17]. Although the review by Chiu [18] encompassed a larger dataset, that report carried a greater risk of bias due to the inclusion of retrospective studies and studies that used PSG and portable monitoring as the reference standard.

This review considered all existing OSA screening questionnaires for inclusion. In contrast, Chui [18] pre-selected four questionnaires, including the ESS, which was not developed as a screening questionnaire, but as a measure of daytime sleepiness.

Similar to Abrishami [12] and Chui [18], our review focused on questionnaires only, in contrast to Ross [16] and Ramachandran [17], who also included portable monitoring and clinical prediction tools, respectively.
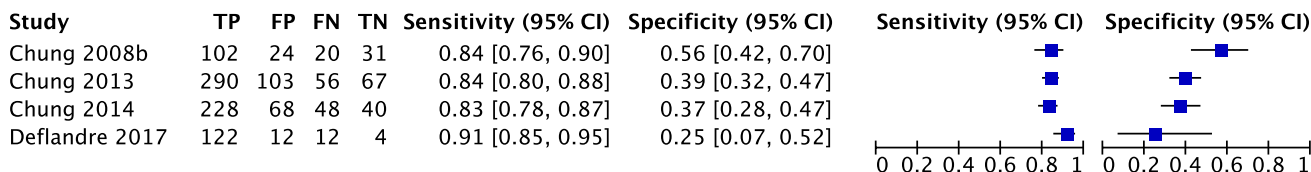
There are a number of limitations to this work. Our findings are influenced by the limitations of the included studies. In several, the true risk of bias was unclear in several of the QUADAS-2 domains due to underreporting in the index test, reference standard, and flow and timing domains. Similarly, it was often unclear if the results of the index test and the reference standard were interpreted independently. Very few studies provided adequate information to determine if the time interval between the index test and the reference standard was appropriate.

Our decision to exclude seven additional clinical cohorts may be considered a limitation; however, in the context of unclear, and possibly substantial, differences among these studies in the patient spectrum and disease
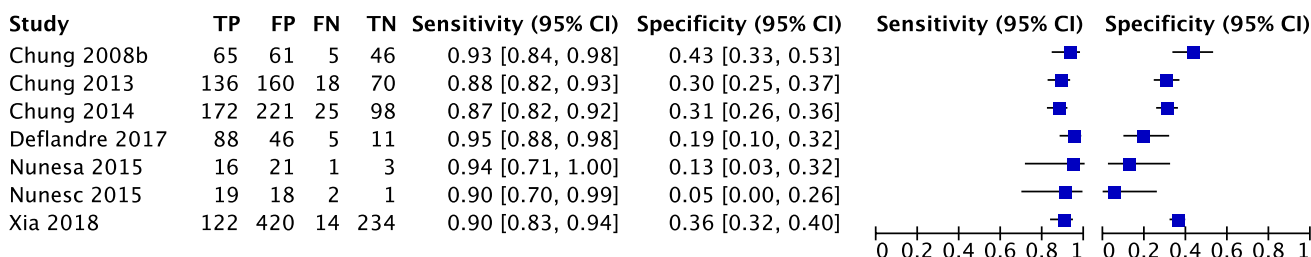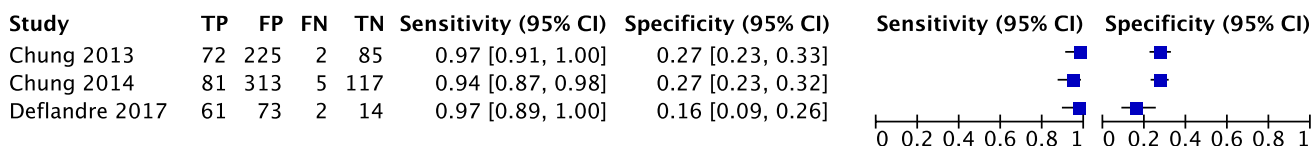
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Chung 2008b	52	50	18	57	0.74 [0.62, 0.84]	0.53 [0.43, 0.63]		
Nunesa 2015	15	21	2	3	0.88 [0.64, 0.99]	0.13 [0.03, 0.32]		
Nunesc 2015	22	0	18	0	0.55 [0.38, 0.71]	Not estimable		
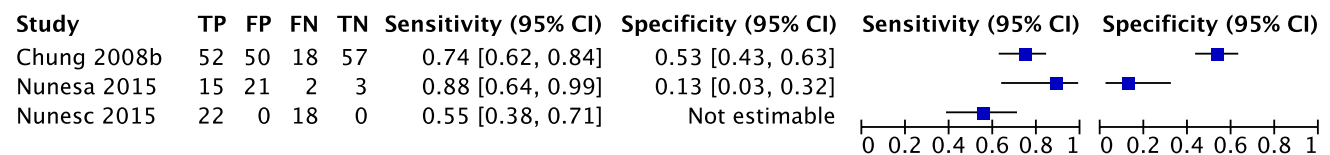
Fig. 12 Forest plot for STOP questionnaire in surgical population for AHI ≥ 15 (generated using the software Review Manager 5.3, The Cochrane Collaboration)

Surgical: STOP–Bang Questionnaire(>3): AHI>5

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chung 2008b	102	24	20	31	0.84 [0.76, 0.90]	0.56 [0.42, 0.70]
Chung 2013	290	103	56	67	0.84 [0.80, 0.88]	0.39 [0.32, 0.47]
Chung 2014	228	68	48	40	0.83 [0.78, 0.87]	0.37 [0.28, 0.47]
Deflandre 2017	122	12	12	4	0.91 [0.85, 0.95]	0.25 [0.07, 0.52]

Surgical: STOP–Bang Questionnaire (>3) AHI >15

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chung 2008b	65	61	5	46	0.93 [0.84, 0.98]	0.43 [0.33, 0.53]
Chung 2013	136	160	18	70	0.88 [0.82, 0.93]	0.30 [0.25, 0.37]
Chung 2014	172	221	25	98	0.87 [0.82, 0.92]	0.31 [0.26, 0.36]
Deflandre 2017	88	46	5	11	0.95 [0.88, 0.98]	0.19 [0.10, 0.32]
Nunesa 2015	16	21	1	3	0.94 [0.71, 1.00]	0.13 [0.03, 0.32]
Nunesc 2015	19	18	2	1	0.90 [0.70, 0.99]	0.05 [0.00, 0.26]
Xia 2018	122	420	14	234	0.90 [0.83, 0.94]	0.36 [0.32, 0.40]

Surgical: STOP–Bang Questionnaire (>3) AHI >30

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chung 2013	72	225	2	85	0.97 [0.91, 1.00]	0.27 [0.23, 0.33]
Chung 2014	81	313	5	117	0.94 [0.87, 0.98]	0.27 [0.23, 0.32]
Deflandre 2017	61	73	2	14	0.97 [0.89, 1.00]	0.16 [0.09, 0.26]

Fig. 13 Forest plots for STOP-Bang questionnaire in surgical population (generated using the software Review Manager 5.3, The Cochrane Collaboration)

prevalence, we felt it appropriate not to include these in the meta-analysis. Because the accuracy of screening tools varies according to the spectrum of disease, this further reiterates the need for validation studies in similar clinical cohorts.

There was a high degree of heterogeneity among included studies with the possibility of selection bias, especially in the sleep clinic population. Consequently, reported sensitivity estimates will be higher than lower-risk populations, making it difficult to extrapolate the true utility of the questionnaire in clinical practice.

In conclusion, our review investigated the accuracy and clinical utility of existing OSA screening questionnaires in different clinical cohorts. While the STOP-Bang questionnaire had a high sensitivity to detect OSA in both the sleep clinic and surgical cohorts, it lacked adequate specificity. This review highlights the issue of low specificity across OSA screening questionnaires. Research is required

Table 5 STOP-Bang questionnaire at different questionnaire cut-offs for moderate–severe OSA (AHI ≥ 15) in the surgical population

Predictive parameters	SBQ ≥ 4, AHI ≥ 15 2 studies; $n=231$	SBQ ≥ 5, AHI ≥ 15	SBQ ≥ 6, AHI ≥ 15	SBQ ≥ 7, AHI ≥ 15
Prevalence (% and range)	56.71 (41.46–62)			
Sensitivity (95% CI)	0.893 (0.828, 0.936)	0.640 (0.480, 0.774)	0.372 (0.211, 0.567)	0.120 (0.033, 0.353)
Specificity (95% CI)	0.310 (0.227, 0.407)	0.575 (0.458, 0.684)	0.807 (0.676, 0.894)	0.958 (0.889, 0.985)
False positive rate (95% CI)	0.690 (0.593, 0.773)	0.425 (0.316, 0.542)	0.193 (0.106, 0.324)	0.042 (0.015, 0.111)
Log LR + ve (95% CI)	1.294 (1.108, 1.481)	1.505 (1.052, 1.958)	1.930 (0.859, 3.000)	2.875 (−2.580, 7.647)
Log LR − ve (95% confidence interval)	0.345 (0.147, 0.543)	0.627 (0.372, 0.881)	2.481 (0.547, 4.414)	0.918 (0.756, 1.080)
DOR (95% CI)	3.755 (1.135, 6.374)	2.402 (0.763, 4.041)	8.406 (2.650, 14.162)	3.131 (−2.580, 8.841)

CI confidence interval, *LR* likelihood ratio, *DOR* diagnostic odds ratio, + positive, − negative, *OSA* obstructive sleep apnoea, *AHI* apnoea-hypopnoea index.

to explore reasons for low specificity and strategies for improvement, ideally without reducing sensitivity. The validation of screening questionnaires in sleep clinic populations is limited by possible selection and spectrum bias, reiterating the need for diagnostic validation studies in clinically similar cohorts. Additionally, further research is needed in resistant hypertension and other at-risk populations that we could not include in the meta-analysis. Improvement in the conduct and reporting of diagnostic validation studies must ensure quality and low risk of bias.

Finally, to enable the extrapolation of the true accuracy and clinical utility of screening questionnaires, validation studies of high methodological quality in comparable, clinically relevant cohorts are required.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

1. Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MSM, Morrell MJ, Nunex CM, Patel SR, Penzel T, Pepin JL, Peppard P, Sinha S, Tufik S, Valentine K, Malhotra A (2019) Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respir Med 7(8):687–698. https://doi.org/10.1016/S2213-2600(19)30198-5

2. Shahar E, Whitney CW, Redline S, Lee ET, Newman AB, Nieto FJ, O'Connor GT, Boland LL, Schwartz JE, Samet JM (2001) Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the Sleep Heart Health Study. Am J Respir Crit Care Med 163(10):19–25. https://doi.org/10.1164/ajrccm.163.1.2001008

3. Pepperell JCT, Davies RJO, Stradling JR (2002) Systemic hypertension and obstructive sleep apnoea. Sleep Med Rev 6(3):157–173. https://doi.org/10.1053/smrv.2001.0189

4. Bradley TD, Floras JS (2009) Obstructive sleep apnoea and its cardiovascular consequences. Lancet 373(9657):82–93. https://doi.org/10.1016/S0140-6736(08)61622-0

5. Kasai T, Bradley TD (2011) Obstructive sleep apnea and heart failure: pathophysiologic and therapeutic implications. J Am Coll Cardiol 57(2):119–127. https://doi.org/10.1016/j.jacc.2010.08.627

6. Drager LF, Togeiro SM, Polotsky VY, Lorenzi-Filho G (2013) Obstructive sleep apnoea: a cardiometabolic risk in obesity and the metabolic syndrome. J Am Coll Cardiol 62:569–76. https://doi.org/10.1016/j.jacc.2013.05.045

7. Ayas, N.T., Taylor, C.M. & Laher, I. (2016) Cardiovascular consequences of OSA. Curr Opinion Cardiol. (6), 599–605. Available from: https://doi.org/10.1097/HCO.0000000000000329

8. Moore, A. (2012) Primary care supplement: the big sleep problem. Health Service J. Available from: https://www.hsj.co.uk/supplement-archive/primary-care-supplement-the-big-sleep-problem/5065470.article. Accessed 17 Mar 2017

9. NHS North of England Specialist Commissioning Group (NW office) 2012: Sleep-related breathing disorders

10. Morsy, N.E., Farrag, N.S., Zaki, N.F.W., Badasy, A.Y., Abdelhafez, S.A., El-Gilany, a.H., El Shafey, M.M., Pandi-Perumal, S.R., Spence, D.W. & BaHammam, A.S. (2019) Obstructive sleep apnea: personal, societal, public health, and legal implications. Rev Environ Health. 34(2):153–169. https://doi.org/10.1515/reveh-2018-0068

11. Rejón-Parrilla, J.C., Garau, M. and Sussex, J. (2014) Obstructive sleep apnoea health economics. Consulting Report for the British Lung Foundation. Available from: https://www.ohe.org/publications/obstructive-sleep-apnoea-health-economics-report. Accessed 30 Jun 2017

12. Abrishami A, Khajehdehi A, Chung F (2010) A systematic review of screening questionnaires for obstructive sleep apnea/Une revue méthodique des questionnaires de dépistage de l'apnée obstructive du sommeil. Can J Anesth 57:423–438. https://doi.org/10.1007/s12630-010-9280-x

13. Costa LE, Uchôa CHG, Harmon RR, Bortolotto LA, Lorenzi-Filho G, Drager LF (2015) Potential underdiagnosis of obstructive sleep apnoea in the cardiology outpatient setting. Heart. 101(16):1288–1292. https://doi.org/10.1136/heartjnl-2014-307276

14. Malaweera A.S.N, Riley J.P., Simonds, A.D.& Cowie, M.C. (2015) Barriers to diagnosis and treatment of sleep disordered breathing in patients with heart failure: perception and experiences of healthcare professionals. Eur J Heart Failure. 17, 171. Available from: ISSN: 1388–9842

15. Rowley JA, Aboussouan LS, Badr MS (2000) The use of clinical prediction formulas in the evaluation of obstructive sleep apnea. Sleep. 23(7):929–938. https://doi.org/10.1093/sleep/23.7.929

16. Ross SD, Sheinhait IA, Harrison KJ, Kvasz M, Connelly JE, Shea SA, Allen IE (2000) Systematic review and meta-analysis of the literature regarding the diagnosis of sleep apnea. Sleep. 23(4):1–14 (PMID: 10875559)

17. Ramachandran SK, Josephs LA (2009) A meta-analysis of clinical screening tests for obstructive sleep apnea. Anesthesiology. 110:928–939. https://doi.org/10.1097/ALN.0b013e31819c47b6

18. Chiu H-Y, Chen P-Y, Chuang L-P, Chen N-H, Tu Y-K, Hsieh Y-J, Wang Y-C, Guilleminault C (2017) Diagnostic accuracy of the Berlin questionnaire, STOP-BANG, STOP, and Epworth sleepiness scale in detecting obstructive sleep apnea: a bivariate meta-analysis. Sleep Med Rev 36:57–70. https://doi.org/10.1016/j.smrv.2016.10.004

19. Mulherin SA, Miller WC (2002) Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Ann Intern Med 137:598–602. https://doi.org/10.7326/0003-4819-137-7-200210010-00011

20. Page M.J., McKenzie J.E., Bossuyt P.M., Boutron I., Hoffmann T., Mulrow C.D., Shamseer L. & Moher D. (2020) Updating guidelines for reporting systematic reviews and meta-analyses: development of the PRISMA 2020 statement. Advances in Evidence Synthesis: special issue Cochrane Database of Systematic Reviews. 9 (supplement 1) https://doi.org/10.1002/14651858.CD202001

21. Freeman, S.C., Kerby, C.R., Patel, A., Cooper, N.J., Quinn T. & Sutton, A.J. (2019) Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA. BMC Med Res Methodol.19 (81). Available from: https://doi.org/10.1186/s12874-019-0724-x

22. RevMan. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014

23. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

24. Macaskill, P., Gatsonis, C., Deeks, J.J., Harbord, R.M. & Takwoingi, Y. Chapter 10: analysing and presenting results. In: Deeks J.J., Bossuyt P.M., Gatsonis C. (2010) Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0. The Cochrane Collaboration. Available from: https://methods.cochrane.org/sdt/handbook-dta-reviews. Accessed 16 Mar 2018

25. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 58:982–90. https://doi.org/10.1016/j.jclinepi.2005.02.022

26. Chu H, Cole SR (2006) Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalised linear mixed model approach. J Clin Epidemiol 12:1331–1332. https://doi.org/10.1016/j.jclinepi.2006.06.011

27. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA (2007) A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics. 8:239–251. https://doi.org/10.1093/biostatistics/kxl004

28. Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1–48. https://doi.org/10.18637/jss.v067.i01

29. Abdullah B, Idris AI, Mohammad ZW, Mohamad H (2018) Validation of Bahasa Malaysia STOP-BANG questionnaire for identification of obstructive sleep apnea. Sleep and Breath. 22:1235–1239. https://doi.org/10.1007/s11325-018-1663-1

30. Alhouqani, S., Al Manhali, M., Al Essa, A. & Al-Houqani, M. (2015) Evaluation of the Arabic version of STOP-Bang questionnaire as a screening tool for obstructive sleep apnea. Sleep Breathing. 9(4), 1235–1240. Available from: http://link.springer.com/https://doi.org/10.1007/s11325-015-1150-x

31. Amra B, Nouranian E, Golshan M, Fietze I, Penzel T (2013) Validation of the Persian version of Berlin Sleep Questionnaire for diagnosing obstructive sleep apnea. Int J Prev Med 4(3):334–339 (PMID: 23626891)

32. Amra B, Javani M, Soltaninejad F, Penzel T, Fietze I, Schoebel C, Farajzadegan Z (2018) Comparison of Berlin questionnaire, STOP-Bang, and Epworth sleepiness scale for diagnosing obstructive sleep apnea in Persian patients. Int J Prev Med 9:1–5. https://doi.org/10.4103/ijpvm.ijpvm_131_17

33. Oktay Arslan B, UçarHoşgör ZZ, Orman MN (2020) Which screening questionnaire is best for predicting obstructive sleep apnea in the sleep clinic population considering age, gender, and comorbidities? Turkish Thoracic J. 21(6):383–389. https://doi.org/10.5152/TurkThoracJ.2019.19024

34. Avincsal, M.O., Dinc, M.E., Ulusoy, S., Dalgic, A., Ozdemir, C. & Develioglu, O.N. (2017) Modified Mallampati score improves specificity of STOP-BANG questionnaire for obstructive sleep Apnea. J Craniofacial Surg. 28(4), 904–908. https://doi.org/10.1097/SCS.0000000000003513

35. BaHammam, A.S., Al-Aqeel, A.M., Alhedyani, A.A., Al-Obaid, G.I., Al-owais, M.M. & Olaish, A.H. (2015) The validity and reliability of an Arabic version of the STOP-Bang Questionnaire for identifying obstructive sleep apnea. The Open Respiratory Medicine Journal. 9(1), 22–9. Available from: http://benthamopen.com/ABSTRACT/TORMJ-9-22

36. Boynton, G., Vahabzadeh, Hammoud, S., Ruzicka, D.L. & Chervin, R.D. (2013) Validation of the STOPBANG Questionnaire among patients referred for suspected obstructive sleep apnea. Journal of Sleep Disorders: Treatment & Care. 2(4). Available at: https://doi.org/10.4172/2325-9639.1000121

37. Deflandre E, Piette N, Bonhomme V, Degey S, Cambron L, Poirrier R, Richant J-F, Joris J (2018) Comparison of clinical scores in their ability to detect hypoxemic severe OSA patients. PLoS ONE. 13(5):1–11. https://doi.org/10.1371/journal.pone.0196270

38. Delgado-Vargas B, Acle-Cervera L, Narciso López G (2020) Validation of the Spanish version of the STOP-Bang Questionnaire: usefulness as a screening tool for obstructive sleep apnea in adults. Ear Nose Throat J 18:1–7. https://doi.org/10.1177/0145561320932334

39. Duarte, R.L.M. de M., Fonseca, L.B. de M., Magalhães-da-Silveira, F.J., Da Silveira, E. A. & Rabahi, M.F. (2017) Validation of the STOP-Bang questionnaire as a means of screening for obstructive sleep apnea in adults in Brazil [Validação do questionário STOP-Bang para a identificação de apneiaobstrutiva do sono em adultos no Brasil]. J Brasileiro Pneumol. 43(6), 456–463. Available from: https://doi.org/10.1590/s1806-37562017000000139

40. Duarte RL, Magalhães-da-Silveira FJ, Oliveira-E-Sá TS, Silva JA, Mello FC, Gozal D (2020) Obstructive sleep apnea screening with a 4-item instrument, named GOAL questionnaire: development, validation and comparative study with No-Apnea, STOP-Bang, and NoSAS. Nat Sci Sleep 23(12):57–67. https://doi.org/10.2147/NSS.S238255

41. El-Sayed IH (2012) Comparison of four sleep questionnaires for screening obstructive sleep apnea. Egypt J Chest Dis Tuberc 61(4):433–441. https://doi.org/10.1016/j.ejcdt.2012.07.003

42. Ha SCN, Lee DLY, Abdullah VJ, Van Hasselt CH (2014) Evaluation and validation of four translated Chinese questionnaires for obstructive sleep apnea patients in Hong Kong. Sleep Breathing. 18(4):715–721. https://doi.org/10.1007/s11325-013-0889-1

43. Hu YY, Yu Y, Wang ZB, Liu C, Cui YH, Xiao WM (2019) Reliability and validity of simplified Chinese STOP-BANG questionnaire in diagnosing and screening obstructive sleep apnea hypopnea syndrome. Curr Med Sci 39(1):127–133. https://doi.org/10.1007/s11596-019-2010-x

44. Kashaninasab F, Alavi K, Farhadi M, Salehi M, Bandi MFGA (2017) Comparative study of four Persian versions of sleep questionnaires for screening obstructive sleep apnea syndrome (OSAS). Med J Islamic Repub Iran. 31(1):810–816. https://doi.org/10.1007/s11596-019-2010-x

45. Khaledi-Paveh B, Khazzaie H, Nasouri M, Ghadami MR, Tahmasian M (2016) Evaluation of Berlin questionnaire validity for sleep apnea risk in sleep clinic populations. Basic Clin Neurosci 7(1):43–48 (PMID: 27303598)

46. Kim B, Lee EM, Chung Y-S, Kim W-S, Lee S-A (2015) The utility of three screening questionnaires for obstructive sleep apnea in a sleep clinic setting. Yonsei Med J 56(3):684–690. https://doi.org/10.3349/ymj.2015.56.3.684

47. Ong TH, Raudha S, Fook-Chong S, Lew N, Hsu AAL (2010) Simplifying STOP-BANG: use of a simple questionnaire to screen for OSA in an Asian population. Sleep Breathing. 14(4):371–376. https://doi.org/10.1007/s11325-010-0350-7

48. Pataka A, Zarogoulidis P, Hohenforst-Schmidt W, Tsiouda T, Tsavlis D, Kioumis DI, Papakala E, Karapantzos I, Karapantzou C, Rapati A, Tsakiridis K, Zarogoulidis K, Argyropoulou P (2016) During economic crisis can sleep questionnaires improve the value of oximetry for assessing sleep apnea? Ann Transl Med 4(22):443–443. https://doi.org/10.21037/atm.2016.11.06

49. Pecotic R, Dodig IP, Valic M, Ivkovic N, Dogas Z (2012) The evaluation of the Croatian version of the Epworth sleepiness scale and STOP questionnaire as screening tools for obstructive sleep apnea syndrome. Sleep Breathing 16(3):793–802. https://doi.org/10.1007/s11325-011-0578-x

50. Pereira EJ, Driver HS, Stewart SC, Fitzpatrick MF (2013) Comparing a combination of validated questionnaires and level III portable monitor with polysomnography to diagnose and exclude sleep apnea. J Clin Sleep Med 9(12):1259–1266. https://doi.org/10.5664/jcsm.3264

51. Perumalsamy, P., Charles, A.K., Reddy, G., Chokkalingam, C. & Balakrishnan, R. (2017) Comparison of three sleep questionnaires in screening obstructive sleep apnoea. J Evol Med Dental Sci. 6(88), 6132–6136. Available from: https://www.jemds.com/data_pdf/preethi-meena-giri-Nov-6-.pdf

52. Sadeghniiat-Haghighi K, Montazeri A, Khajeh-Mehrizi A, Ghajarzadeh M, Alemohammad ZB, Aminian O, Sedaghat M (2015) (2015) The STOP-BANG questionnaire: reliability and validity of the Persian version in sleep clinic population. Qual Life Res 24(8):2025–2030. https://doi.org/10.1007/s11136-015-0923-9

53. Saleh ABM, Ahmad MA, Awadalla NJ (2011) Development of Arabic version of Berlin questionnaire to identify obstructive sleep apnea at risk patients. Ann Thorac Med 6(4):212. https://doi.org/10.4103/1817-1737.84775

54. Sangkum, L., Klair, I., Limsuwat, C., Bent, S., Myers, L. & Thammasitboon, S. (2017) Incorporating body-type (apple vs. pear) in STOP-BANG questionnaire improves its validity to detect OSA. J Clin Anaesthesia. 41, 126–131. Available from: https://doi.org/10.1016/j.jclinane.2016.12.019 https://linkinghub.elsevier.com/retrieve/pii/S0952818016311369

55. Suksakorn S, Rattanaumpawan P, Banhiran W, Cherakul N, Chotinaiwattarakul W (2014) Reliability and validity of a Thai version of the Berlin questionnaire in patients with sleep disordered breathing. J Med Assoc Thai 97(3):S46-56 (PMID: 24772580)

56. Vana KD, Silva GE, Goldberg R (2013) Predictive abilities of the STOP-Bang and Epworth Sleepiness Scale in identifying sleep clinic patients at high risk for obstructive sleep apnea. Res Nurs Health. 36(1):84–94. https://doi.org/10.1002/nur.21512

57. Yüceege M, Firat H, Sever Ö, Demir A, Ardiç S (2015) The effect of adding gender item to Berlin Questionnaire in determining obstructive sleep apnea in sleep clinics. Ann Thorac Med 10(1):25–28. https://doi.org/10.4103/1817-1737.146856

58. Chung F, Yeneswaran B, Liao P, Chung SA, Vairavananthan S, Islam S, Khajehdehi A, Shapiro CM (2008) Validation of the Berlin questionnaire and American Society of Anesthesiologists checklist as screening tools for obstructive sleep apnea in surgical

59. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, Khajehdehi A, Shapiro CM (2008) STOP Questionnaire: a tool to screen patient for obstructive sleep apnea. Anesthesiology. 108(5):812–821. https://doi.org/10.1097/ALN.0b013e31816d83e4

60. Chung F, Chau E, Yang Y, Liao P, Hall R, Mokhlesi B (2013) Serum bicarbonate level improves specificity of STOP-Bang screening for obstructive sleep apnea. Chest 143(5):1284–1293. https://doi.org/10.1378/chest.12-1132

61. Chung F, Yang Y, Brown R, Liao P (2014) Alternative scoring models of STOP-Bang questionnaire improve specificity to detect undiagnosed obstructive sleep apnea. J Clin Sleep Med 10(9):951–958. https://doi.org/10.5664/jcsm.4022

62. Deflandre E, Degey S, Brichant J-F, Conneau A-F, Frognier R, Poirrier R, Bonhomme V (2017) Pre-operative ability of clinical scores to predict obstructive sleep apnea (OSA) severity in susceptible surgical patients. Obesity Surgery. 27(3):716–729. https://doi.org/10.1007/s11695-016-2352-4

63. Nunes FS, Danzi-Soares NJ, Genta PR, Drager LF, Cesar LAM, Lorenzi-Filho G (2014) Critical evaluation of screening questionnaires for obstructive sleep apnea in patients undergoing coronary artery bypass grafting and abdominal surgery. Sleep Breathing. 19(1):115–122. https://doi.org/10.1007/s11325-s11325-014-0971-3

64. Xia, M., Liu, S., Ji, N., Xu, J., Zhou, Z., Tong, J. & Zhang Y. (2018) BMI 35 kg/m^2 does not fit everyone: a modified STOP-Bang questionnaire for sleep apnea screening in the Chinese population. *Sleep and Breathing*. 1–8. Available from: https://doi.org/10.1007/s11325-017-1610-6

65. Giampá SQC, Pedrosa RP, Gonzaga CC, Bertolami A, Amodeo C, Furlan SF, Bortolotto LA, Lorenzi-Filho G, Drager LF (2018) Performance of NoSAS score versus Berlin questionnaire for screening obstructive sleep apnoea in patients with resistant hypertension. J Hum Hypertens 32(7):518–523. https://doi.org/10.1038/s41371-018-0072-z

66. Margallo VS, Muxfeldt ES, Guimarães GM, Salles GF (2014) Diagnostic accuracy of the Berlin questionnaire in detecting obstructive sleep apnea in patients with resistant hypertension. J Hypertens 32(10):2030–2037. https://doi.org/10.1097/HJH.0000000000000297

67. Lu H, Fu C, Li W, Jiang H, Wu Y, Li S (2017) Screening for obstructive sleep apnea syndrome in asthma patients: a prospective study based on Berlin and STOP-Bang questionnaires. J Thoracic Dis 9(7):1945–1958. https://doi.org/10.21037/jtd.2017.06.03

68. Gantner D, Ge J-Y, Antic N, Windler S, Wong K, Heeley E, Huang S-G, Ciu P, Anderson C, Wang J-G, McEvoy D (2010) Diagnostic accuracy of a questionnaire and simple home monitoring device in detecting obstructive sleep apnoea in a Chinese population at high cardiovascular risk. Respirology 15(6):952–960. https://doi.org/10.1111/j.1440-1843.2010.01797.x

69. Firat H, Yuceege M, Demir A, Ardic S (2012) Comparison of four established questionnaires to identify highway bus drivers at risk for obstructive sleep apnea in Turkey. Sleep and Biological Rhythms. 10(3):231–236. https://doi.org/10.1111/j.1479-8425.2012.00566.x

70. Elkholy SH, Amer HA, Nada MM, Nada MAF, Labib A (2012) Sleep-related breathing disorders in cerebrovascular stroke and transient ischemic attacks: a comparative study. J Clin Neurophysiol 29(2):194–198. https://doi.org/10.1097/WNP.0b013e31824d96b0

71. Bouloukaki I, Komninos ID, Mermigkis C, Micheli K, Komninou M, Moniaki V, Mauroudi E, Siafakas NM, Schiza S (2013)

Translation and validation of Berlin questionnaire in primary health care in Greece. BMC Pulmonary Medicine 13(1):6. https://doi.org/10.1186/1471-2466-13-6

72. Yunus, A., Seet, W., Mohamad Adam, B. & Haniff, J. (2014) Validation of the Malay version of Berlin questionnaire to identify Malaysian patients for obstructive sleep apnea. Malaysian Family Physic. 8(1), 5–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25606261

73. Banhiran, W., Durongphan, A., Saleesing, C.K. & Chongkolwatana, C. (2014) Diagnostic properties of the STOP-bang and its modified version in screening for obstructive sleep apnea in Thai patients. J Med Assoc Thailand. 97(6), 644–654. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25137882

74. Deeks JJ, Macaskill P, Irwig L (2005) The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 58(9):882–893. https://doi.org/10.1016/j.jclinepi.2005.01.016

75. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP (1999) Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med 131(7):485–491. https://doi.org/10.7326/0003-4819-131-7-199910050-00002

76. Nagappa M, Liao P, Wong J, Auckley D, Ramachandran SK, Memtsoudis S, Mokhlesi B, Chung F (2015) Validation of the STOP-Bang questionnaire as a screening tool for obstructive sleep apnea among different populations: a systematic review and meta-analysis. PLoS ONE 10(12):e0143697. https://doi.org/10.1371/journal.pone.0143697

77. Senaratna CV, Perret JL, Matheson MC, Lodge CJ, Lowe AJ, Cassim R, Russell MA, Burgess JA, Hamilton GS, Dharmage SC (2017) Validity of the Berlin questionnaire in detecting obstructive sleep apnea: a systematic review and meta-analysis. Sleep Med Rev 36:116–124. https://doi.org/10.1016/j.smrv.2017.04.001

78. Kaw R, Michota F, Jaffer A, Ghamande S, Auckley D, Golish J (2006) Unrecognized sleep apnea in the surgical patient: implications for the peri-operative setting. Chest 129:198-e205. https://doi.org/10.1378/chest.129.1.198

79. Liao P, Yegneswaran B, Vairavanathan S, Zilberman P, Chung F (2009) Postoperative complications in patients with obstructive sleep apnea: a retrospective matched cohort study. Canadian Journal of Anaesthesia. 56:819-e828. https://doi.org/10.1007/s12630-009-9190-y

80. Van Oosten EM, Hamilton A, Pesikas D, Payne D, Redfearn DP, Zhang S, Hopman WM, Baranchuk A (2014) Effect of preoperative obstructive sleep apnea on the frequency of atrial fibrillation after coronary artery bypass grafting. Am J Cardiol 113(6):919–923. https://doi.org/10.1016/j.majcard.2013.11.047

81. Peker, Y., Celik, Y., Arbatli, S., Isik, S.R., Balcan, B., Karataş, F., Uzel, F.I., Tabak, L., Çetin, B., Baygül, A., Öztürk, A.B., Altuğ, E., İliaz, S., Atasoy, C., Kapmaz, M., Yazici, D., Bayram, H., Çetin, B.D. & Çağlayan, B. The OSACOVID-19 Study Collaborators (2021) Effect of high-risk obstructive sleep apnea on clinical outcomes in adults with coronavirus disease 2019: a multicenter, prospective, observational cohort study. Ann Am Thorac Soc. Available from https://doi.org/10.1513/AnnalsATS.202011-1409OC

82. Bossuyt PMM, Reitsma JB, Linnet K, Moons KGM (2012) Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin Chem 58(12):1636–1643. https://doi.org/10.1373/clinchem.2012.182576

83. American Academy of Sleep Medicine Task Force (1999) 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 22:667–89

84. Iber, C., Ancoli-Israel, S., Chesson, A.L. & Quan, S.F. for the American Academy of Sleep Medicine (2007) The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. First Edition. Westchester, IL: Am Acad Sleep Med. Available from https://doi.org/10.1186/s41606-018-0023-1

85. Berry RB, Budhiraja R, Gottlieb DB, Gozal D, Iber C, Kapur VK, Marcus CL, Mehra R, Parthasarathy S, Quan SF, Redline S, Strohl KP, Davidson SL, Tangredi MM (2012) Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. J Clin Sleep Med 8(5):597–619. https://doi.org/10.5664/jcsm.2172

86. Shamim-Uzzaman A, Singh S, Chowdhuri S (2018) Hypopnea definitions, determinants and dilemmas: a focused review. Sleep Sci Practice 2(7):1–12

87. Guilleminault C, Hagen CC, Huynh NT (2009) Comparison of hypopnea definitions in lean patients with known obstructive sleep apnea hypopnea syndrome (OSAHS). Sleep Breath 13:341–7. https://doi.org/10.1007/s11325-009-0253-7

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.