Sleep Medicine Reviews 36 (2017) 116–124

Contents lists available at ScienceDirect

Sleep Medicine Reviews

journal homepage: www.elsevier.com/locate/smrv

CLINICAL REVIEW

Validity of the Berlin questionnaire in detecting obstructive sleep apnea: A systematic review and meta-analysis

Chamara V. Senaratna [a, b], Jennifer L. Perret [a, c], Melanie C. Matheson [a, d], Caroline J. Lodge [a, d], Adrian J. Lowe [a, d], Raisa Cassim [a, d], Melissa A. Russell [a, d], John A. Burgess [a], Garun S. Hamilton [e, f, g, *, 1], Shyamali C. Dharmage [a, d, 1]

[a] School of Population and Global Health, The University of Melbourne, Melbourne, Australia
[b] Department of Community Medicine, University of Sri Jayewardenepura, Nugegoda, Sri Lanka
[c] Institute for Breathing & Sleep (IBAS), Heidelberg, Melbourne, Victoria 3084, Australia
[d] Murdoch Children's Research Institute, Parkville, VIC 3052, Australia
[e] Department of Lung and Sleep Medicine, Monash Health, Clayton, Australia
[f] School of Clinical Sciences, Monash University, Clayton, Australia
[g] Monash Partners – Epworth, Melbourne, Australia

ARTICLE INFO

Article history:
Received 23 November 2016
Received in revised form
31 March 2017
Accepted 4 April 2017
Available online 8 April 2017

Keywords:
Berlin questionnaire (BQ)
Obstructive sleep apnea (OSA)
Screen
Validity
Validation
Diagnostic utility
Sensitivity
Specificity
Diagnostic odds ratio

SUMMARY

We aimed to systematically review the Berlin questionnaire as a screening tool for obstructive sleep apnea. We systematically searched PubMed, Embase, and Scopus databases, reviewed articles reporting the Berlin questionnaire's diagnostic utility as measured against type-1 polysomnography, and performed meta-analyses where possible. Thirty five eligible articles showed that the Berlin questionnaire's diagnostic utility varied by study population, definition of hypopnea used, and apnea-hypopnea index threshold used. It had good sensitivity and specificity for detecting clinically relevant obstructive sleep apnea as well as any obstructive sleep apnea in the sleep clinic population. Despite limited evidence, it showed modest to high sensitivity for detecting clinically relevant obstructive sleep apnea or any obstructive sleep apnea in other clinical and general population subgroups. Its specificity was relatively low. Possible reasons for variability in reported diagnostic utility of the Berlin questionnaire are multifaceted. We conclude that the Berlin questionnaire is useful as a clinical screening test and epidemiological tool in the sleep clinic population. Despite limited evidence, it likely has potential clinical and research utility in other populations. Adopting more consistent methodological definitions and focussing more on the general population and specific clinical populations to determine its usefulness as a clinical or epidemiological screening tool are recommended.

Introduction

Obstructive sleep apnea (OSA) is the most common phenotype of the serious sleep-related breathing disorders [1]. Despite being frequently underdiagnosed [2–4], its prevalence in the population is high [5]. The prevalence of moderate to severe OSA is reported to be 6%–17%, being as high as 49% with advanced age [5]. In various populations and their subgroups the prevalence varies markedly by age, sex, and body-mass index (BMI) of the respective population, and the methods and definitions used in the diagnosis of OSA [5–8].

OSA is diagnosed using various methods and diagnostic definitions. The gold standard is laboratory-based, overnight, attended polysomnography, i.e., the type 1 sleep study [9]. Sleep studies

Abbreviations: AHI, Apnea-hypopnea index; BMI, Body-mass index; CI, Confidence interval; OSA, Obstructive sleep apnea; QUADAS-2, Quality assessment of diagnostic accuracy studies-2.

* Corresponding author. School of Clinical Sciences, Monash University, Monash Medical Centre, 246 Clayton Road, Clayton VIC 3168, Australia. Fax: +61 3 9594 6415.

E-mail addresses: chamaravs@sjp.ac.lk, senaratnab@unimelb.edu.au (C.V. Senaratna), jperret@unimelb.edu.au (J.L. Perret), mcmat@unimelb.edu.au (M.C. Matheson), clodge@unimelb.edu.au (C.J. Lodge), lowea@unimelb.edu.au (A.J. Lowe), raisa.cassim@unimelb.edu.au (R. Cassim), melissar@unimelb.edu.au (M.A. Russell), jburgess@unimelb.edu.au (J.A. Burgess), garun.hamilton@monashhealth.org (G.S. Hamilton), s.dharmage@unimelb.edu.au (S.C. Dharmage).

[1] Equally contributed.

http://dx.doi.org/10.1016/j.smrv.2017.04.001
1087-0792/

using portable instruments are also widely used. These include: type 2 sleep studies that use the same parameters as type 1 but are performed at the patient's home without an attending technician and; type 3 and 4 sleep studies that record limited breathing-related parameters [10]. Sleep studies using portable instruments have variable sensitivity and specificity against the gold standard for diagnosing OSA [11,12] and thus are not comparable with gold standard measures for testing the sensitivity and specificity of screening questionnaires. Although widely used in the diagnosis of or screening for OSA, the utility of portable measures is the subject of an ongoing discussion [9,11,13–15]. Sleep related breathing events that indicate OSA are recorded in sleep studies as apnea (breathing pauses lasting for at least ten seconds) and hypopnea (reduction in respiratory airflow (without apnea) associated with oxygen desaturation or arousal from sleep). A composite index of these two measures, the apnea-hypopnea index (AHI), is used to determine the presence or absence of OSA. Methods for measuring airflow (relevant for accurately identifying hypopneas) are also variable and have changed over time. It has been measured using oro-nasal thermistors, inductive plethysmography, or nasal cannula/pressure transducer systems (where the latter is the preferred and more sensitive method) [9]. The diagnostic AHI thresholds used are arbitrary rather than evidence-based. Commonly used AHI thresholds are ≥5 events/h, ≥15 events/h, and ≥30 events/h, which (by consensus) indicate mild (AHI ≥5–15 events/h), moderate (AHI ≥15–30 events/h), and severe (AHI ≥30 events/h) disease [16,17]. AHI ≥5 events/h is often reported in epidemiological studies [5], although AHI of ≥15 events/h and ≥30 events/h are usually considered clinically relevant (requiring treatment). However, AHI ≥5 events/h is also considered clinically significant if accompanied by symptoms such as excessive daytime sleepiness, non-restorative sleep or substantial fatigue [17]. How hypopnea is defined has changed serially over time. For example, successive versions of the American Academy of Sleep Medicine (AASM) criteria have defined hypopnea as a) >50% drop in airflow or a lower airflow reduction with oxygen desaturation of ≥3% or an arousal [17]; b) ≥30% reduction in nasal pressure signal plus ≥4% oxygen desaturation, or c) ≥50% reduction in nasal pressure signal plus ≥3% oxygen desaturation or an arousal [18]; and d) ≥30% reduction in airflow plus ≥3% oxygen desaturation or an arousal [19]. Classification of a patient as having OSA or not is significantly influenced by the definition used for hypopnea [6,8,20].

Due to cost, human resources, and other logistics required, it is difficult to offer gold standard polysomnography to everyone. In addition, the use of polysomnography in large epidemiological research studies is constrained by its cost and the logistics of investigating a large number of participants around the same time. This leads to difficulty in accurately assessing the prevalence of OSA due to few studies using gold standard polysomnography [5]. A suitable screening method is potentially advantageous to detect those at high risk of OSA for subsequent assessment with polysomnography. Although some portable devices can also be used for this purpose, cost and human resources required remain constraints for large studies.

Several questionnaires have been developed to detect those at high risk of OSA [21]. These include the Berlin questionnaire, which is a commonly used questionnaire in epidemiological and clinical research with the highest number of validation studies [22]. Although its validity has been examined in a variety of populations, the reported sensitivity and specificity varies from study to study. The variability in diagnostic utility is probably in part due to differences in the type of sleep study and the definitions used in the reference standard. It is in this context that a scientific synthesis of the current evidence related to the validity of this screening instrument is timely. A recent systematic review article [22] compared the Berlin questionnaire with STOP, STOP-BANG, and Epworth sleepiness scale questionnaires. However, this review meta-analysed the estimates without regard to the heterogeneity in the reference standard used, for example the variability in the type of sleep study, variation in AHI cut-offs, or the hypopnea definition. Furthermore, when an article included in the above review reported the diagnostic utility of the Berlin questionnaire for more than one AHI threshold, only the one with highest Youden's index was included in the review, thereby not revealing the within-study variation of the diagnostic utility of the Berlin questionnaire at different AHI thresholds. To overcome these limitations in the current literature, our aim was to systematically review, document and attempt to explain the variability in the diagnostic utility of the Berlin questionnaire as measured against the gold standard - type 1 polysomnography and different hypopnea definitions, and reported for different AHI thresholds. An additional aim was to meta-analyse its validity parameters only where possible.

Methodology

Berlin questionnaire

The Berlin questionnaire consists of ten questions plus information on height and weight arranged in three categories: snoring and cessation of breathing (category 1; five questions); symptoms of excessive daytime sleepiness (Category 2; four questions); and BMI and hypertension (category 3; one question and height and weight information). Positive scores in 2 or more categories suggest that the respondent has a high risk for OSA [23].

Search strategy

The final search was conducted on 20th November 2016, and we searched three databases, namely, PubMed, Embase, and Scopus, using search criteria shown in Table S1. Broadly, the search terms addressed four aspects: OSA as a disease of interest, the Berlin questionnaire as the index test, Type 1 sleep studies as the reference test, and measures of diagnostic utility (sensitivity, specificity, predictive values, and likelihood ratios) as the validity parameters. Only studies conducted among adults and published in English were included.

Systematic search

Two authors (CS and RC) independently conducted the systematic search, screening of articles and selecting those eligible for review, quality assessment of the selected articles, and data extraction. When there were disputes, these were referred to a third author (SD) for resolution.

Screening of articles

We removed the duplicates and screened the articles using titles and abstracts to determine their eligibility for full article screening. Further evaluation of full texts determined eligibility for inclusion in the review and meta-analysis.

Eligibility criteria

We included all studies that investigated the validity of the Berlin questionnaire and reported at least one of the chosen main validity parameters i.e., sensitivity and/or specificity, or information from which these could be derived. In addition, we included those articles that reported the validity of the Berlin questionnaire as a secondary outcome. Studies that had not reported the

diagnostic utility of the Berlin questionnaire (or data that allowed its derivation) were excluded.

Quality assessment of the selected articles

We used the quality assessment of diagnostic accuracy studies-2 (QUADAS-2) [24] criteria to assess the quality of the selected articles. This tool has been widely used to assess the quality of studies reporting the validity of screening questionnaires. It has fourteen items that assess four domains (patient selection, index test, reference standard, and flow and timing of the tests) that appear under two main categories (risk of bias and applicability judgement). Each item is a question that is answered 'yes', 'no' or 'unclear', where 'yes' indicates a low risk of bias or low concerns of applicability. For example, the items included in the 'patient selection' domain of the 'risk of bias' category are: 'was a consecutive or random sample of patients enrolled?'; 'was a case −control design avoided?'; and 'did the study avoid inappropriate exclusions?'. The responses to these items do not generate a summary score. Instead, if items in all domains have been answered 'yes' this categorises the respective study as having low risk of bias or concerns of applicability, and if one or more questions (domains) were answered 'no' or 'unclear', the respective study is judged to be 'at risk of bias' or as having 'concerns regarding applicability'.

Data extraction

Data from the eligible articles were extracted to predefined tables. The extracted data included the name/s of author/s, year of publication, study setting and country, nature of data collection (prospective or retrospective), source population, size of the sample used for the validation study, sampling method/s, age of the sample, percentage of males in the sample, type of sleep study used as a reference standard, the definitions used for apnea and hypopnea, the number of participants in different AHI categories (AHI <5 events/h, ≥5 events/h, ≥10 events/h, ≥15 events/h, and ≥30 events/h), and the sensitivity/specificity/predictive values/likelihood ratios reported for different AHI threshold levels. Some validity parameters were calculated for the studies which did not report these, but provided sufficient information to allow these parameters to be calculated. The reported sensitivity, specificity, and the prediction values were rounded off to zero decimal points and the likelihood ratios and the diagnostic odds ratios to one decimal point (diagnostic odds ratio is used to discriminate subjects with a given disorder from subjects without it [25]). It is useful as a compound measure of diagnostic utility (taking into account both the sensitivity and specificity) but has the disadvantage of being unable to determine the true positive and false positive rate separately. Its value ranges from 0 to infinity, with higher values indicating better discriminatory ability of the test. Several studies had reported an odds ratio as a validity parameter, which is different from the diagnostic odds ratio. To be consistent across studies, we re-calculated the diagnostic odds ratios for all studies afresh using the following formula; diagnostic odds ratio = (sensitivity/(1 − sensitivity))/(1 − specificity/specificity) [25]. The validity parameters of the Berlin questionnaire were categorized based on the study population, type of sleep study, the definition of hypopnea used (3% oxygen desaturation vs 4% oxygen desaturation), and the different levels of AHI.

Meta-analysis

We grouped studies based on the type of study population (e.g., sleep clinic patients, surgical patients etc.), and the AHI threshold for which the validity parameters had been reported (AHI ≥5 events/h, AHI ≥15 events/h etc.). We used Stata Statistical Software: Release 13.1 (StataCorp, College Station (TX) to obtain pooled, summary measures for each subgroup. Due to specific data requirements, meta-analysis could be performed only for groups containing at least four studies. The summary estimates for sensitivity, specificity, likelihood ratios, and diagnostic odds ratios were presented together with forest plots to graphically express the distribution of these parameters.

Results

Article selection process

Our search in PubMed, Embase, and Scopus yielded 663 articles. After screening the titles and abstracts, 53 were eligible for full paper screening, of which 35 were eligible for the review. The article selection process is shown in Fig. 1. A summary of the articles that were excluded during the full paper review and the reasons for their exclusion are given in Table S2.

Quality assessment of the selected studies

The summary of the quality assessment of the studies selected for the review is shown in Table S3. Twenty seven of the thirty-five studies had minimal risk of bias, and none of the studies were deemed inapplicable to the review question.

Characteristics of the eligible studies

The details of the studies selected for the review are shown in Tables S4–S7. The summary of these characteristics is given in Table 1.

Most of the validation studies had been conducted in North America (n = 11). There were no studies in Central Asia or Africa (other than Egypt). Over half of the studies were in sleep clinic patients or patients suspected to have OSA, and there were only 2 studies in the general population. The definitions used for diagnosis

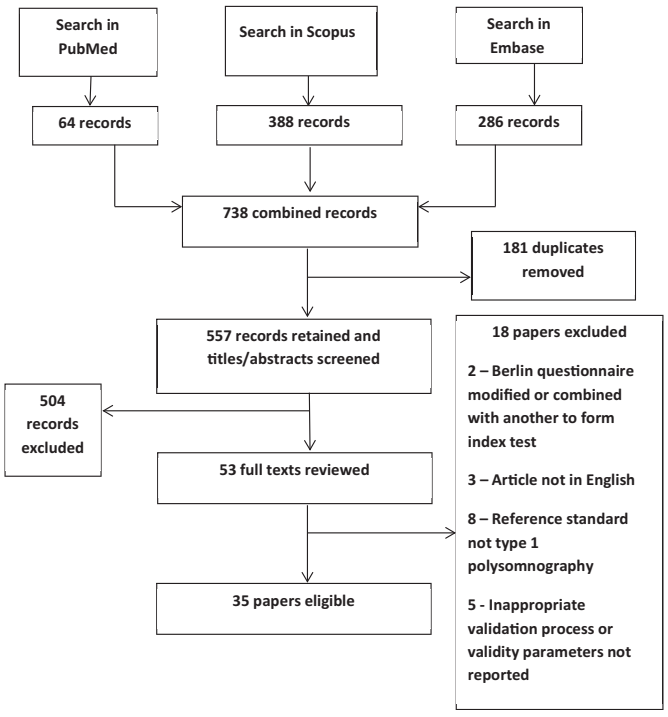

Fig. 1. Screening process of the papers included in the analysis.

Table 1
Summary characteristics of the eligible articles.

Characteristic	Number of studies (N = 35)
Region in the world	
North America [29,32,34,37,39,42,46–48,50,68]	11
West Asia or North Africa [27,36,40,41,44,49,69,70]	8
East or South east Asia [31,43,51,52,71–73]	7
South America [26,33,45,74]	4
Europe [28,30,38,75]	3
Australia or New Zealand [35]	1
South Asia [76]	1
Type of study population[a]	
Sleep clinic patients [28,29,34,36,40–44,48–51,69,71–73,76]	18
Patients with cardio- or cerebro-vascular disease or risk factors for such disease [26,32,33,39,45,52,70,74]	8
General population [31,38]	2
Occupational groups (drivers or nurses) [27,68]	2
Surgical patients [33,46]	2
Pregnant women [35]	1
Other groups (primary care patients [30], or patients with arthritis [37], or intracranial hypertension [47])	3
Exclusion of previously-diagnosed OSA patients	
Excluded [27,29,30,32,35,39,48,51,68,69,74]	11
Not excluded or not mentioned [26,28,31,33,34, 36–38,40–47,49,50,52,70–73,76]	24
Definition of hypopnea used in the diagnosis of OSA	
≥50% airflow reduction with a ≥3% oxygen desaturation [26–28,31–36,50,69]	11
≥30% airflow reduction with a ≥4% oxygen desaturation [29,38,39,42–45,49]	8
≥50% airflow reduction with a ≥4% oxygen desaturation [48,52,68]	3
Not reported or unclear [30,37,40,41,46,47,51,70–74,76]	13

[a] One study included surgical patients and patients with cardiovascular diseases as separate groups; OSA = obstructive sleep apnea.

of a hypopnea differed substantially between the studies. Nearly one third used ≥50% airflow reduction with a ≥3% oxygen desaturation to define hypopnea, while eight used ≥30% airflow reduction with a ≥4% oxygen desaturation as the definition. In addition, eleven [26–36] out of fourteen studies [26–39] that scored respiratory-related arousals used this in their definitions of hypopnea. Thirteen studies had not clearly reported the criteria used.

Reported validity parameters

The selected studies had reported the sensitivity, specificity, predictive values, likelihood ratios, and diagnostic odds ratios of the Berlin questionnaire for one or more conventionally accepted AHI thresholds (AHI ≥5 events/h, AHI ≥10 events/h, AHI ≥15 events/h, or AHI ≥30 events/h). These are shown in Tables S4–S7 (diagnostic utility in some populations are summarized in Tables 2 and 4). These reported validity parameters varied based on the type of study population, the definition of hypopnea used, and the AHI threshold for which they were reported.

Sleep clinic patients

Most studies were small in sample size with less than 150 participants (range 100 [40,41] to 1853 [42]) except for four [29,42–44]. The average age ranged from 42 to 52 y (see Tables S4–S7). Reported sensitivity and specificity of the Berlin questionnaire varied across different AHI thresholds and different definitions of hypopnea, as shown in Tables 2 and S4–S7.

The sensitivity was higher when hypopnea was defined as ≥3% oxygen desaturation rather than ≥4% (see Table 2). No such relationship with hypopnea definition was seen for specificity or diagnostic odds ratio. Diagnostic utility of the Berlin questionnaire also did not show any trend across the standard AHI thresholds.

Meta-analysis of the validity parameters in sleep clinic patients

As the commands used by the Stata for meta-analyses of diagnostic accuracy studies require at least four studies to execute, meta-analyses could only be performed for the sleep clinic patients or patients with suspected OSA. At all three conventional AHI thresholds, and when different definitions were used for hypopnea, the pooled sensitivity values remained close to each other ranging from 79% to 82% (see Table 3). Respective pooled specificity values ranged from 32% to 39%, except at threshold of AHI ≥5 events/h where the value was higher (53%) when hypopnea was defined at ≥3% oxygen desaturation. The diagnostic odds ratio was highest for threshold of AHI ≥5 events/h, when hypopnea was defined as ≥3% oxygen desaturation (4.2, 95% CI 2.1–8.4) and lowest for AHI ≥5 events/h when hypopnea was defined as ≥4% oxygen desaturation (1.7, 95% CI 1.1–2.7). Figs. S1–S4 give a graphical representation of the distribution of sensitivity, specificity, and diagnostic odds ratio for each component study and the pooled estimates. However, when hypopnea was defined as >4% oxygen desaturation, the I^2 value was very high for the pooled summary sensitivity and specificity at both AHI thresholds and for the negative likelihood ratio at AHI ≥5 events/h, indicating a high heterogeneity in these data. All diagnostic odds ratios at all AHI thresholds also have extremely high heterogeneity.

Patients with cardio- or cerebro-vascular disease or risk factors

All but one [45] of these studies had sample sizes less than 100 (see Tables S4–S7). The participants were older than the sleep clinic patients, their average ages ranging from 47 to 68 y. Similar to

Table 2
Summary of the validity parameters of Berlin questionnaire for the sleep clinic patients.

Study reference	AHI threshold (events/h)	Sensitivity %		Specificity %		DOR	
		Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Hypopnea defined at OD ≥3%	≥5	68 [50]	86 [34]	25 [34]	73 [28]	2.0 [50]	8.4 [69]
	≥10	62 [50]	88 [34]	25 [34]	63 [28]	1.2 [50]	4.4 [28]
	≥15	57 [50]	91 [34]	28 [34]	61 [28]	0.9 [50]	4.9 [28]
	≥30	71 [28]	89 [34]	18 [34]	53 [28]	1.8 [34]	2.8 [28]
Hypopnea defined at OD ≥4%	≥5	62 [48]	96 [41]	14 [29]	90 [41]	0.5 [48]	306.0 [41]
	≥10	–	–	–	–	–	–
	≥15	68 [51]	89 [49]	23 [49]	52 [51]	1.4 [71]	3.3 [42]
	≥30	69 [48]	90 [42]	19 [49]	54 [51]	1.4 [49]	3.5 [51]

AHI = apnea-hypopnea index; DOR = diagnostic odds ratio; OD = oxygen desaturation.

Table 3
Summary scores for the validity parameters of Berlin questionnaire administered to sleep clinic patients or patients with suspected OSA.

AHI threshold; N; n	Hypopnea definition	Summary scores									
		Sensitivity		Specificity		LR+		LR-		DOR	
		%	I^2	%	I^2		I^2		I^2		I^2
≥5 events/h; 4; 538	OD ≥ 3%	79 (71–86)	74.4 (48.2–100.0)	53 (34–70)	79.4 (59.3–99.6)	1.7 (1.2–2.4)	71.8 (71.8–98.6)	0.4 (0.3–0.6)	71.9 (42.5–100.0)	4.2 (2.1–8.4)	100.0 (99.9–100.0)
	OD ≥ 4%	78 (69–86)	95.9 (94.1–97.8)	32 (22–45)	92.4 (88.2–96.6)	1.2 (1.0–1.3)	81.9 (81.9–95.6)	0.7 (0.5–0.9)	93.3 (89.8–96.8)	1.7 (1.1–27)	100.0 (100.0–100.0)
≥15 events/h; 5; 971	OD ≥ 3%	82 (72–89)	82.2 (67.3–94.0)	39 (29–50)	79.6 (62.0–97.2)	1.4 (1.2–1.6)	47.7 (47.7–97.8)	0.5 (0.3–0.7)	81.2 (65.3–97.1)	2.9 (1.8–4.9)	100.0 (100.0–100.0)
	OD ≥ 4%	82 (71–89)	94.8 (91.3–98.3)	35 (27–43)	85.1 (71.6–98.6)	1.2 (1.2–1.3)	63.9 (63.6–99.2)	0.5 (0.4–0.7)	88.9 (79.5–98.2)	2.4 (1.7–3.5)	100.0 (100.0–100.0)

AHI = apnea-hypopnea index; DOR = diagnostic odds ratio; LR = likelihood ratio; N = number of studies; n = number of total participants; OD = oxygen desaturation.

Table 4
Summary of the validity parameters of Berlin questionnaire for the patients with cardio- or cerebro-vascular disease or risk factors.

Study reference	AHI threshold (events/h)	Sensitivity %		Specificity %		DOR	
		Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Hypopnea defined at OD ≥3%	≥5	72 [26]	93 [74]	39 [32]	59 [74]	2.1 [26]	18.4 [74]
	≥10 [32]	85		37		3.3	
	≥15	67 [33]	74 [26]	26 [33]	34 [26]	0.7 [33]	1.2 [26]
	≥30	–	–	–	–	–	–
Hypopnea defined at OD ≥4%	≥5	68 [39,45]	81 [52]	44 [52]	46 [39,45]	1.9 [39,45]	3.4 [52]
	≥10	–	–	–	–	–	–
	≥15	40 [45]	89 [52]	35 [52]	76 [45]	1.0 [39]	4.6 [52]
	≥30	71 [39]	91 [52]	28 [52]	40 [45]	1.4 [39]	3.9 [52]

AHI = apnea-hypopnea index; DOR = diagnostic odds ratio; OD = oxygen desaturation.

the sleep clinic patients, the sensitivity was consistently higher when hypopnea was defined as ≥3% oxygen desaturation than ≥4% oxygen desaturation, but this was not true for specificity or diagnostic odds ratio. No discernible trends of the diagnostic utility across the different AHI thresholds were seen. As shown in Tables 4 and S4–S7, the reported sensitivity and specificity varied across different AHI thresholds and different definitions of hypopnea.

General population

Tables S4–S6 show the reported diagnostic utility of the Berlin questionnaire for the general population across different AHI thresholds and definitions of hypopnea. At a threshold of AHI ≥5 events/h the sensitivity, specificity, and the diagnostic odds ratio of the Berlin questionnaire for the general population were 69%, 83%, and 11.3 when hypopnea was defined as ≥3% oxygen desaturation [31] and 37%, 84%, and 1.8 when hypopnea was defined as ≥4% oxygen desaturation [38]. At a threshold of AHI ≥10 events/h these were 79%, 67%, and 7.5 when hypopnea was defined as ≥3% oxygen desaturation [31]. At AHI ≥15 events/h and when hypopnea was defined as ≥3% oxygen desaturation, the sensitivity, specificity, and the diagnostic odds ratio were 89%, 63% and 12.9 [31]. When hypopnea was defined at ≥4% oxygen desaturation at the same AHI threshold, these were 43%, 80%, and 3.0 [38]. The diagnostic utility of the Berlin questionnaire in the general population was not reported for the AHI threshold of ≥30 events/h.

Surgical population

When hypopnea was defined as ≥3% oxygen desaturation, the sensitivity, specificity, and the diagnostic odds ratio were 69%, 56%, and 2.8 [46] at a threshold of AHI ≥5 events/h and 79% [46] to 82% [33], 50% [46] to 62% [33], and 3.8 [46] to 7.8 [33] at AHI ≥15 events/h. At a threshold of AHI ≥30 events/h, these were 87%, 46%, and 5.7 [46]. Diagnostic utility of the Berlin questionnaire for the surgical populations was not available when hypopnea was defined as ≥4% oxygen desaturation.

Other populations

The sensitivity, specificity, predictive values, and the diagnostic odds ratio of the Berlin questionnaire for other populations are shown in Tables S4–S7. These studies had younger participants than the sleep clinic patients, except for the rheumatoid arthritis patients who were aged 60 y on the average [37]. The youngest population was those with idiopathic intracranial hypertension whose median age was 23 y [47] and pregnant women whose mean age was 33 y [35]. Similar to other population subgroups, among pregnant women, both sensitivity and specificity in the 2nd as well as 3rd trimesters were less at the threshold of AHI ≥5 events/h (93% and 50% for 2nd trimester and 87% and 32% for 3rd trimester) compared to AHI ≥10 events/h (86% and 39% for 2nd trimester and 86% and 28% for 3rd trimester) [35]. Among the primary care patients, the sensitivity and diagnostic odds ratio increased consistently from the threshold of AHI ≥5 events/h (76% and 2.7) to AHI ≥30 events/h (93% and 10.2) although the specificity did not follow this trend [30].

An interesting observation across all study populations (except for [28,34,39,45,48,49]) was that, within the same study, the sensitivity increased gradually as the AHI threshold increased. However, neither the specificity nor the diagnostic odds ratio followed this trend. In two studies, this sensitivity trend was reversed [35,50].

Discussion

This review has important clinical and research implications. Our meta-analysis showed the Berlin questionnaire to have good sensitivity for detecting clinically relevant OSA (AHI ≥15 events/h) in the sleep clinic population. Although the evidence from other

populations is limited and precluded meta-analyses, we found the Berlin questionnaire to have modest-high sensitivity for detecting clinically relevant OSA in patients with: cardio- or cerebro-vascular disease or risk factors, surgical patients, and the general population when hypopnea was defined as 3% oxygen desaturation. In contrast, its specificity in all the populations was relatively low. The main limitation for interpretation of the findings was that the gold standard definition of OSA varied between studies. While all the studies used attended, laboratory polysomnography as the reference standard, the AHI cut off used to define OSA as well as oxygen desaturation threshold used to define hypopnea varied across studies.

The sensitivity of the Berlin questionnaire for detecting clinically relevant OSA in the sleep clinic population ranged from 57% [50] to 91% [34] when hypopnea was defined as 3% oxygen desaturation and from 68% [51] to 90% [42] when hypopnea was defined as 4% oxygen desaturation. Despite this variability (which was highlighted by the high I^2 values in the meta-analysis) and the low specificity reported in individual studies, the summary scores showed the Berlin questionnaire to have a good sensitivity for detecting clinically relevant OSA in this population regardless of the hypopnea definition used in the gold standard. The pooled sensitivity in our review, however, is only slightly different from the one reported by Chiu et al. [22] (80%) and reflects the variety of methodologically heterogeneous studies in the literature.

The limited studies performed in those with cardio- or cerebro-vascular disease or risk factors showed a smaller variation in sensitivity from 67% [33] to 74% [26] for detecting clinically relevant OSA when hypopnea was defined as 3% oxygen desaturation than when it was defined as 4% (range from 40% [45] to 91% [52]). In two studies from the surgical population [33,46], the Berlin questionnaire had a sensitivity of 79%–82% to detect clinically relevant OSA when hypopnea was defined as 3% oxygen desaturation. However, this evidence is too scarce to draw meaningful conclusions.

Out of 36 studies eligible for this review, only two were conducted in the general population. The difficulty in conducting a validation study using type 1 polysomnography as the gold standard test in the general population is understandable given the cost and logistics involved. Such studies will become increasingly limited given the advancement of portable polysomnography and related technology, as indicated by recent derivation and validation studies for new screening tools such as NoSAS questionnaire [53–56]. However, given that even portable polysomnography has variable sensitivity and specificity against type 1 sleep studies [11,12,55], more studies that explore the Berlin questionnaire's validity in the general population measured against the gold standard reference test are required to determine its suitability to screen the general population for possible OSA. Notwithstanding, the limited data available show that the sensitivity of the Berlin questionnaire for clinically relevant OSA in the general population is reasonably high (89% for AHI $\geq$15 events/h) when hypopnea is defined as 3% oxygen desaturation [31], but undesirably low when hypopnea is defined as 4% oxygen desaturation (43% for AHI $\geq$15 events/h) [38]. This is noteworthy in the backdrop of the recommendations of the American Academy of Sleep Medicine which currently recommends using a 3% oxygen desaturation criterion in its 2012 edition of respiratory event scoring rules [19], and this extent of oxygen desaturation with hypopnea is in more common clinical use than the more stringent 4% criterion. However, the single study that assessed the performance of the Berlin questionnaire in primary care patients [30] used a 4% oxygen desaturation threshold, but showed a similar sensitivity (86% for AHI $\geq$15 events/h) to the general population study [31] that used the 3% threshold. If further studies show similar results and confirm good sensitivity in general and primary care populations, then the Berlin questionnaire may perform well as a population screening test, identifying those who need consideration for polysomnography.

Compared with the low sensitivity of the Berlin questionnaire when hypopnea was defined as 4% oxygen desaturation, the STOP-BANG questionnaire has a higher sensitivity (87%) for clinically relevant OSA [56] (when matched for sleep study definition). However, a recent systematic review of the STOP-BANG questionnaire showed that this tool also has been validated in the general population only once [57]. Furthermore, the diagnostic utility of the STOP-BANG in the general population at AHI $\geq$5 events/h is not known and it is also unknown how well it performs when the more common 3% oxygen desaturation for hypopnea definition is used. Research directly comparing these two questionnaires in the general population is needed to determine how best to use one or both in screening for OSA. In particular, the cost effectiveness of this approach needs to be fully appraised.

Sensitivity and specificity of a screening tool are usually inversely related, and the high sensitivity often comes at the cost to specificity. When the cost of the gold standard diagnostic test is very high or the screening test is associated with a known risk to the patient, minimizing false positives during the screening (which happens when the specificity of the screening tool is high) who would unnecessarily undergo the diagnostic test is a high priority. For a disease with known serious consequences such as OSA [58], this is of a secondary importance as the costs of diagnostic tests are gradually coming down [59] and there are no risks associated with the gold standard. Therefore, in terms of sleep apnea, it is more important that a screening test has a high sensitivity, and does not miss patients with sleep apnea, rather than a high specificity. Despite the lesser importance of specificity measures in OSA screening tests, we found the Berlin questionnaire had modest or low specificity for clinically relevant OSA in general, across all populations. Although the cost of sleep studies for the resulting false positives from the Berlin questionnaire would be high, this needs to be compared with the cost of missing the true cases (due to reduced sensitivity which usually occurs when the specificity is increased) in the general population or other clinical populations as undiagnosed OSA potentially incurs heavy healthcare costs [60] and other indirect costs due to loss of productivity and poor quality of life etc. [61–63]. Evidence is clearly needed to formally assess the cost effectiveness of using the Berlin questionnaire when used to determine the need for sleep testing.

A threshold of AHI $\geq$5 events/h to define OSA of any severity is often used in epidemiological research. Although this threshold may be useful in research, it is considered to be of uncertain clinical significance given the extraordinary high population prevalence based on this threshold, and the lack of association between mild OSA and either symptoms or cardiovascular co-morbidities [5,6]. At this AHI threshold, the Berlin questionnaire showed moderate sensitivity among sleep clinic patients regardless of the hypopnea definition used, but specificity was modest-low when hypopnea was defined as 4% oxygen desaturation. Among the patients with cardio- or cerebro-vascular disease or risk factors, it showed a moderate or high sensitivity and relatively low specificity regardless of the hypopnea definition. Although the Berlin questionnaire showed a moderate sensitivity and specificity in surgical patients [46] and in the general population [31] when a hypopnea definition of 3% oxygen desaturation was used, the sensitivity in the general population reduced to 37% [38] when a hypopnea definition of 4% oxygen desaturation was used. Despite being constrained by the scarcity of evidence, these findings suggest that the Berlin questionnaire is potentially useful in epidemiological research.

The observed variability in the Berlin questionnaire performance estimates in the included studies arises from being

measured against a differing gold standard; AHI ≥5 to ≥30 events/h used to define OSA and the level of oxygen desaturation (and airflow reduction) used to define a hypopnea (see Table 1). Although use of differing screening test thresholds are often reported in validation studies, a differing gold standard is rather unusual and makes interpretation of and comparison between results in different studies difficult. Part of the variability may also be attributable to differences in population characteristics related to severity and differential reporting (e.g., age and sex). Spectrum effects may be observed if the chance of screening positive depends on the severity of the disease. If the Berlin questionnaire captured the entire spectrum of OSA (mild to severe), then the sensitivity would not change when populations with the full spectrum of disease (general population) are compared with populations with predominantly severe disease (high-risk populations). Whereas, if the Berlin questionnaire preferentially identified severe OSA, its sensitivity would decrease in populations where mild and moderate disease predominated [64]. Lack of a clear relationship between OSA severity (as measured by AHI thresholds) and the "high risk of OSA" as determined by the Berlin questionnaire in some populations in this review also likely reflects a poor relationship between OSA severity and self-reported symptoms. This is supported by recent evidence which indicates that a large proportion of those with OSA have minimal self-reported symptoms or no symptoms [65] and that there is no association between most of the self-reported sleep-related symptoms and the AHI [66]. It is likely that many people with OSA might not be aware of their nocturnal symptoms, especially when they do not have a bed-partner. In the general population, however, the sensitivity was higher at higher AHI thresholds [31,38]. The small number of studies in the general population and their lack of validation information at an AHI threshold of ≥30 events/h, however, prevent making definitive conclusions based on these observations.

Due to adverse health consequences of untreated OSA and the healthcare costs of these consequences, there is increasing interest in screening the general population, primary care patients, and other clinical populations for OSA [58,60,67,69]. However, due to uncertainty about the accuracy, clinical utility and cost effectiveness of all potential screening tools, including the Berlin questionnaire, the US Preventive Services Task Force recently stopped short of recommending screening of asymptomatic adults for OSA [14]. Our review also demonstrated the inadequacy of validation studies of the Berlin questionnaire in the general population and most clinical populations (excluding the sleep clinic population) that prevents making definitive conclusions about the usefulness of the Berlin questionnaire in them. Despite the higher number of validation studies performed in sleep clinic patients, the actual usefulness of the Berlin questionnaire in this population is questionable. Patients are referred to sleep clinics due to the presence of significant clinical features which puts them "at risk" for OSA or other sleep disorders; these clinical features are largely included as questions in the Berlin questionnaire, so it is unclear what is to be additionally gained by using the Berlin questionnaire in this population. This is true for any screening questionnaire when used in clinical settings, unless it has perfect specificity. Therefore, the Berlin questionnaire could be more useful in populations (general or clinical) whose OSA-related clinical features are yet to be elicited, as a means of determining who should be referred to a sleep clinic and/or undergo polysomnography. Given the insufficient evidence to support widespread screening of the general population for OSA [14], the Berlin questionnaire could be used opportunistically in the primary care setting to detect patients with symptoms who may need referral [58,60]. Nevertheless, as we have discussed, more research is needed to assess the utility of this approach.

In summary, using the gold standard for diagnosing OSA and stratifying study groups by OSA definition and study population, we found that the Berlin questionnaire showed some evidence of usefulness as a clinical screening test and an epidemiological tool in the sleep clinic population, those with cardio- or cerebrovascular disease or risk factors, surgical population, and primary care or general population. However, the limited number of studies available for this review precluded meta-analyses, except for the sleep clinic population which suggested that the Berlin questionnaire may be a valid option to facilitate screening of sleep patients. The variability of the available limited evidence on its performance in other populations did not allow definitive conclusions about its clinical or research utility in these specific groups. Focusing future validation studies of the Berlin questionnaire on the general population and these specific clinical populations will facilitate the assessment of its usefulness as a clinical screening tool in those settings.

Practice points

1) The Berlin questionnaire has modest-high sensitivity and low specificity to detect clinically relevant OSA in sleep clinic patients.
2) Limited evidence suggests that the Berlin questionnaire is potentially useful in screening those with cardio- or cerebro-vascular diseases or risk factors, surgical population, and primary care or the general population, but more validation studies are needed to draw firm conclusions.
3) The diagnostic utility of the Berlin questionnaire varies according to the definitions used in gold-standard polysomnography (e.g., definition of hypopnea, AHI threshold).

Research agenda

1. This review highlights the need for consensus on consistent definitions for gold-standard polysomnography to measure and diagnose obstructive sleep apnea. In addition, reporting the diagnostic utility for multiple reference standards (e.g., different diagnostic criteria recommended by the American Academy of Sleep Medicine in their successive revisions) and for multiple apnea-hypopnea thresholds could help make valid comparisons between different validation studies.
2. More validation studies are needed to determine the diagnostic utility and cost effectiveness of using the Berlin questionnaire in primary care and in the general population – the situations where obstructive sleep apnea screening tools are most needed.
3. Validation studies which directly compare the Berlin questionnaire to other OSA screening tools (such as the STOP-BANG questionnaire) in the same population are needed.

Conflicts of interest

Prof Hamilton has received equipment for research from Resmed, Philips Respironics, and Air Liquide Healthcare.

Acknowledgements

No financial or other support was used for conducting this systematic review. Dr. Chamara Senaratna is supported by the Australian Government's Research Training Program.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.smrv.2017.04.001.

References

[1] Thurnheer R. Diagnostic approach to sleep-disordered breathing. Expert Rev Respir Med 2011;5(4):573–89.
[2] Fuhrman C, Fleury B, Nguyên X-L, Delmas M-C. Symptoms of sleep apnea syndrome: high prevalence and underdiagnosis in the French population. Sleep Med 2012;13(7):852–8.
[3] Kapur V, Strohl KP, Redline S, Iber C, O'connor G, Nieto J. Underdiagnosis of sleep apnea syndrome in US communities. Sleep Breath 2002;6(02):049–54.
[4] Simpson L, Hillman DR, Cooper MN, Ward KL, Hunter M, Cullen S, et al. High prevalence of undiagnosed obstructive sleep apnea in the general population and methods for screening for representative controls. Sleep Breath 2013;17(3):967–73.
*[5] Senaratna CV, Perret JL, Lodge C, Lowe A, Campbell BE, Matheson MC, et al. Prevalence of obstructive sleep apnea in the general population: a systematic review. Sleep Med Rev 2016. http://dx.doi.org/10.1016/j.smrv.2016.07.002.
*[6] Heinzer R, Vat S, Marques-Vidal P, Marti-Soler H, Andries D, Tobback N, et al. Prevalence of sleep-disordered breathing in the general population: the HypnoLaus study. Lancet Respir Med 2015;3(4):310–8.
[7] Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. Increased prevalence of sleep-disordered breathing in adults. Am J Epidemiol 2013;177(9):1006–14.
[8] BaHammam AS, Obeidat A, Barataman K, Bahammam SA, Olaish AH, Sharif MM. A comparison between the AASM 2012 and 2007 definitions for detecting hypopnea. Sleep Breath 2014;18(4):767–73.
[9] Patil SP, Schneider H, Schwartz AR, Smith PL. Adult obstructive sleep apnea: pathophysiology and diagnosis. Chest 2007;132(1):325–37.
[10] Collop NA, Anderson WM, Boehlecke B, Claman D, Goldberg R, Gottlieb D, et al. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. J Clin Sleep Med 2007;3(7):737–47.
[11] Qaseem A, Dallas P, Owens DK, Starkey M, Holty J-EC, Shekelle P. Diagnosis of obstructive sleep apnea in adults: a clinical practice guideline from the american college of physicians. Ann Intern Med 2014;161(3):210–20.
[12] Jonas DE, Amick HR, Feltner C, Weber RP, Arvanitis M, Stine A, et al. Screening for obstructive sleep apnea in adults: evidence report and systematic review for the us preventive services task force. JAMA 2017;317(4):415–33.
[13] Walia R, Achilefu A, Crawford S, Jain V, Wigley SD, McCarthy LH. Are at-home sleep studies performed using portable monitors (PMs) as effective at diagnosing obstructive sleep apnea (OSA) in adults as sleep laboratory-based polysomnography (PSG)? J Okla State Med Assoc 2014;107(12):642–4.
[14] US Preventive Services Task Force. Screening for obstructive sleep apnea in adults: us preventive services task force recommendation statement. JAMA 2017;317(4):407–14.
[15] Cooksey JA, Balachandran JS. Portable monitoring for the diagnosis of OSA. Chest 2016;149(4):1074–81.
[16] Kimoff RJ. When to suspect sleep apnea and what to do about it. Can J Cardiol 2015;31(7):945–8.
[17] Sleep-related breathing disorders in adults. Recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. Sleep 1999;22(5):667–89.
[18] Iber C, Ancoli-Israel Jr S, Chesson A, Quan SF. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Westchester, IL: Am Acad Sleep Med; 2007. p. 1–57.
[19] Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus CL, Voughn BV. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Darien, IL: American Academy of Sleep Medicine; 2012.
[20] Ruehland WR, Rochford PD, O'Donoghue FJ, Pierce RJ, Singh P, Thornton AT. The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. Sleep 2009;32(2):150–7.
[21] Abrishami A, Khajehdehi A, Chung F. A systematic review of screening questionnaires for obstructive sleep apnea. Can J Anaesth 2010;57(5):423–38.
*[22] Chiu H-Y, Chen P-Y, Chuang L-P, Chen N-H, Tu Y-K, Hsieh Y-J, et al. 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 2016. http://dx.doi.org/10.1016/j.smrv.2016.10.004.
*[23] Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med 1999;131(7):485–91.
[24] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529–36.
[25] Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003;56(11):1129–35.
[26] Danzi-Soares NJ, Genta PR, Nerbass FB, Pedrosa RP, Soares FSN, Cesar LAM, et al. Obstructive sleep apnea is common among patients referred for coronary artery bypass grafting and can be diagnosed by portable monitoring. Coron Artery Dis 2012;23(1):31–8.
[27] Firat H, Yuceege M, Demir A, Ardic S. Comparison of four established questionnaires to identify highway bus drivers at risk for obstructive sleep apnea in Turkey. Sleep Biol Rhythms 2012;10(3):231–6.
[28] Sagaspe P, Leger D, Taillard J, Bayon V, Chaumet G, Philip P. Might the Berlin Sleep Questionnaire applied to bed partners be used to screen sleep apneic patients? Sleep Med 2010;11(5):479–83.
[29] Subramanian S, Hesselbacher SE, Aguilar R, Surani SR. The NAMES assessment: a novel combined-modality screening tool for obstructive sleep apnea. Sleep Breath 2011;15(4):819–26.
[30] Bouloukaki I, Komninos ID, Mermigkis C, Micheli K, Komninou M, Moniaki V, et al. Translation and validation of Berlin questionnaire in primary health care in Greece. BMC Pulm Med 2013;13(1).
[31] Kang K, Park KS, Kim JE, Kim SW, Kim YT, Kim JS, et al. Usefulness of the Berlin Questionnaire to identify patients at high risk for obstructive sleep apnea: a population-based door-to-door study. Sleep Breath 2013;17(2):803–10.
[32] Laporta R, Anandam A, El-Solh AA. Screening for obstructive sleep apnea in veterans with ischemic heart disease using a computer-based clinical decision-support system. Clin Res Cardiol 2012;101(9):737–44.
[33] Nunes FS, Danzi-Soares NJ, Genta PR, Drager LF, Cesar LAM, Lorenzi-Filho G. Critical evaluation of screening questionnaires for obstructive sleep apnea in patients undergoing coronary artery bypass grafting and abdominal surgery. Sleep Breath 2015;19(1):115–22.
[34] Pereira EJ, Driver HS, Stewart SC, Fitzpatrick MF. Comparing a combination of validated questionnaires and level III portable monitor with polysomnography to diagnose and exclude sleep apnea. J Clin Sleep Med 2013;9(12):1259–66.
[35] Wilson DL, Walker SP, Fung AM, O'Donoghue F, Barnes M, Howard M. Can we predict sleep-disordered breathing in pregnancy? The clinical utility of symptoms. J Sleep Res 2013;22(6):670–8.
[36] Yuceege M, Firat H, Sever O, Demir A, Ardic S. The effect of adding gender item to Berlin Questionnaire in determining obstructive sleep apnea in sleep clinics. Ann Thorac Med 2015;10(1):25–8.
[37] Gjevre JA, Taylor-Gjevre RM, Nair BV, Lim HJ. Do sleepy rheumatoid arthritis patients have a sleep disorder? Musculoskelet Care 2012;10(4):187–95.
[38] Hrubos-Strom H, Randby A, Namtvedt SK, Kristiansen HA, Einvik G, Benth J, et al. A Norwegian population-based study on the risk and prevalence of obstructive sleep apnea the Akershus Sleep Apnea Project (ASAP). J Sleep Res 2011;20(1 PART II):162–70.
[39] Sert Kuniyoshi FH, Zellmer MR, Calvin AD, Lopez-Jimenez F, Albuquerque FN, Van Der Walt C, et al. Diagnostic accuracy of the Berlin questionnaire in detecting sleep-disordered breathing in patients with a recent myocardial infarction. Chest 2011;140(5):1192–7.
[40] Khaledi-Paveh B, Khazaie H, Nasouri M, Ghadami MR, Tahmasian M. Evaluation of Berlin questionnaire validity for sleep apnea risk in sleep clinic populations. Basic Clin Neurosci 2016;7(1):43–8.
[41] Saleh ABM, Ahmad MA, Awadalla NJ. Development of Arabic version of Berlin questionnaire to identify obstructive sleep apnea at risk patients. Ann Thorac Med 2011;6(4):212–6.
[42] Pataka A, Daskalopoulou E, Kalamaras G, Fekete Passa K, Argyropoulou P. Evaluation of five different questionnaires for assessing sleep apnea syndrome in a sleep clinic. Sleep Med 2014;15(7):776–81.
[43] Kim B, Lee EM, Chung YS, Kim WS, Lee SA. The utility of three screening questionnaires for obstructive sleep apnea in a sleep clinic setting. Onsei Med J 2015;56(3):684–90.
[44] Ulasli SS, Gunay E, Koyuncu T, Akar O, Halici B, Ulu S, et al. Predictive value of Berlin Questionnaire and Epworth Sleepiness Scale for obstructive sleep apnea in a sleep clinic population. Clin Resp J 2014;8(3):292–6.
[45] Margallo VS, Muxfeldt ES, Guimarães GM, Salles GF. Diagnostic accuracy of the Berlin questionnaire in detecting obstructive sleep apnea in patients with resistant hypertension. J Hypertens 2014;32(10):2030–7.
[46] Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, et al. Validation of the Berlin questionnaire and American Society of Anesthesiologists checklist as screening tools for obstructive sleep apnea in surgical patients. Anesthesiology 2008;108(5):822–30.
[47] Thurtell MJ, Bruce BB, Rye DB, Newman NJ, Biousse V. The Berlin questionnaire screens for obstructive sleep apnea in idiopathic intracranial hypertension. J Neuroophthalmol 2011;31(4):316–9.

* The most important references are denoted by an asterisk.

[48] Friedman M, Wilson MN, Pulver T, Pandya H, Joseph NJ, Lin HC, et al. Screening for obstructive sleep apnea/hypopnea syndrome: subjective and objective factors. Otolaryngol Head Neck Surg 2010;142(4):531–5.
[49] Karakoc O, Akcam T, Genc H, Yetkin S, Piskin B, Gerek M. Use of the Berlin Questionnaire to screen at-risk patients for obstructive sleep apnea. B-ent 2014;10(1):21–5.
[50] Ahmadi N, Chung SA, Gibbs A, Shapiro CM. The Berlin questionnaire for sleep apnea in a sleep clinic population: relationship to polysomnographic measurement of respiratory disturbance. Sleep Breath 2008;12(1):39–45.
[51] Luo J, Huang R, Zhong X, Xiao Y, Zhou J. STOP-bang questionnaire is superior to epworth sleepiness scales, Berlin questionnaire, and STOP questionnaire in screening obstructive sleep apnea hypopnea syndrome patients. Chin Med J 2014;127(17):3065–70.
[52] Wu X, Liu Z, Chang SC, Fu C, Li W, Jiang H, et al. Screening and managing obstructive sleep apnea in nocturnal heart block patients: an observational study. Respir Res 2016;17(1) (no pagination)(16).
[53] Marti-Soler H, Hirotsu C, Marques-Vidal P, Vollenweider P, Waeber G, Preisig M, et al. The NoSAS score for screening of sleep-disordered breathing: a derivation and validation study. Lancet Respir Med 2016;4(9):742–8.
[54] McNicholas WT. Screening for sleep-disordered breathing: the continuing search for a reliable predictive questionnaire. Lancet Respir Med 2016;4(9):683–5.
[55] Corral-Peñafiel J, Pepin J-L, Barbe F. Ambulatory monitoring in the diagnosis and management of obstructive sleep apnea syndrome. Eur Respir Rev 2013;22(129):312–24.
[56] Silva GE, Vana KD, Goodwin JL, Sherrill DL, Quan SF. Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth Sleepiness Scales. J Clin Sleep Med 2011;7(5):467–72.
*[57] Nagappa M, Liao P, Wong J, Auckley D, Ramachandran SK, Memtsoudis S, et al. 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 2015;10(12):e0143697.
*[58] Redline S. Screening for Obstructive sleep apnea: implications for the sleep health of the population. JAMA 2017;317(4):368–70.
[59] Safadi A, Etzioni T, Fliss D, Pillar G, Shapira C. The effect of the transition to home monitoring for the diagnosis of OSAS on test availability, waiting time, patients' satisfaction, and outcome in a large health provider system. Sleep Disord 2014;2014:1–6. http://dx.doi.org/10.1155/2014/418246 (Article ID 418246).
[60] Ronald J, Delaive K, Roos L, Manfreda J, Bahammam A, Kryger MH. Health care utilization in the 10 years prior to diagnosis in obstructive sleep apnea syndrome patients. Sleep 1999;22(2):225–9.
[61] Appleton SL, Vakulin A, McEvoy RD, Vincent A, Martin SA, Grant JF, et al. Undiagnosed obstructive sleep apnea is independently associated with reductions in quality of life in middle-aged, but not elderly men of a population cohort. Sleep Breath 2015;19(4):1309–16.
[62] Bjornsdottir E, Keenan BT, Eysteinsdottir B, Arnardottir ES, Janson C, Gislason T, et al. Quality of life among untreated sleep apnea patients compared with the general population and changes after treatment with positive airway pressure. J Sleep Res 2015;24(3):328–38.
[63] Santaolalla CJE, del Campo Matias F. Work-related accidents, absenteeism and productivity in patients with sleep apnea. A future consideration in occupational health assessments? Arch Bronconeumol 2015;51(5):209–10.
[64] Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Ann Intern Med 2002;137(7):598–602.
[65] Bixler EO, Vgontzas AN, Gaines J, Fernandez-Mendoza J, Calhoun SL, Liao D. Moderate sleep apnea: a “silent” disorder, or not a disorder at all? Eur Respir J 2016;47(1):23–6.
[66] Arnardottir ES, Bjornsdottir E, Olafsdottir KA, Benediktsdottir B, Gislason T. Obstructive sleep apnea in the general population: highly prevalent but minimal symptoms. Eur Respir J 2016;47(1):194–202.
[67] Westlake K, Polak J. Screening for Obstructive sleep apnea in type 2 diabetes patients – questionnaires are not good enough. Front Endocrinol 2016;7(Article 124):1–3.
[68] Geiger-Brown J, Rogers VE, Han K, Trinkoff A, Bausell RB, Scharf SM. Occupational screening for sleep disorders in 12-h shift nurses using the Berlin Questionnaire. Sleep Breath 2013;17(1):381–8.
[69] Amra B, Nouranian E, Golshan M, Fietze I, Penzel T. Validation of the Persian version of berlin sleep questionnaire for diagnosing obstructive sleep apnea. Int J Prev Med 2013;4(3):334–9.
[70] Elkholy SH, Amer HA, Nada MM, Nada MAF, Labib A. Sleep-related breathing disorders in cerebrovascular stroke and transient ischemic attacks: a comparative study. J Clin Neurophysiol 2012;29(2):194–8.
[71] Ha SCN, Lee DLY, Abdullah VJ, van Hasselt CA. Evaluation and validation of four translated Chinese questionnaires for obstructive sleep apnea patients in Hong Kong. Sleep Breath 2014;18(4):715–21.
[72] Suksakorn S, Rattanaumpawan P, Banhiran W, Cherakul N, Chotinaiwattarakul W. Reliability and validity of a Thai version of the Berlin questionnaire in patients with sleep disordered breathing. J Med Assoc Thai 2014;97(3 Suppl. 3):S46–56.
[73] Yunus A, Seet W, Mohamad AB, Haniff J. Validation of the Malay version of Berlin questionaire to identify Malaysian patients for obstructive sleep apnea. Malays Fam Physician 2013;8(1):5–11.
[74] Drager LF, Genta PR, Pedrosa RP, Nerbass FB, Gonzaga CC, Krieger EM, et al. Characteristics and predictors of Obstructive sleep apnea in patients with systemic hypertension. Am J Cardiol 2010;105(8):1135–9.
[75] Sforza E, Chouchou F, Pichot V, Herrmann F, Barthelemy JC, Roche F. Is the Berlin questionnaire a useful tool to diagnose obstructive sleep apnea in the elderly? Sleep Med 2011;12(2):142–6.
[76] Gupta R, Ali R, Dhyani M, Das S, Pundir A. Hindi translation of Berlin questionnaire and its validation as a screening instrument for obstructive sleep apnea. J Neurosci Rural Pract 2016;7(2):244–7.