Review Article

Screening Tools for Obstructive Sleep Apnea in Pregnant Women: An Extended and Updated Systematic Review and Meta-analysis

Babak Amra, Masoud Mansouri[1], Forogh Soltaninejad, Awat Feizi[2], Marta Kaminska[3]

Bamdad Respiratory and Sleep Research Center, Pulmonary and Sleep Ward, Department of Internal Medicine, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran, [1]Department of Internal Medicine, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran, [2]Department of Epidemiology and Biostatistics, School of Health, Isfahan University of Medical Sciences, Isfahan, Iran, [3]Respiratory Division, McGill University Health Centre, Montreal, Quebec, Canada

Abstract

The prevalence of obstructive sleep apnea syndrome (OSA) increases in women during pregnancy and negatively affects maternal and fetal outcomes. The updated systematic review and meta-analysis aimed to evaluate the validity of the Berlin, STOP-Bang, and Epworth sleepiness scale (ESS) questionnaires in detecting OSA in pregnant women. PubMed, Embase, and Web of Science were searched systematically up to March 2022. After eligible studies inclusion, two independent reviewers extracted demographic and clinical data. Bivariate random effects models were used to estimate the pooled accuracy measures including sensitivity and specificity, positive (PPV) and negative predictive values (NPVs), diagnostic odds ratio (DOR), and receiver operating characteristic curve (ROC) curve. We included 8 studies including 710 pregnant women with suspected OSA. The performance values of Berlin, STOP-Bang, and ESS questionnaires were as follows: the pooled sensitivity were 61% (95% confidence interval (CI): 40%–80%), 59% (95% CI: 49%–69%), and 29%, (95% CI: 10%–60%); pooled specificity were 61% (95% CI: 42%–78%), 80% (95% CI: 55%–93%), and 80% (95% CI: 50%–94%); pooled PPVs were 60% (95% CI: 0.49–0.72), 73% (95% CI: 61%–85%), and 59% (95% CI: 31%–87%); pooled NPVs were 60% (95% CI: 0.49–0.71), 65% (95% CI: 54%–76%), and 53% (95% CI: 41%–64%); and pooled DORs were 3 (95% CI: 1–5), 6 (95% CI: 2–19), and 2 (95% CI: 1–3), respectively. It seems that the Berlin, STOP-Bang, and ESS questionnaires had poor to moderate sensitivity and specificity in pregnancy, with the ESS showing the worst characteristics. Further studies are required to evaluate the performance of alternative screening methods for OSA in pregnancy.

Keywords: *Berlin questionnaire, Epworth sleepiness Scale, obstructive sleep apnea, pregnant women, STOP-bang questionnaire*

Introduction

Sleep-disordered breathing (SDB) is a major public health problem with a prevalence of 5%–20%. Obstructive sleep apnea (OSA) is the most common form of SDB in adults.[1,2] OSA is a chronic condition manifested by recurrent collapse of the upper airway during sleep. The episodes result in oxygen desaturation, cortical sleep arousal, sleep fragmentation, and sympathetic activation.[3] Patients with OSA may report symptoms such as snoring, choking during sleep, excessive daytime sleepiness, non-restorative sleep, poor sleep quality, fatigue, and morning headaches.[4,5] Additionally, untreated OSA is associated with chronic physical and mental health conditions such as cardiovascular disease, hypertension, diabetes mellitus, depression, and cognitive impairment.[6-9]

Previously, it was believed that OSA is much more prevalent in men. However, it has recently been detected that OSA is not a rare sleep disorder in women but remains underdiagnosed because of its different manifestations in the population.[10,11] Furthermore, the prevalence of OSA probably increases during pregnancy because of physiological and hormonal changes.[12,13] Previous research indicated that OSA is associated with an increased risk of adverse maternal (e.g., gestational diabetes, preeclampsia, postoperative wound complication, pulmonary edema, and cesarian section) and fetal outcomes (e.g., pre-term birth and increased rate of neonatal intensive care unit admission).[14] Thus, early diagnosis and efficacious management of OSA in pregnant women may be helpful in decreasing unfavorable maternal and fetal outcomes.[15]

Polysomnography (PSG) is the gold standard diagnostic test for OSA. Generally,

How to cite this article: Amra B, Mansouri M, Soltaninejad F, Feizi A, Kaminska M. Screening tools for obstructive sleep apnea in pregnant women: An extended and updated systematic review and meta-analysis. Int J Prev Med 2024;15:31.

Address for correspondence:
Dr. Forogh Soltaninejad, Bamdad Respiratory and Sleep Research Center, Pulmonary and Sleep Ward, Department of Internal Medicine, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. E-mail: Soltaninejad.fg@gmail.com; f.soltaninejad@med.mui.ac.ir

Access this article online

Website: www.ijpvmjournal.net/www.ijpm.ir

DOI: 10.4103/ijpvm.ijpvm_88_23

Quick Response Code:

PSG requires a sleep laboratory and trained staff to set up the sleep study and supervise it overnight, rendering it a resource-intensive and costly test.[16] Therefore, there is a need to establish a reliable, quick, and cost-effective screening tool for OSA. Previous researchers have investigated the validity of some tools such as the Berlin questionnaire,[17] STOP-Bang questionnaire,[18] Epworth Sleepiness Scale (ESS),[19] and American Anesthesiologists checklist[20] in OSA screening among pregnant women.[21,22] Validation studies assessing the performance of these tools during pregnancy have shown inconsistent results. Furthermore, the validity of the tests greatly depends on the severity of SDB itself and the trimester of pregnancy.[13,18-21]

Despite the multitude of studies in the field of OSA screening tools, there is no certainty in using these tools and choosing the best and simplest tool for diagnosing OSA in pregnant women. In addition, so far, few studies have evaluated the different diagnostic accuracy of these tools in pregnant women; the higher the diagnostic accuracy of the selected tool, the better and faster it is possible to diagnose the disease and control related complications in this population. Therefore, this systematic review and meta-analysis study aims to evaluate the validity of different screening tools for OSA (including Berlin, ESS, and STOP-Bang questionnaires) in pregnant women by reviewing previous studies in order to introduce the most appropriate tool.

Methods

The protocol of the systematic review and meta-analysis was registered with the International Prospective Register of Systematic Review (PROSPERO) under the registration number CRD42021283101.

Search strategy

Two independent authors comprehensively and independently conducted a complete and systematic search of the published manuscripts from 1980 to March 2022 in PubMed, Medline, Scopus, Embase, Web of Science, and Google Scholar.

The search was performed using the keywords obtained from the Medical Subject Headings (MeSH) database and the keywords of other articles published in this field by combining AND and OR operators. The search was done using the following keywords: "Pregnancy," "Pregnant women," "Obstructive sleep apnea syndrome," "Sleep apnea syndrome," "Obstructive sleep apnea," "Sleep hypopnea," "Sleep-disordered breathing," "Screening tool," "Validation," "Questionnaire." Multiple searches were performed during the writing of this manuscript. The general search strategy for all databases is as ("Pregnancy" OR "Pregnant women") AND ("obstructive sleep apnea syndrome" OR "sleep apnea syndrome" OR "obstructive sleep apnea" OR "sleep hypopnea" OR "sleep-disordered breathing") AND ("screening tool" OR "validation" OR "questionnaire").

After removing duplicates, the reference list of included studies and relevant review articles was manually assessed. The titles and abstracts of publications were checked against eligibility criteria by two independent authors and full texts of potentially relevant studies were obtained. Any disagreement was resolved by consensus.

Inclusion and exclusion criteria

Results of the systematic search were included in the systematic review and meta-analysis if they met the following inclusion criteria: 1) the study evaluated screening tools for OSA in pregnant women, 2) OSA diagnosis was confirmed by a standard PSG or home sleep apnea test (AASM level 3), 3) OSA was defined based on apnea-hypopnea index (AHI) or respiratory disturbance index (RDI). All case reports, editorials, and reviews were excluded from the study. In addition, conference abstracts and references of the presented articles were also searched to find any relevant data to add to the review. In this study, only articles in the English language were included.

Quality assessment

Two independent reviewers appraised the quality of included studies according to the quality assessment of diagnostic accuracy studies-2 (QUADAS-2).[23] It comprises four main sections including patient selection, index test, reference standards, and the flow and timing of the tests.

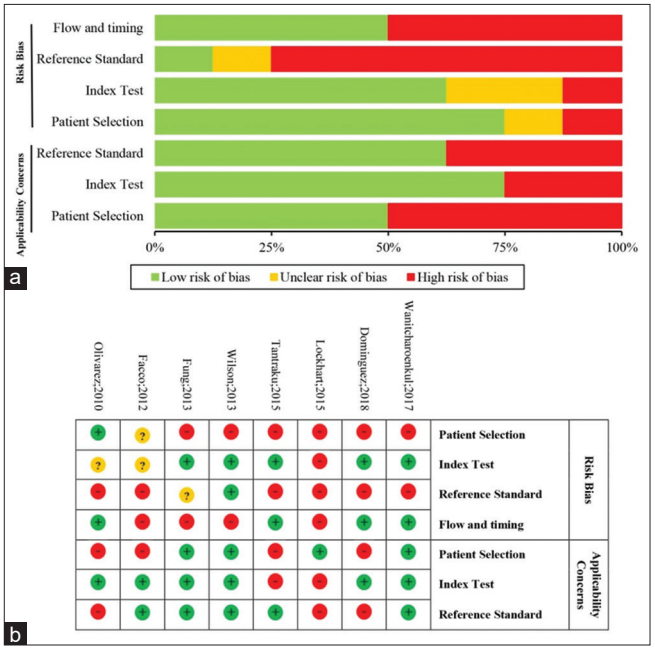

Figure 1: (a) The risk of bias graph showing the authors' judgements about each risk of bias item presented as percentages across all included studies. (b) The risk of bias summary based on the authors' judgements about each risk of bias item for each included study

Each part involves the assessment of the risk of bias and clinical applicability. Any disagreement was resolved by consensus [Figure 1].

Study selection and data collection

The extracted data comprised the name of the first author, year and country of publication, type of study, sample size, type of reference test, OSA definition, age, gestational age, as well as screening tools. A 2 × 2 contingency table was created to calculate validity parameters if they were not reported in the article. Data extraction from the selected articles was performed by two authors independently. In case of differences in information recordings, they were discussed and the best information extracted from each study was approved following a final decision.

Data synthesis and statistical analysis

We followed the standard methods recommended for diagnostic accuracy meta-analysis studies.[24] For each study, the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) was retrieved and then sensitivity and specificity were calculated for each index sleep apnea test screening, when available. The sensitivity of a diagnostic test is defined as the probability that the index test result will be positive in a patient with definite diagnosed OSA in pregnant women based on the gold standard while specificity is defined as the probability that the diagnostic test result will be negative in non-affected pregnant women by sleep apnea. For each study, the sensitivity and specificity of each test along with a 95% confidence interval (CI) were calculated by using the exact binomial method.[25] The diagnostic odds ratio (DOR) is a summary estimate of how many times higher the odds are of obtaining a positive test result in a diseased rather than a non-diseased pregnant woman with OSA. DOR is a useful measure if there is no preference for either superior sensitivity or specificity and the focus is on global performance for comparing different tests. If the DOR is less than one, then the test is uninformative and that test has no clinical value.[26]

A hierarchical summary receiver operating characteristic (HSROC) curve in a bivariate setting in order to consider the inter-correlation between sensitivity and specificity was created for each studied test to provide an overall summary of the diagnostic test accuracy data. The estimates from the HSROC model are used to plot a summary ROC curve of sensitivity versus specificity (expressed as 1 − specificity), the 95% confidence region around this summary estimate, and a 95% prediction region taking into account unobserved heterogeneity: if a new study was conducted, we would expect the "true" sensitivity and specificity to lie within the prediction region with a 95% confidence level. The prediction region can be wider than the confidence region as it goes beyond the uncertainty in the available data.[26]

We also estimated the likelihood ratio (LR) for each index text by using a bivariate meta-analysis. The positive LR >1 for a positive test result is associated with the presence of disease, and negative LR <1 for a negative test result is associated with the absence of disease.[27]

Heterogeneity was evaluated by the I^2 measure and tested by the Cochran Q Chi-squared test and visually explored by forest plot for the DOR of index tests with 95% CIs for each individual study. I^2 values of 25% or less, 50% or less, and 75% or less are used to denote low, moderate, and high heterogeneity, respectively.[28] We conducted a bivariate meta-regression analysis to assess what proportion of heterogeneity in the estimated pooled accuracy measures is explained by the potential confounding effects of age and gestational age during pregnancy. Sensitivity analysis was also conducted to assess the impact of each study on pooled estimated measures by removing studies one at a time. The presence of publication bias was tested by using Deeks' funnel plots.[29] All data synthesis was done using the *metandi* and midas commands package in STATA MP v11.2.[26,30]

Results

Our initial systematic search yielded 6,170 studies. After excluding duplicates and irrelevant studies, 23 full texts were screened for eligibility. Finally, 8 studies were included in the systematic review and meta-analysis [Figure 2].

The basic and validation characteristics of the included studies have been summarized in Table 1. The included

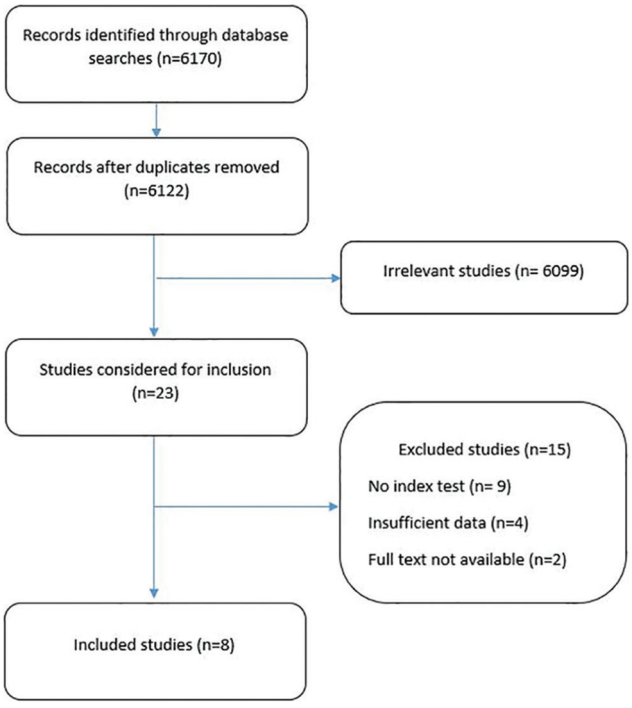

Figure 2: Flowchart of study selection for the systematic review and meta-analysis

Table 1: Characteristics of included studies

First author, year	Olivarez, 2010[32]	Facco, 2010[13]	Fung, 2013[34]	Wilson, 2013[33]	Tantrakul, 2015[36]	Lockhar, 2015[21]	Wanitcharoenkul, 2017[35]	Dominguez, 2018[22]
Location	USA	USA	Australia	Australia	Thailand	USA	Thailand	USA
Study design	Prospective cohort	Prospective cohort	Prospective cohort	Prospective cohort	Prospective observational	Prospective cohort	Prospective cohort	Prospective cohort
Number	100	100	41	43	72	248	82	24
Age (Year)	26.6±7.1	33.0±6.5	31.2±NR	33.5±5.1	33.1±5.2	28.0±6.3	31.4±6.1	30±1.0
Gestational age (week)	32.3±3.5	16.5±3.7	21.4±2.4	22.3±4.0	22.8±9.2	32.0±3.1	29.0±0.96	29.85±0.85
BMI; kg/m^2	Pre-pregnancy: NA	Pre-pregnancy: 31.9±9.1	Pre-pregnancy: 26.1±6.4	Pre-pregnancy: 32.2±8.0	Pre-pregnancy: 24.2±5.3	Pre-pregnancy: NA	Pre-pregnancy: 28.9 (26.1-31.7)†	Pre-pregnancy: NA
	Pregnancy: 27.5±7.2	Pregnancy: NA	Pregnancy: NA	Pregnancy: 37.5±7.9	Pregnancy: 26.9±5.3	Pregnancy: 31 (27-36)†	Pregnancy: NA	Pregnancy: 54.4 (44.2-59.6)†
Validation tool	Apnea link	Watch-PAT100 (WP100)	***PSG	PSG	Watch-PAT200 (WP200)	Apnea link	Watch-PAT200 (WP200)	PSG
Severity OSA	NA	Overall AHI: 1.5 (0.5-6)	RDI with OSA: 6.2 (4.9-11.7)†	AHI with OSA: 6.2 (4.9-13.2)†	Overall AHI: 2.4±8.0	NA	AHI with OSA: 9.4 (6.4-12.4)	AHI with OSA: 8.7 (6.1-12.1)†
			Without OSA: 1.4 (0.6-2.6)†	Without OSA: 1.5 (0.6-2.7)†			Without OSA: 1.9 (1.3-3.7)†	Without OSA: 1.3 (0.5-2.8)†
*OSA definition	**AHI ≥5	AHI ≥5	RDI ≥5	#RDI ≥5	AHI ≥5	AHI ≥5	AHI ≥5	AHI ≥5
OSA prevalence	20%	28%	34%	35%	31.9%	12%	52.4%	24%
Applied questionnaire								
Berlin	2 or more categories	2 or more categories	At least 2 out of 3 categories	At least 2 out of 3 categories	At least 2 out of 3 categories	At least 2 out of 3 categories	≥1 categories	Median score=2
Epworth sleepiness scale	-	Score 10 or more	-	-	≥ 10	≥10	-	Median score=3
STOP-Bang	-	-	-	-	2 or more categories	2 or more categories	-	Median score=4

*OSA; Obstructive Sleep Apnea, **AHI; Apnea Hypopnea Index, ***PSG; Polysomnography, #RDI; Respiratory Disturbance Index. †: Median (IQR)

studies contained 710 pregnant women with suspected OSA. The mean age ranged from 26.6 to 33.5 years and the mean gestational age ranged from 16.5 to 32.3 weeks. The prevalence of OSA (defined as AHI or RDI ≥5) ranged from 12% to 52.4%. Studies were performed in the USA,[21,22,31,32] Australia,[33,34] and Thailand.[20,35]

Predictive parameters of Berlin questionnaire in pregnant women

For the final analyses of the Berlin questionnaire eight studies were included.[20-22,31-35] Heterogeneity assessment showed a significant heterogeneity across studies (I^2 = 98%, P < 0.001); accordingly, the pooled estimation of diagnostic characteristics was done by a bivariate random effect model. The sensitivity of the questionnaire ranged from 16.3% to 93% across various studies [Figure 3a]. The pooled sensitivity of Berlin questionnaire was 61% (95% CI: 40%–80%, I2 = 87.02, P < 0.001) [Table 2 and Figure 3a]. The specificity of the questionnaire ranged from 19.7% to 89.7% [Figure 3a] and its pooled specificity was 61% (95% CI: 42%–78%, I2 = 91.64, P < 0.001) [Table 2 and Figure 3a]. Additionally, the pooled positive predictive value (PPV) and negative predictive value (NPV) of the Berlin questionnaire were 60% (95% CI: 49%–72%) and 60% (95% CI: 49%–71%), respectively. The pooled LR+ and LR- were 1.6 (95% CI: 1.10–2.20) and 0.63 (95% CI: 0.43–0.93), respectively. Pooled DOR was 3.00 (95% CI: 1.00–5.00) [Table 2]. The area under the ROC curve was 0.65 (95%CI: 0.61–0.65) [Figure 3b]; most of the points clustered around a bit higher than the reference line on the top left of the graph, indicating the moderate accuracy of the test.

To find the probable effect of age and gestational age on significant heterogeneity of pooled estimated sensitivity and specificity of the questionnaire, a bivariate meta-regression was conducted. The resulted sensitivity and specificity of meta-regression for age were 63 (95% CI: 39–82), and 63 (95% CI: 42–80), respectively (P > 0.05). The sensitivity and specificity of meta-regression for gestational age were 61 (95% CI: 39–79), and 61 (95% CI: 42–78), respectively (P > 0.05). The findings showed that neither age nor gestational age could significantly explain the heterogeneity in estimated pooled sensitivity and specificity of the Berlin questionnaire (P > 0.05). Findings from Deeks' funnel plot test showed a non-significant publication bias (P = 0.59). The results of pooled analysis by excluding the Dominguez study[22] were as follows: the pooled sensitivity: 0.59 (95% CI: 0.35–0.79), pooled specificity: 0.67 (95% CI: 0.51–0.80), pooled LR+: 1.80 (95% CI: 1.30–2.40), pooled LR-: 0.62 (95% CI: 0.39–0.96), and pooled DOR: 3.00 (95% CI: 1.00–6.00). Furthermore, analysis was performed once again by excluding the Wanitcharoenkul study.[35] The results were as follows: the pooled sensitivity: 0.68 (95% CI: 0.50–0.82), pooled specificity: 0.55 (95% CI: 0.38–0.72), pooled LR+: 1.50 (95% CI: 1.10–2.10), pooled LR-: 0.58 (95% CI: 0.37–0.89), and pooled DOR: 3.00 (95% CI: 1.00–5.00). The mentioned results showed

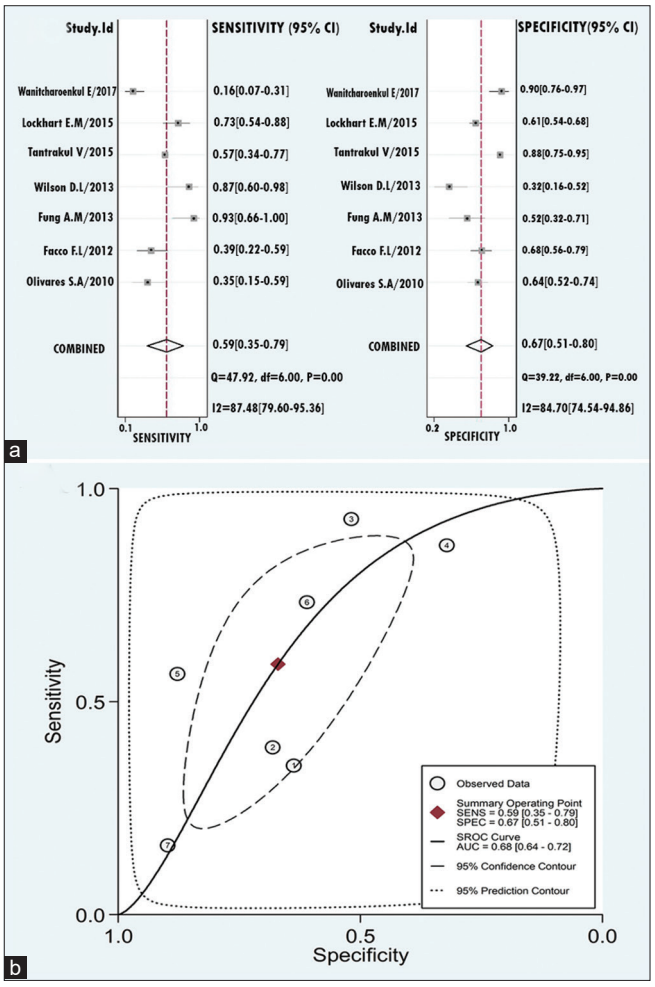

Figure 3: Forest plot (a) and the summary receiver operating characteristic curve (b) for Berlin questionnaire

Table 2: Summary scores for the validity parameters of Berlin, STOP-Bang, and ESS questionnaires administered to pregnant women with suspected OSA

Questionnaire	Sensitivity (95% CI)	Specificity (95% CI)	LR + (95% CI)	LR- (95% CI)	DOR (95% CI)	Chi-square	I^2	P
Berlin	0.61 (0.40–0.80)	0.61 (0.42–0.78)	1.6 (1.1–2.2)	0.63 (0.43–0.93)	3 (1–5)	87.10	98	<0.001
STOP-Bang	0.59 (0.49–0.69)	0.80 (0.55–0.93)	2.9 (1.2–7.4)	0.51 (0.37–0.70)	6 (2–19)	19.42	90.95	<0.001
ESS	0.29 (0.10–0.60)	0.80 (0.50–0.94)	1.5 (0.8–2.6)	0.88 (0.72–1.07)	2 (1–3)	43.46	95.95	<0.001

ESS=Epworth Sleepiness Scale, LR=likelihood ratio, DOR=diagnostic odds ratio, P<0.05 was considered significant

that the omission of any of the studies had no significant effect on the results.

Predictive values of ESS questionnaire in pregnant women

Four studies were included in the final analysis of ESS.[20-22,31] Significant heterogeneity was found among studies (I2 = 95.95%, $P < 0.001$). The sensitivity and specificity of the questionnaire ranged from 0% and 57.5% to 57.7% and 100%. The results of the bivariate random effect model showed that the pooled sensitivity and specificity of ESS were 29% (95% CI: 10%–60%, $I^2 = 78.15$, $P < 0.001$) and 80% (95%CI: 50%–94%, $I^2 = 93.37$, $P < 0.001$), respectively [Table 2 and Figure 4a]. The pooled PPV and NPV of ESS were 59% (95% CI: 31%–87%) and 53% (95% CI: 41%–64%), respectively. The pooled LR+ and LR- were 1.5 (95% CI: 0.8-2.6) and 0.88 (95% CI: 0.72-1.07), respectively. Additionally, pooled DOR was 2 (95% CI: 1–3) [Table 2], and the area under the ROC curve was 0.56 (95%CI: 0.52–0.60); suggesting a moderate accuracy of the tool [Figure 4b].

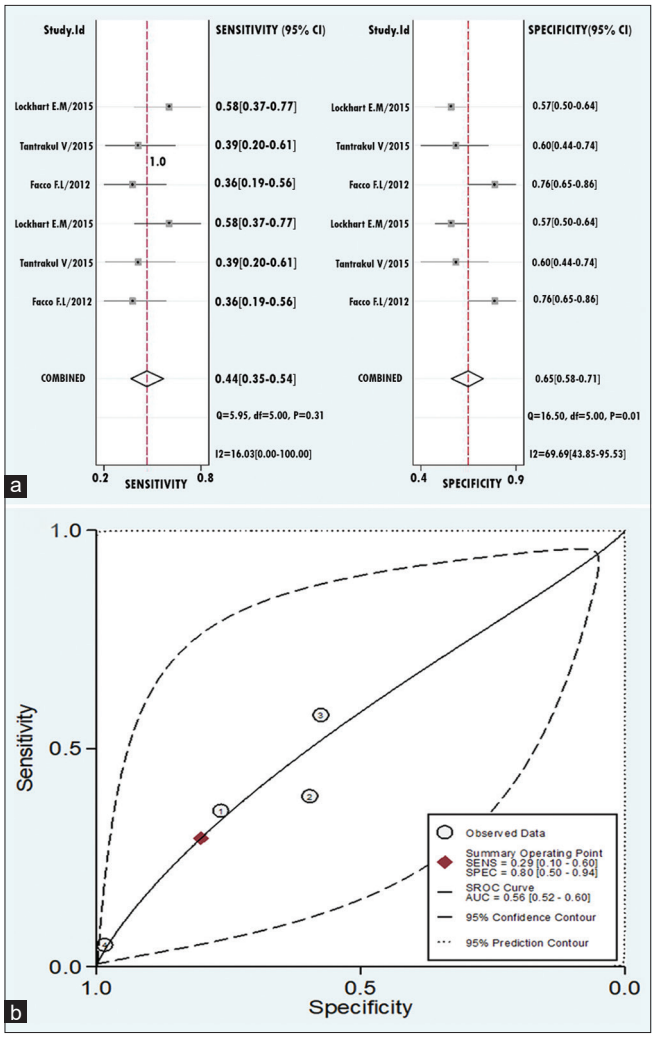

Figure 4: Forest plot (a) and the summary receiver operating characteristic curve (b) for Epworth Sleepiness Scale

The results of the meta-regression showed that age and gestational age did not contribute to the heterogeneity of pooled sensitivity and specificity of ESS ($P > 0.05$). The resulted sensitivity and specificity of meta-regression for age were 30 (95% CI: 10–61), and 79 (95% CI: 44–95), respectively ($P > 0.05$). The sensitivity and specificity of meta-regression for gestational age were 25 (95% CI: 10–50), and 83 (95% CI: 62–94) respectively ($P > 0.05$). The publication bias was not significant according to Deeks' funnel plot test ($P = 0.97$). No significant effect on the results was observed by the omission of studies in the sensitivity analysis. The results of analysis with exclusion of Dominguez study[22] were as follows: the pooled sensitivity: 0.44 (95% CI: 0.35–0.54), pooled specificity: 0.65 (95% CI: 0.58–0.71), pooled LR+: 1.30 (95% CI: 1.00–1.60), pooled LR-: 0.86 (95% CI: 0.73–1.01), and pooled DOR: 1.00 (95% CI: 1.00–2.00).

Predictive values of STOP-Bang questionnaire in pregnant women

Three studies were included in the final analysis of the STOP-Bang questionnaire.[16-18] Significant heterogeneity was observed across studies ($I^2 = 90.95\%$, $P < 0.001$). The pooled sensitivity and specificity were 59% (95% CI: 49%–69%, $I^2 = 0.60$, $P = 0.90$) and 80% (95% CI: 55%–93%, $I^2 = 1.34$, $P = 0.72$), respectively [Figure 5a]. The pooled PPV and NPV of STOP-Bang were 73% (95%CI: 61%–85%) and 65% (95% CI: 54%–76%) [Table 2]. Additionally, the pooled LR+ and LR- were 2.90 (95% CI: 1.20–7.40) and 0.51 (95% CI: 0.37–0.70) [Table 2]. Pooled DOR of the STOP-Bang questionnaire was 6.00 (95% CI: 2.00–19.00) which was higher than that of the Berlin questionnaire and ESS [Table 2]. Moreover, the area under the ROC curve was 0.61 (95% CI: 0.57–0.65); indicating the moderate accuracy of the questionnaire for the diagnosis of OSA in pregnant women [Figure 5b]. The meta-regression results showed that age and gestational age did not contribute to heterogeneity in the estimated pooled sensitivity and specificity of the STOP-Bang questionnaire ($P > 0.05$). The resulted sensitivity and specificity of meta-regression for age were 61 (95% CI: 47–73), and 75 (95% CI: 38–93) respectively ($P > 0.05$). The sensitivity and specificity of meta-regression for gestational age were 59 (95% CI: 48–69), and 76 (95% CI: 57–90) respectively ($P > 0.05$). Furthermore, publication bias was not statistically significant based on Deeks' funnel plot asymmetry test ($P = 0.82$). No significant effect on the results was observed by the omission of studies in the sensitivity analysis. The results of analysis with exclusion of Dominguez study[22] were as follows: the pooled sensitivity: 0.57 (95% CI: 0.47–0.66), pooled specificity: 0.86 (95% CI: 0.83–0.89), pooled LR+: 4.10 (95% CI: 3.10–5.30), pooled LR-: 0.50 (95% CI: 0.40–0.63), and pooled DOR: 8.00 (95% CI: 5.00–13.00).

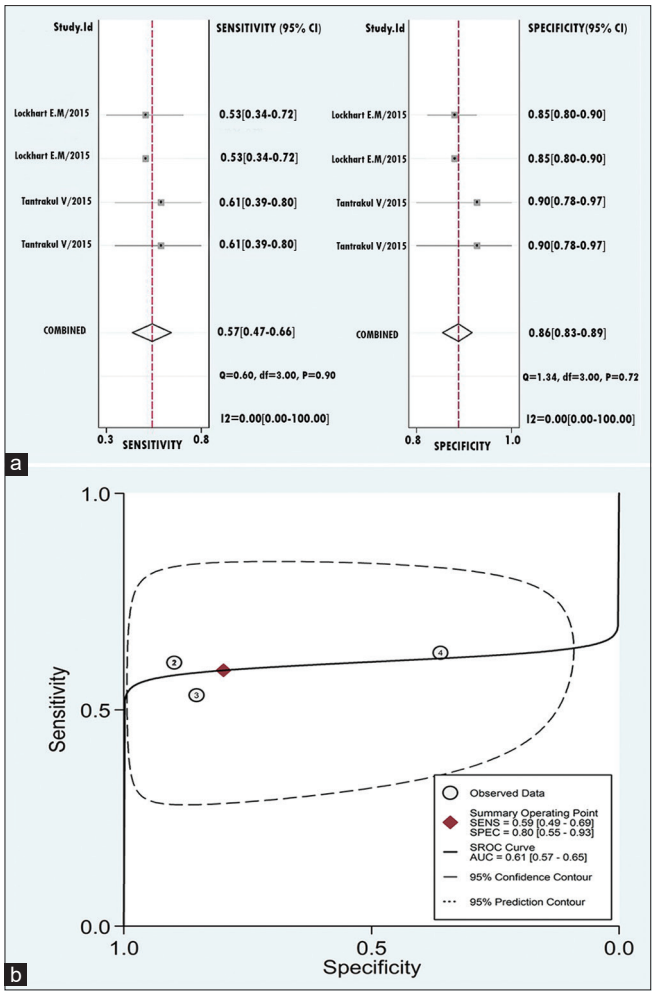

Figure 5: Forest plot (a) and the summary receiver operating characteristic curve (b) for STOP-Bang questionnaire

Discussion

OSA has been known as a frequent health issue in pregnant women with an overall prevalence of 15% in mid- and late pregnancy.[14] Many changes such as weight gain and increasing abdominal girth with gestational age make pregnant women more prone to develop OSA.

Increased risk of adverse maternal and infant health effects is associated with OSA.[14] The lack of institutional screening guidelines for pregnant women results in underdiagnosis of OSA in the population.[37] Thus, it is required to develop accurate and accessible tools for OSA screening in the population. The common OSA screening questionnaires such as Berlin and STOP-Bang have been assessed for OSA screening in the population, as well as the ESS, the most frequently used scale for excessive daytime sleepiness which is a frequent symptom of OSA; however, the results are not consistent. Thus, this updated systematic review and meta-analysis incorporated several recent validation studies to evaluate the performance of OSA screening tools in pregnant women.

In previous studies, the gold standard cut-point values of OSA screening tools have been examined and they have shown that a STOP-Bang score of 5–8 can be considered a high risk of OSA, which has a sensitivity of 100% and a specificity of more than 95%. In the Berlin questionnaire scale, having at least two categories out of three has a sensitivity and specificity of 81.46% and 82.35%, respectively. Additionally, the ESS score of 11 or more is known as a standard cut-off point, so that the sensitivity and specificity obtained from this test were 59% and 76.47%. Therefore, STOP-Bang and Berlin questionnaires had higher sensitivity than ESS.[38-40] It should be noted that these cut points have been determined in the general population, and this is not evaluated in the group of pregnant women. Now, due to the complexity and dynamic changes of OSA and pregnancy, the screening of this condition during pregnancy is complicated and the current OSA screening questionnaires have performed poorly during pregnancy. In this regard, in a previous systematic review in this group of people, it was shown that ESS compared to BQ and ESS had a weak diagnostic value during pregnancy.[36]

Our systematic review and meta-analysis indicated that the Berlin questionnaire had moderate sensitivity (61%) and specificity (61%) with an LR+ of 1.6 for screening clinically relevant OSA in pregnant women. Moderate specificity of the tool predisposes patients to false-positive results and costly sleep studies such as PSG. Additionally, the moderate NPV of 60% suggested that the questionnaire may be of limited screening value in pregnant women at a high risk of OSA. The performance of the Berlin questionnaire has been assessed in various populations such as the elderly, surgical, and sleep clinic populations.[37-41] However, its validity in OSA diagnosing has only been confirmed in the sleep clinic population.[36]

Previously, the performance of the STOP-Bang questionnaire has been investigated widely in the detection of OSA in various populations.[42-44] A recent meta-analysis has demonstrated that the performance of the STOP-Bang questionnaire in OSA screening is superior to ESS and Berlin questionnaires in different populations.[45] The STOP-Bang is a more widely used test in the general population but was not included in the meta-analysis of OSA screening tools in pregnancy by Tantrakul *et al*.[36] According to our findings, the DOR of the STOP-Bang questionnaire was higher than those of the Berlin and ESS questionnaires. The combination of sensitivity and specificity for estimating the likelihood of disease detection compared to the disease prevalence displayed by LR. A higher LR+ value indicates that a positive test result will increase the likelihood of illness. While a lower LR- value means a lower likelihood of illness in the presence of a negative test result. The questionnaire had a higher LR+ and lower LR- compared to the Berlin and EES questionnaires. However, the poor sensitivity of

the STOP-Bang questionnaire (59%), despite its better DOR and LR compared to Berlin and ESS questionnaires, does not make it an ideal tool for OSA screening in the population.

We found that ESS had good specificity (80%) but poor sensitivity (29%) for OSA screening during pregnancy. Previous studies in various populations have also confirmed a lower predictive value of ESS compared to other OSA screening tools.[46-49] The ESS questionnaire has been developed for daytime sleepiness assessment. Therefore, it seems that the use of this tool is not desirable in the screening of OSA, as many sleep disturbances and disorders other than OSA may contribute to hypersomnia.[47] The different results for ESS sensitivity in the study of Dominguez[22] are probably due to the small sample size. It seems that ESS is not a good test to rule out OSA in pregnancy.

Clinically, the high sensitivity of a tool for OSA screening in pregnant women is crucial as it can prioritize women for objective testing and promote early treatment that may reduce associated comorbidities and their sequelae. Missing true positive cases, related to low questionnaire sensitivity, may impose a higher direct and indirect economic burden on the health care system and individuals.

Our meta-regression test for subjects' age and gestational age failed to demonstrate a significant improvement in sensitivity and specificity of examined tools. However, in different approaches, the mother's age, BMI and frequent snoring were strong predictors for OSA in pregnancy.[50]

Facco *et al*.[51] found that symptoms such as frequent snoring, chronic hypertension, older age, and higher BMI can be predictors of OSA in high-risk women at 6–26 weeks of pregnancy. Another study showed that the Berlin questionnaire has a sensitivity of 56.5% and a specificity of 87.8%.[52] In fact, they stated that pre-pregnancy BMI of women was a significant predictor of OSA in the first trimester, while frequent snoring and weight gain during pregnancy predicted OSA in the second to third trimester, respectively.[52] Our study is different because BMI was not an independent predictor in predicting OSA. Perhaps the reason for this difference is the lack of remarkable changes in the BMI range of the studied women, the mean age group of 30 years, and the lack of consideration of arterial oxygen saturation and other clinical symptoms of these individuals.

Although, the prediction models using these variables had good performance for diagnosis of gestational OSA in a meta-analysis, similarly to our study, the heterogeneity and high risk of bias have been mentioned as limitations.[53] Traditional factors like age, BMI, snoring, and hypertension are the most frequently used variables, but other items like tongue enlargement also seemed relevant for OSA prediction in pregnant women. However, contrary to the general population, hypersomnolence does not appear to be strongly associated with OSA in pregnancy, which is congruent with the poor performance of the ESS that we have found.

Presumably, factors such as differences in the type of reference test, studied population, socioeconomic status, and cultural differences may be associated with the significant heterogeneity of our results. The limited number of included studies with small sample sizes and their significant heterogeneity due to different genetic makeup prevent us from definitive conclusions on the performance of investigated questionnaires in pregnant women. A limitation of several studies in our meta-analysis is that level 3 devices have been used as the reference test instead of PSG. These devices may not have been validated in pregnancy and may not be accurate enough to diagnose OSA in pregnancy.[54,55] An exception appears to be the Watch-PAT device, which showed good accuracy compared with PSG in one study,[56] but these results need to be replicated. Our data suggest that alternative screening methods or the combination of screening questionnaires should be developed in future studies and examined compared to PSG as a gold standard test or a well-validated alternative. In addition, one of the necessary policies in order to identify and prevent OSA in pregnant women is to pay attention to their demographic and clinical characteristics (such as age, BMI, daytime respiratory symptoms, and sleep-related complaints, etc.) along with the evaluation by one of the mentioned screening tools.

Conclusion

The STOP-Bang, Berlin, and ESS questionnaires had a suboptimal performance for OSA screening in pregnant women. There is an urgent need to develop alternative screening tools for this population and examining their performance and validity in large-scale studies.

Ethics approval

The project was approved by the ethical committee of the Institutional Board Review of Isfahan University of Medical Sciences, Isfahan, Iran (IR MUI.MED.REC.1400.117).

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Received: 21 Mar 23 **Accepted:** 30 Aug 23
Published: 06 Aug 24

References

1. Durán J, Esnaola S, Rubio R, Iztueta Á. Obstructive sleep apnea–hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 yr. Am J Respir Crit Care Med 2001;163:685-9.
2. Punjabi NM. The epidemiology of adult obstructive sleep apnea. Proc Am Thorac Soc 2008;5:136-43.
3. Pack AI. Advances in sleep-disordered breathing. Am J Respir Crit Care Med 2006;173:7-15.
4. Malhotra A, White DP. Obstructive sleep apnoea. Lancet

2002;360:237-45.
5. Lévy P, Kohler M, McNicholas WT, Barbé F, McEvoy RD, Somers VK, *et al*. Obstructive sleep apnoea syndrome. Nat Rev Dis Primers 2015;1:1-21.
6. Wang X, Ouyang Y, Wang Z, Zhao G, Liu L, Bi Y. Obstructive sleep apnea and risk of cardiovascular disease and all-cause mortality: A meta-analysis of prospective cohort studies. Int J Cardiol 2013;169:207-14.
7. Kendzerska T, Gershon AS, Hawker G, Tomlinson G, Leung RS. Obstructive sleep apnea and incident diabetes. A historical cohort study. Am J Respir Crit Care Med 2014;190:218-25.
8. Torres G, Sánchez-de-la-Torre M, Barbé F. Relationship between OSA and hypertension. Chest 2015;148:824-32.
9. Kerner NA, Roose SP. Obstructive sleep apnea is linked to depression and cognitive impairment: Evidence and potential mechanisms. Am J Geriatr Psychiatry 2016;24:496-508.
10. Jordan A, McEvoy RD. Gender differences in sleep apnea: Epidemiology, clinical presentation and pathogenic mechanisms. Sleep Med Rev 2003;7:377-89.
11. Ye L, Pien GW, Weaver TE. Gender differences in the clinical manifestation of obstructive sleep apnea. Sleep Med 2009;10:1075-84.
12. Izci B, Vennelle M, Liston WA, Dundas KC, Calder AA, Douglas NJ. Sleep-disordered breathing and upper airway size in pregnancy and post-partum. Eur Respir J 2006;27:321-7.
13. Facco FL, Kramer J, Ho KH, Zee PC, Grobman WA. Sleep disturbances in pregnancy. Obstet Gynecol 2010;115:77-83.
14. Liu L, Su G, Wang S, Zhu B. The prevalence of obstructive sleep apnea and its association with pregnancy-related health outcomes: A systematic review and meta-analysis. Sleep Breath 2019;23:399-412.
15. Balserak BI. Sleep disordered breathing in pregnancy. Breathe 2015;11:268-77.
16. Jordan AS, McSharry DG, Malhotra A. Adult obstructive sleep apnoea. Lancet 2014;383:736-47.
17. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med 1999;131:485-91.
18. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, *et al*. STOP questionnaire: A tool to screen patients for obstructive sleep apnea. Anesthesiology 2008;108:812-21.
19. Johns MW. A new method for measuring daytime sleepiness: The Epworth sleepiness scale. Sleep 1991;14:540-5.
20. Gross JB, Bachenberg KL, Benumof JL, Caplan RA, Connis RT, Cote CJ, *et al*. Practice guidelines for the perioperative management of patients with obstructive sleep apnea: A report by the American Society of Anesthesiologists Task Force on Perioperative Management of Patients with Obstructive Sleep Apnea. Anesthesiology 2006;104:1081–93.
21. Lockhart EM, Abdallah AB, Tuuli MG, Leighton BL. Obstructive sleep apnea in pregnancy: assessment of current screening tools. Obstet Gynecol 2015;126:93-102.
22. Dominguez JE, Grotegut CA, Cooter M, Krystal AD, Habib AS. Screening extremely obese pregnant women for obstructive sleep apnea. Am J Obstet Gynecol 2018;219:613-e1.
23. 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;155:529-36.
24. Deeks JJ, Wisniewski S, Davenport C. Guide to the contents of a Cochrane Diagnostic Test Accuracy Protocol. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. 0. The Cochrane Collaboration, 2013. Available from: methods.cochrane. org/sdt/handbook-dta-reviews. 2013. [Last accessed on 2023 Mar 21].
25. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934;26:404-13.
26. Harbord RM, Whiting P. Metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. Stata J 2009;9:211-29.
27. Deeks JJ, Altman DG. Diagnostic tests 4: Likelihood ratios. BMJ 2004;329:168-9.
28. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
29. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58:882-93.
30. Dwamena B. MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies. 2009.
31. Facco FL, Ouyang DW, Zee PC, Grobman WA. Development of a pregnancy-specific screening tool for sleep apnea. J Clin Sleep Med 2012;8:389-94.
32. Olivarez SA, Maheshwari B, McCarthy M, Zacharias N, van den Veyver I, Casturi L, *et al*. Prospective trial on obstructive sleep apnea in pregnancy and fetal heart rate monitoring. Am J Obstet Gynecol 2010;202:552-e1-7.
33. 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;22:670-8.
34. Fung AM, Wilson DL, Lappas M, Howard M, Barnes M, O'Donoghue F, *et al*. Effects of maternal obstructive sleep apnoea on fetal growth: A prospective cohort study. PloS One 2013;8:e68057.
35. Wanitcharoenkul E, Chirakalwasan N, Amnakkittikul S, Charoensri S, Saetung S, Chanprasertyothin S, *et al*. Obstructive sleep apnea and diet-controlled gestational diabetes. Sleep Med 2017;39:101-7.
36. Tantrakul V, Numthavaj P, Guilleminault C, McEvoy M, Panburana P, Khaing W, *et al*. Performance of screening questionnaires for obstructive sleep apnea during pregnancy: A systematic review and meta-analysis. Sleep Med Rev 2017;36:96-106.
37. Malhamé I, Bublitz MH, Bourjeily G. The challenge of screening for obstructive sleep apnea in pregnancy. Ann Am Thorac Soc 2019;16:1242-4.
38. Amra B, Rahmati B, Soltaninejad F, Feizi A. Screening questionnaires for obstructive sleep apnea: An updated systematic review. Oman Med J 2018;33:184.
39. Amra B, Javani M, Soltaninejad F, Penzel T, Fietze I, Schoebel C, *et al*. Comparison of Berlin questionnaire, STOP-Bang, and Epworth sleepiness scale for diagnosing obstructive sleep apnea in Persian patients. Int J Prev Med 2018;9:28.
40. Gantner D, GE JY, LI LH, Antic N, Windler S, Wong K, *et al*. Diagnostic accuracy of a questionnaire and simple home monitoring device in detecting obstructive sleep apnoea in a Chinese population at high cardiovascular risk. Respirology 2010;15:952-60.
41. 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;7:43-8.
42. Tan A, Yin JD, Tan LW, van Dam RM, Cheung YY, Lee CH.

Predicting obstructive sleep apnea using the STOP-Bang questionnaire in the general population. Sleep Med 2016;27:66-71.

43. Chung F, Yang Y, Liao P. Predictive performance of the STOP-Bang score for identifying obstructive sleep apnea in obese patients. Obes Surg 2013;23:2050-7.
44. Chiu HY, Chen PY, Chuang LP, Chen NH, Tu YK, Hsieh YJ, *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 2017;36:57-70.
45. 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:e0143697.
46. Godoy PH, dos Santos Nucera AP, de Paiva Colcher A, de-Andrade JE, Alves DD. Screening for obstructive sleep apnea in elderly: Performance of the Berlin and STOP-Bang questionnaires and the Epworth Sleepiness Scale using polysomnography as gold standard. Sleep Sci 2022;15:203.
47. Miller JN, Kupzyk KA, Zimmerman L, Pozehl B, Schulz P, Romberger D, *et al*. Comparisons of measures used to screen for obstructive sleep apnea in patients referred to a sleep clinic. Sleep Med 2018;51:15-21.
48. Mahmoud H, Sallma S, Mohammad A. Validation of the Epworth Sleepiness Scale, Berlin, STOP-Bang Questionnaires and American Society of Anesthesiologists checklist as screening tools for obstructive sleep apnea in patients with chronic obstructive pulmonary disease, asthma, and cardiovascular diseases. Chest. 2014;146:944A.
49. 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;127:3065-70.
50. Louis JM, Koch MA, Reddy UM, Silver RM, Parker CB, Facco FL, *et al*. Predictors of sleep-disordered breathing in pregnancy. Am J Obstet Gynecol 2018;218:521.e1-12.
51. Facco FL, Parker CB, Reddy UM, Silver RM, Koch MA, Louis JM, *et al*. Association between sleep-disordered breathing and hypertensive disorders of pregnancy and gestational diabetes mellitus. Obstet Gynecol 2017;129:31-41.
52. Tantrakul V, Sirijanchune P, Panburana P, Pengjam J, Suwansathit W, Boonsarngsuk V, *et al*. Screening of obstructive sleep apnea during pregnancy: Differences in predictive values of questionnaires across trimesters. J Clin Sleep Med 2015;11:157-63.
53. Siriyotha S, Tantrakul V, Plitphonganphim S, Rattanasiri S, Thakkinstian A. Prediction models of obstructive sleep apnea in pregnancy: A systematic review and meta-analysis of model performance. Diagnostics 2021;11:1097.
54. Massie F, Van Pee B, Bergmann J. Correlations between home sleep apnea tests and polysomnography outcomes do not fully reflect the diagnostic accuracy of these tests. J Clin Sleep Med 2022;18:871-6.
55. Facco FL, Lopata V, Wolsk JM, Patel S, Wisniewski SR. Can we use home sleep testing for the evaluation of sleep apnea in obese pregnant women? Sleep Disord 2019;2019:3827579.
56. O'Brien LM, Bullough AS, Shelgikar AV, Chames MC, Armitage R, Chervin RD. Validation of Watch-PAT-200 against polysomnography during pregnancy. J Clin Sleep Med 2012;8:287-94.