Performance of screening questionnaires for obstructive sleep apnea during pregnancy: A systematic review and meta-analysis

Visasiri Tantrakul a,b, Pawin Numthavaj a,*, Christian Guilleminault c, Mark McEvoy d, Panyu Panburana e, Win Khaing f, John Attia d, Ammarin Thakkinstian a

a Section for Clinical Epidemiology and Biostatistics, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
b Ramathibodi Hospital Sleep Disorder Center and Division of Pulmonary and Critical Care, Medicine Department, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
c Stanford Center for Sleep Sciences and Medicine, Stanford University, Redwood City, CA, USA
d Hunter Medical Research Institute and School of Medicine and Public Health, Faculty of Health and Medicine, The University of Newcastle, Australia
e Ramathibodi Hospital Sleep Disorder Center and Division of Pulmonary and Critical Care, Medicine Department, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
f Department of Preventive and Social Medicine, University of Medicine, Mandalay, Myanmar

This review aims to evaluate the performance of obstructive sleep apnea (OSA) screening questionnaires during pregnancy. A systematic review and meta-analysis was performed using MEDLINE Scopus, CINAHL, and the Cochrane library. A bivariate meta-analysis was applied for pooling of diagnostic parameters. Six of the total 4719 articles met the inclusion criteria. The Berlin questionnaire (BQ, N = 604) and Epworth sleepiness scale (ESS, N = 420) were the most frequently used screening tools during pregnancy. The pooled prevalence of OSA during pregnancy was 26.7% (95%CI: 16.9%, 34.4%, I² = 83.1%), BQ performance was poor to fair with pooled sensitivity and specificity of 0.66 (95%CI: 0.45, 0.83; I² = 78.65%) and 0.62 (95%CI: 0.48, 0.75; I² = 81.55%), respectively. BQ performance was heterogeneous depending on type of reference test and pregnancy. Sensitivity increased if diagnosis was based on polysomnography (0.90), and respiratory disturbance index (0.90). However, sensitivity decreased if screening was performed in early pregnancy (≥20 weeks gestation: 0.47), and high-risk pregnancy (0.44). Performance of ESS was poor with pooled sensitivity and specificity of 0.44 (95%CI: 0.33, 0.56; I² = 32.8%) and 0.62 (95%CI: 0.48, 0.75; I² = 81.55%), respectively. In conclusion, BQ and ESS showed poor performance during pregnancy, hence a new OSA screening questionnaire is needed.

Registration: PROSPERO registration CRD42015025848.
© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
in general non complicated pregnancies [15]. Moreover, prevalence of SDB increases as a pregnancy progresses from the 1st trimester (30%) to the 3rd trimester (47%) in a study with serial screening [7].

Data from meta-analyses also indicate that SDB during pregnancy is associated with adverse maternal and fetal outcomes including gestational hypertension (GHT)/pre-eclampsia, gestational diabetes (GDM) and low infant birth-weight [16,17]. Early screening for OSA in pregnancy may help in reducing these adverse outcomes. However, currently there is no standard guideline for OSA screening during pregnancy despite its potential impact. Although in-laboratory polysomnography (PSG) is the gold standard diagnostic test, long waiting periods for appointments, particularly considering the short time-window to perform the test early during the pregnancy, and the discomfort induced by the sleep test may lead to absence of investigation [18]. Thus, simple and accurate screening strategies should be investigated. Risk stratification for the probability of OSA in pregnant women will also help in prioritizing the need for further diagnostic sleep testing given the limited resources in many places throughout the world. Early diagnosis and treatment of OSA in pregnancy should be implemented given the potential benefit on pregnancy outcomes [19,20].

The Berlin and STOP-BANG questionnaires have been developed and used to identify those at high risk of OSA in non-pregnant populations with fair to good performance [21,22]. For example, the Berlin questionnaire has a sensitivity and specificity ranging from 68% to 86% and 46%–95% [21], respectively, with corresponding pooled values of 72% (95%CI: 66%–78%) and 61% (95%CI: 55%–67%) [21,23,24].

Validation studies assessing the performance of the Berlin and STOP-BANG questionnaires 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 [9–15].

Considering the urgent need for a simple and feasible screening tool for OSA during pregnancy, a systematic review and meta-analysis were performed to clarify the performance of OSA questionnaires when applied to pregnant women. This study aimed to determine the sensitivity, specificity, positive likelihood ratio (LR+), negative likelihood ratio (LR−), and area under the receiver operator curve (AUROC) of available tests and the prevalence of OSA during pregnancy from existing studies.

**Methods**

**Search strategy and study location**

Our study protocol was registered with the International prospective register of systematic reviews (PROSPERO) (registration number: CRD42015025848). In order to include all available evidence, a systematic search of the literature was performed through MEDLINE (from 1996 to January 2016), Scopus (from 1980 to January 2016), CINAHL, and Cochrane library. Search terms were built according to the research question which was modelled on the PICO principle, i.e., patient, intervention, comparator, and outcome. These included: pregnancy(MeSH), “pregnant women”, parturient, gestation, obstetric: “sleep questionnaire”, Berlin, STOP-BANG “Epworth sleepiness scale” or ESS, “Pittsburgh sleep quality index”, PSQI, screening, validation, prevalence, predictors; “sleep test”, polysomnography, PSG, Watch-PAT; and obstructive sleep apnea (MeSH) sleep apnea, obstructive(MeSH), “sleep apnea”, OSA, “sleep disordered breathing”, SDB, snoring. Details of the search strategy are described in the Appendix.

**Inclusion and exclusion criteria**

Any type of observational study (cross-sectional, cohort, or case-control) or randomized-controlled trial published in any language was included in the review if it met all the following criteria: 1) studied in women during pregnancy 2) used at least one of the OSA screening questionnaires (e.g., Berlin questionnaire, STOP-BANG questionnaire, ESS, etc.), and 3) had outcome of interest as OSA/SDB by objective sleep tests including PSG, Watch-PAT [25], or any type 3 home monitoring.

Studies were excluded if they: 1) were reviews or case reports 2) provided insufficient data for pooling despite several attempts to contact authors; or 3) were multiple publications of the same original study.
All citations were then combined and duplicates were excluded. The literature search results were evaluated independently by two reviewers (VT and PN) to locate eligible studies for inclusion. Irrelevant studies were excluded at the first step of the abstracts reviews. The full-text articles of the remaining studies were then retrieved and reviewed thoroughly to determine eligibility. Disagreement between the 2 reviewers was resolved by a third party (AT).

Data extraction

Data from the included studies were extracted using standardized data extraction forms. General characteristic of studies and subjects were extracted. For data used for pooling, cross-tabulated numbers of studied sleep questionnaires and standard tests including true positive (TP), false positive (FP), false negative (FN), and true negative (TN) were extracted from individual studies. Additionally, usage of cut-off thresholds for sleep questionnaires and the standard test were also extracted.

Studied questionnaires

The Berlin questionnaire consists of 10 self-administered questions with 3 categories regarding risk factors of OSA as follows: snoring component, daytime sleepiness, and obesity (BMI ≥ 30 kg/m²) or chronic hypertension. Being positive for 2 out of 3 categories is considered as high risk for OSA [21]. The ESS is used to assess daytime sleepiness, consisting of 8 questions regarding propensity to doze off in certain situations. Total ESS scores range from 0 to 24, daytime sleepiness, consisting of 8 questions regarding propensity to doze off in certain situations. Total ESS scores range from 0 to 24, and a score ≥10 is defined as excessive daytime sleepiness [26]. Details of all other questionnaires are shown in Table S1.

Reference test

The outcome of interest was OSA/OSDB diagnosed by performing PSG, Watch-PAT, or any type 3 home monitoring. The diagnosis of OSA/OSDB was, according to the original studies, based on an apnea-hypopnea index (AHI) or respiratory disturbance index (RDI) ≥5 [18].

Assessment of methodological quality

Methodological quality of each study was assessed independently by 2 reviewers (VT and PN) based on the QUADAS-2 (quality assessment of diagnostic accuracy studies). The QUADAS-2 tool assesses risk of bias (internal validity) and applicability (external validity) with several domains consisting of patient selection, index test, reference standard, and the flow and timing between index test and reference standard. Each domain is graded as low, high, or unclear risk of bias [27].

Statistical analysis

A bivariate meta-analysis with a random-effect model was applied for pooling diagnostic parameters (i.e., sensitivity, specificity, likelihood ratio positive/negative (LR+/LR-), and diagnostic odds ratio (DOR) using metandi and midas commands in STATA [28]). Hierarchical summary receiver operating curves (HSROC) analysis was applied to construct summary receiver operating curves (SROC) and assess threshold effects [29,30]. Pre-test probability (i.e., prevalence of OSA) and positive predictive value were pooled using the user-provided pmeta command in STATA. In addition, a post-test probability was performed for further estimation based on Bayes’ theorem and depicted visually with Fagan’s nomogram [31]. Heterogeneity was considered present if F > 25% or Q test <0.1. Exploration of sources of heterogeneity was performed by fitting the co-variables one by one on diagnostic odd ratios (DOR) incorporating both sensitivity and specificity in a meta-regression, and subgroup and sensitivity analyses were carried out accordingly [32,33]. Publication bias was assessed using Deek’s funnel plot [34]. Threshold effects from a different cut-off point in each diagnostic test were also assessed.

Results

Our extensive search yielded 4915 citations, as shown in the PRISMA flow diagram in Fig. 1 [35]. After thorough screenings, 31 studies were included in the qualitative assessment, but only 6 studies were eligible for the meta-analysis [9–14]. Characteristics of included studies are summarized in Table 1, and reasons for exclusion of studies are summarized in Table S2 [15,19,20,36–57].

Among the 6 studies, 4 studies enrolled subjects through outpatient antenatal care services (ANC) [9,11–13], whereas 1 study enrolled subjects during obstetric admission [10] and 1 study enrolled subjects both during in-patient and ANC visits [14].

Screening of subjects was performed during the 2nd or 3rd trimesters in 4 studies [9,10,12,14]; gestational age (GA) ≤20 weeks in 1 study [13]; and at all trimesters in 1 study [11]. One study serially screened subjects with questionnaires during 2nd and 3rd trimesters, but validation PSG was performed only during 3rd trimester [12].

Pregnant women were recruited from the general ANC services in 3 studies [9,12,14]. The other 2 studies included pregnant women during their ANC visit with conditions considered to be high risk for either OSA or obstetric complications, e.g., chronic hypertension, pre-gestational diabetes, GDM, obesity, or a prior history of pre-eclampsia [11,13]. One study enrolled pregnant women who were admitted to the hospital with conditions, e.g., preterm labor, GDM, pre-eclampsia, preterm premature rupture of membranes, chronic hypertension, trauma, GHT, pregestational diabetes, and urinary tract infection, and thus was considered as a high risk pregnancy study [10].

A total of 604 singleton pregnant women from 6 studies were included in the meta-analysis. Mean age ranged from 26.6 to 33.5 years with their mean gestational age ranging from 16.5 to 32.3 weeks. The prevalence of OSA (either AHI or RDI ≥5 events/hour) in overall and low risk pregnancy studies ranged from 12% to 35%, and ranged from 20% to 31.9% in high risk pregnancy studies.

Obstructive screening questionnaire characteristics

Among the 6 included studies [9–14], 5 questionnaires had been used for OSA screening, namely the Berlin questionnaire, STOP-BANG questionnaire, STOP questionnaire, ESS, and American Society of Anesthesiologist’s checklist (ASA) [21,22,26,58–60]. Two clinical prediction scores for OSA including multivariate apnea prediction index (MAP index) and the Flemons index were used in 3 studies [61,62]. However, results from the MAP index, and Flemons index were not reported and there were insufficient 2 × 2 contingency table data for pooling [9,12,14].

In terms of questionnaire performance, 6 studies used Berlin questionnaire, 3 studies used ESS, 2 studies used STOP-BANG questionnaire, 1 study used STOP questionnaire, and 1 study used ASA checklist. Given the limited data, the meta-analysis was performed only for the Berlin and ESS questionnaires. The qualitative descriptions of reported results of these questionnaires are shown in Table 2 and Table S3.

Reference tests

All 6 studies administered at least one of the OSA screening questionnaires, which were validated against reference sleep
tests. PSG, considered as the gold standard for OSA diagnosis was performed in only 2 studies [9,12] whereas either Watch-PAT (2 studies) [11,13], or ApneaLink (2 studies) were performed in the other studies [10,14]. However, only Watch-PAT has been validated against PSG for the diagnosis of OSA during pregnancy [63].

The 2 studies which performed PSG, used respiratory scoring criteria according to the American academy of sleep medicine scoring manual, 2007 [9,12]. Specifically, apnea was defined as a reduction of thermistor signal of at least 90% from baseline lasting ≥10 s, and hypopnea (alternative rule) was defined as a reduction in nasal pressure transducer signal of at least 50% from baseline which is associated with either oxygen desaturation ≥3% or arousal. Respiratory effort related arousal (RERA) was defined as a sequence of breaths lasting ≥10 s characterized by increasing respiratory effort or nasal airflow flattening leading to arousal [18].

One of the 2 studies using ApneaLink defined apneas and hypopneas as the reduction of airflow by 0–20% lasting ≥10 s, and reduction of airflow by 50% lasting ≥10 s, respectively [14]. The other ApneaLink study did not specify the criteria used [10].

Scoring of the Watch-PAT studies was based on the automatic algorithm provided by the manufacturer’s propriety software based on heart rate and finger plethysmography [11,13].

Tabulation of AHI consisted of numbers of apneas and hypopneas divided by total sleep time (hours). Additionally, respiratory disturbance index (RDI) was defined as numbers of apneas, hypopneas, and RERAs divided by total sleep time (hours) [18].

Four studies diagnosed OSA based on AHI ≥5 events/hour cut-off threshold, the other 2 studies used a RDI ≥5 events/hour as the diagnostic cut-off threshold for OSA. Additionally, result based on a RDI ≥10 events/hour cut-off was also reported in 1 study [12].

**Risk of bias assessment**

A summary of methodological quality assessment according to the QUADAS-2 tool is shown as a percentage for each domain in Figs. 2, 3 and Table S4. Details of the assessment criteria are shown in the Appendix. The agreement among independent reviewers was excellent with k-statistics of 0.96.

For the patient selection domain, a high risk of bias was documented in 4 studies: convenience sampling in all studies [9,11,12,14] and selection of subsets of subjects at either end of the SDB risk spectrum in 1 study [12]. A high risk of bias for applicability for patient selection was documented in 3 studies due to selection of high risk pregnant women [10,11,13]. This resulted in spectrum bias, i.e., selection of the subjects at either the high or low risk ends of the spectrum of the condition. Risk of bias for the index test domain was high in 1 study because the criteria for diagnosis were not pre-specified [14]. Also, the blinding of the index test interpretation was unclear in 2 studies [10,13]. Concerns regarding applicability of the index test were high in 2 studies due to differences in body mass index (BMI) cut-off threshold for obesity, and timing of measurements during pregnancy [11,14]. A cut-off threshold of BMI ≥30 kg/m² was used in 4 studies [9,10,12,13], whereas a cut-off threshold of 27.5 kg/m² was used in 1 Asian study [11], and ≥35 kg/m² in another study [14]. Pre-pregnancy BMI was used in 2 studies [11,13], early pregnancy BMI was used in 1 study [9], and pregnancy BMI was used in 3 studies [10,12,14].
In terms of reference test, a high risk of bias was documented in 4 studies due to the fact that the testing was performed either with ApneaLink or Watch-PAT instead of standard PSG [9,11,13,14]. An unclear risk of bias was present in 1 study due to unclear information on previous knowledge of the result of the test prior to study interpretation. Applicability of the reference test was rated high bias in 2 studies using ApneaLink because the criteria used for hypopnea scoring was not according to the American academy of sleep medicine sleep scoring manual [10,14].

Appropriate timing between questionnaire (index test) administration and sleep testing (reference standard) is considered to be under 2 weeks, see Appendix. Risk of bias assessment in the flow

---

**Table 1**

Characteristics of the included studies.

<table>
<thead>
<tr>
<th>Author, year/questionnaire</th>
<th>Study characteristics</th>
<th>Patient characteristics</th>
<th>Diagnosis</th>
<th>OSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tantrakul et al. 2015 [11]</td>
<td>Study design: Cross sectional study</td>
<td>Age mean (SD), y</td>
<td>Pre-pregnancy AHI 24.2 (5.3)</td>
<td>31.9%</td>
</tr>
<tr>
<td>Berlin STOP-BANG ESS</td>
<td>Patients: Singleton pregnancy</td>
<td>GA mean (SD), wk</td>
<td>Pregnancy 26.3 (5.3)</td>
<td>Median (IQR)</td>
</tr>
<tr>
<td></td>
<td>Setting: high risk ANC clinic</td>
<td>BMI mean (SD), kg/m²</td>
<td>Watch-PAT200</td>
<td>AHI overall</td>
</tr>
<tr>
<td></td>
<td>Pregnancy risk: High</td>
<td></td>
<td>AHI ≥5</td>
<td>2.4 (8.0)</td>
</tr>
<tr>
<td></td>
<td>GA: Any</td>
<td></td>
<td>Automated analysis on PAT software algorithm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Country: Thailand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lockhart et al. 2015 [14]</td>
<td>Study design: Prospective Cohort</td>
<td>Age mean (SD), y</td>
<td>Pre-pregnancy n/a</td>
<td>12%</td>
</tr>
<tr>
<td>STOP STOP-BANG ESS ASA - checklist</td>
<td>Patients: Volunteer singleton pregnancy</td>
<td>GA mean (SD), wk</td>
<td>Pregnancy median (IQR) 31 (27–36)</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Setting: OPD and IPD ANC service</td>
<td>BMI mean (SD), kg/m²</td>
<td>ApneaLink</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pregnancy risk: General GA: ≥27 weeks</td>
<td></td>
<td>AHI ≥5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Country: USA</td>
<td></td>
<td>Apnea – reduction of airflow to 0–20% lasting ≥10 s</td>
<td></td>
</tr>
<tr>
<td>Wilson et al. 2013 [12]</td>
<td>Study design: Prospective Cohort</td>
<td>Age mean (SD), y</td>
<td>Pre-pregnancy 32.2 (8.0)</td>
<td>35%</td>
</tr>
<tr>
<td>Berlin MAP Index</td>
<td>Patients: Singleton pregnancy N = 43</td>
<td>GA mean (SD), wk</td>
<td>Pregnancy 37.5 (7.9)</td>
<td>Median (IQR)/AHI</td>
</tr>
<tr>
<td></td>
<td>Setting: ANC clinic</td>
<td>BMI mean (SD), kg/m²</td>
<td>PSG</td>
<td>Without OSA:</td>
</tr>
<tr>
<td></td>
<td>Pregnancy risk: General GA: 2nd trimester with 3rd trimester follow-up Country: Australia</td>
<td></td>
<td>RDI ≥10</td>
<td>1.5 (0.6–2.7)</td>
</tr>
<tr>
<td></td>
<td>Country: Australia</td>
<td></td>
<td>Scoring based on AASM 2007 criteria, alternative hypopnea rule (defined as reduction of nasal pressure signal at least 50% lasting ≥10 s with either oxygen desaturation ≥3% or arousal)</td>
<td>With OSA:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.2 (4.9–11.2)</td>
</tr>
<tr>
<td>Fung et al. 2013 [9]</td>
<td>Study design: Prospective Cohort</td>
<td>Age mean (SD), y</td>
<td>Pre-pregnancy 26.1 (6.4)</td>
<td>34%</td>
</tr>
<tr>
<td>Berlin MAP Index</td>
<td>Patients: Convenient sample of singleton pregnancy N = 41</td>
<td>GA mean (SD), wk</td>
<td>Pregnancy n/a</td>
<td>Median (IQR)/RDI</td>
</tr>
<tr>
<td></td>
<td>Setting: ANC clinic</td>
<td>BMI mean (SD), kg/m²</td>
<td>PSG</td>
<td>Without OSA:</td>
</tr>
<tr>
<td></td>
<td>Pregnancy risk: General GA: 2nd trimester with 3rd trimester follow-up Country: Australia</td>
<td></td>
<td>RDI ≥5</td>
<td>1.4 (0.6–2.6)</td>
</tr>
<tr>
<td></td>
<td>Country: Australia</td>
<td></td>
<td>Scoring based on AASM 2007 criteria, alternative hypopnea (defined as reduction of nasal pressure signal at least 50% lasting ≥10 s with either oxygen desaturation ≥3% or arousal)</td>
<td>With OSA:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.2 (4.9–11.7)</td>
</tr>
<tr>
<td>Facco et al. 2012 [13]</td>
<td>Study design: Prospective Cohort</td>
<td>Age mean (SD), y</td>
<td>Pre-pregnancy 31.9 (9.1)</td>
<td>28%</td>
</tr>
<tr>
<td>Berlin ESS</td>
<td>Patients: Singleton pregnancy N = 100</td>
<td>GA mean (SD), wk</td>
<td>Pregnancy n/a</td>
<td>Median (IQR)/AHI</td>
</tr>
<tr>
<td></td>
<td>Setting: ANC Clinic</td>
<td>BMI mean (SD), kg/m²</td>
<td>Watch-PAT100</td>
<td>Overall</td>
</tr>
<tr>
<td></td>
<td>Pregnancy Risk: High risk for OSA GA: 6–20 week Country: USA</td>
<td></td>
<td>AHI ≥5</td>
<td>1.5 (0.5–6.0)</td>
</tr>
<tr>
<td></td>
<td>Country: USA</td>
<td></td>
<td>Automated analysis on PAT software algorithm</td>
<td></td>
</tr>
<tr>
<td>Olivarez et al. 2010 [10]</td>
<td>Study design: Prospective Cohort</td>
<td>Age mean (SD), y</td>
<td>Pre-pregnancy n/a</td>
<td>20%</td>
</tr>
<tr>
<td>Berlin</td>
<td>Patients: Singleton pregnancy N = 100</td>
<td>GA mean (SD), wk</td>
<td>Pregnancy 27.5 (7.2)</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Setting: antepartum obstetric admission</td>
<td>BMI mean (SD), kg/m²</td>
<td>ApneaLink</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pregnancy risk: High GA: ≥26 weeks Country: USA</td>
<td></td>
<td>AHI ≥5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Apnea and hypopnea definitions were not specified</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: AASM – American academy of sleep medicine; AHI – apnea-hypopnea index; ANC – antenatal care; ASA checklist – American society of anesthesiologist’s checklist; BMI – body mass index; ESS – Epworth sleepiness scale; GA – gestational age; IQR – interquartile range; IPD – in-patient department; MAP index – multivariate apnea prediction index; n/a – not applicable; OPD – out-patient department; OSA – obstructive sleep apnea; RDI – respiratory disturbance index; SD – standard deviation.
and timing domain were high in 4 studies. The reasons were inappropriate interval between index test and reference standard (>2 weeks) in 2 studies [9,12]; not all subjects underwent the objective sleep testing in 2 studies [12,14]; not all subjects underwent the same type of recording in 1 study [9]; and not all subjects were included in the analysis in 3 studies [12–14]. Verification bias occurred if not all of the included subjects were subjected to an objective nocturnal sleep test [64–66].

High risk of bias and applicability occurred because of differences in methodology, especially the diagnostic measure of OSA and pregnancy risks. However, these factors were also included in the meta-regression to clarify the potential effect on questionnaire performance.

**Performance of Berlin questionnaire**

Among 6 included studies (n = 604), the overall pooled prevalence of OSA was 26.7% (95%CI: 16.9%, 34.4%). When categorized into low and high risk pregnancy, the pooled prevalence was approximately 23.5% (95%CI: 12.8%, 34.2%, respectively) and 29.6% (95%CI: 22.8%, 36.4%, respectively). The specificity and sensitivity of the Berlin questionnaires varied highly across studies, i.e., ranging from 0.35 to 0.93 and 0.32 to 0.88, respectively (see Fig. 4). The pooled specificity and sensitivity were 0.66 (95%CI: 0.45, 0.83; \( I^2 = 78.65 \) and Q test \( p < 0.001 \)) and 0.62 (95%CI: 0.48–0.75; \( I^2 = 81.55 \) and Q test \( p < 0.001 \)), respectively.

The \( LR^+ \) and \( LR^- \) were heterogeneous ranging from 0.97 to 4.62 and 0.14 to 1.02, respectively. The pooled \( LR^+ \) and \( LR^- \) were 1.75 (95%CI: 1.28–2.38; \( I^2 = 28.85 \) and Q test \( p = 0.02 \)) and 0.54 (95%CI: 0.33, 0.90; \( I^2 = 69.91 \) and Q test \( p = 0.01 \)), respectively (see Figure S1).

The DORs varied greatly across studies and ranged from 0.95 to 14.0, leading to a pooled DOR of 3.23 (95%CI: 1.54–6.77; \( I^2 = 96.49 \) and Q test \( p < 0.001)\), see Figure S2). This suggests that the odds of having OSA are 3 times higher in those with a positive Berlin questionnaire than those with a negative Berlin questionnaire. In addition, the SROC was estimated and graphed, which yielded an area under the curve (AUC) of 0.68 (95%CI: 0.58–0.78), see Fig. 5.

Finally, the post-test probability was estimated using Fagan’s nomogram (see Fig. 6), which yielded a post-test probability of 38% given a pre-test probability of 26% and \( LR^+ \) of 1.75. This suggests that the overall probability of OSA occurrence is 38% if a pregnant woman has a positive Berlin questionnaire result.

**Assessment of heterogeneity**

Across studies, the Berlin questionnaire showed extremely high heterogeneity for the DOR with an \( I^2 \) of 91.0% (95%CI: 83%, 100%). The estimated threshold effect was \(-0.02\) (95%CI: \(-0.70, 0.66\)), which suggests that the threshold cutoff of the Berlin score was only minimally influenced by heterogeneity.

Other sources of heterogeneity were explored with meta-regression, as shown in Figure S3. Results showed that...
differences in the performances of the Berlin questionnaire (i.e., sensitivity and specificity) were accounted for by the use of PSG (yes/no), RDI vs AHI for OSA diagnosis, and risk of pregnancy (low vs high). Accordingly, subgroup analyses were performed, which revealed improvements in sensitivity among those studies which used PSG and RDI as the reference tests, with a pooled sensitivity of 0.90 (95%CI:0.78–1.00), 0.90 (95%CI:0.78–1.00), respectively (See Table 3). In addition, the sensitivity was also higher in pregnant women in general ANC clinic and when the GA was above 20 weeks, with pooled sensitivities of 0.81 (95%CI: 0.71, 0.91) and 0.74 (95%CI: 0.56–0.93), respectively.

Publication bias assessment

Publication bias was assessed based on the DOR using Deek’s funnel plot and test, see Figure S4. The funnel is symmetrical and this corresponded with Deek’s test (coefficient = 11.47, SE = 19.01, p-value of 0.60) suggesting that a study size effect (publication bias) was unlikely.

ESS performance

Meta-analysis of ESS performance was analyzed from only 3 studies (N = 420), see Table 2. The sensitivity and specificity of ESS across different studies ranged from 0.36 to 0.58 and 0.58 to 0.76, respectively. The pooled sensitivity was 0.44 (95%CI: 0.33, 0.56; $I^2 = 32.8\%$ and Q test $p = 0.23$), with a pooled specificity of 0.62 (95%CI: 0.57, 0.67; $I^2 = 76.9\%$ and Q test $p = 0.013$). The pooled LR+ was 1.29 (95%CI: 0.973, 1.71; $I^2 = 0\%$ Q test $p = 0.56$), and the pooled LR– was 0.86 (95%CI: 0.70, 1.07; $I^2 = 0\%$ and Q test $p = 0.55$). The pooled DOR was 1.53 (95%CI: 0.89, 2.60; $I^2 = 0\%$ and Q test $p = 0.56$). These results indicate a low heterogeneity.

Discussion

We performed a systematic review and meta-analysis of how screening questionnaires for OSA performed during pregnancy. Our findings indicated that the Berlin questionnaire has poor to fair performance with a pooled sensitivity, specificity, LR+, and LR– of 0.66, 0.62, 1.75, and 0.68, respectively. Using the Berlin questionnaire could change the pre-test probability of being recognized with OSA from 26% to 38%. The ESS had a very poor predictive value. A Fagan plot obtained using the Berlin questionnaire to detect OSA during pregnancy showed that it did not help clinical decision making either on treatment threshold (confirmation) or test threshold (exclusion). There was a high degree of heterogeneity in the usage of the Berlin questionnaire, but only a small threshold effect from a hierarchical SROC curve analysis. We attempted to explore the source/s of the noted heterogeneity using meta-regression analysis. The results showed a higher sensitivity of the Berlin questionnaire. This analysis indicated improvement in the sensitivity of the Berlin questionnaire if the reference test was a PSG, and looking at RDI. The sensitivity was also higher in the general pregnancy population compared to high risk pregnancy, and if it was performed after 20 weeks of gestation.

The explanation for the poor discriminative values of the conventional OSA screening questionnaires during pregnancy might be...
related to the facts that 1) both OSA and pregnancy lead to similar sleep complaints; 2) that there is a continuous change in symptomatology and the severity of the sleep-disordered-breathing with progression of pregnancy; 3) that the standard and threshold to diagnose SDB during pregnancy has not been defined; 4) finally, that the optimal timing of the questionnaire administration has not been standardized.

The Berlin questionnaire and ESS were the most frequently used screening tools during pregnancy. There are several other available OSA screening questionnaires such as the Nordic sleep questionnaire, and sleep disorder questionnaire, but they have not been tested or validated on pregnant women. The majority of pregnant women experienced poor sleep quality, insufficient nocturnal sleep and daytime sleepiness. Sleep disruption occurs from early pregnancy and increases, particularly during the third trimester. Our result showed higher sensitivity of the Berlin questionnaire after 20 weeks gestational age. Our findings indicate that more sensitive and specific sleep questionnaires in pregnancy are needed.

The ESS was of very little help in suggesting OSA during pregnancy. Excessive daytime sleepiness (EDS) is highly prevalent (31.0%–45.5%) even in early pregnancy and increased significantly as pregnancy progressed, probably due to the increase in progesterone and sleep disruption from pregnancy per se. EDS (ESS>10) is not associated with GDM and GHT. However, an ESS score >16 was associated with GDM. There is a correlation between an increasing ESS score and the other symptoms of SDB such as loud snoring, and apnea; such findings raise the question of the optimal cut-off threshold for ESS in pregnant women. In addition, ESS had been found to be poorly predictive of OSA in non-pregnant population, so it is not surprising that it does not predict OSA in pregnancy. Given that ESS is a better predictor of depression compared to other health outcomes, it might be worth looking at this association in pregnancy cohorts.

Worsening of OSA severity and symptoms occurred as pregnancy progressed and the prevalence of OSA increased from early pregnancy to 3rd trimester. During the 3rd trimester, the incidence of new-onset SDB was reported at 20%, in association with twin pregnancy and new-onset of self-reported frequent snoring. Pregnancy-onset snoring was also reported to be associated with pre-eclampsia. This finding highlights the potential for SDB to impact adversely on maternal and fetal outcomes during pregnancy. However, there is still a lack of evidence for SDB during pregnancy, particularly with respect to the role of screening for SDB and potential benefits of treatment, which highlights the urgent need for more research in this area.

Recently, the STOP-BANG questionnaire has become more popular, given its simplicity and greater consistency across different AHI severity, when compared to the Berlin questionnaire. During the 3rd trimester, the incidence of new-onset SDB was reported at 20%, in association with twin pregnancy and new-onset of self-reported frequent snoring. Pregnancy-onset snoring was also reported to be associated with pre-eclampsia. This finding highlights the potential for SDB to impact adversely on maternal and fetal outcomes during pregnancy. However, there is still a lack of evidence for SDB during pregnancy, particularly with respect to the role of screening for SDB and potential benefits of treatment, which highlights the urgent need for more research in this area.

Please cite this article in press as: Tantrakul V, et al., Performance of screening questionnaires for obstructive sleep apnea during pregnancy: A systematic review and meta-analysis, Sleep Medicine Reviews (2016), http://dx.doi.org/10.1016/j.smrv.2016.11.003
screening tools may not perform well in pregnancy. However, this article was not a systematic review or meta-analysis [78].

Interestingly, from our review the performance of OSA screening questionnaires from all the included studies was compared against the diagnosis of OSA with either AHI or RDI ≥ 5 events/hour. There were not enough data available on moderate to severe OSA in pregnancy. The majority of SDB during pregnancy were mild as reported in a high risk cohort [7]. However, even mild SDB during pregnancy might be important as adverse outcomes have been reported [9,79].

In contrast to the non-pregnant population, OSA screening questionnaires were mostly performed based on the diagnosis of moderate to severe severity (AHI ≥ 15 events/hour) [23,24], given the unclear impact of mild OSA and its treatment on the cardiovascular outcomes [80,81].

There were some limitations in our study. First, there was only a small number of available studies. Second, presence of verification and spectrum biases in some of the included studies might have caused overestimation of questionnaire performance, particularly for sensitivity [64–66]. Third, there was a high degree of heterogeneity among studies regarding recruited population, and diagnosis measures, therefore caution must be taken in generalization of the results.

In conclusion, with the complexity and the dynamic change of OSA and pregnancy, screening for this condition during pregnancy is complicated. Current OSA screening questionnaires perform poorly during pregnancy. There is a need for tools that take into account the changes over time of the symptoms of OSA, and pregnancy. New screening tools and strategies specific to pregnancy that enable us to serially screen and monitor for OSA throughout pregnancy should be developed.

**Practice points**

- Diagnosis of OSA during pregnancy is challenging despite its adverse effect on maternal and fetal outcomes.
- Early diagnosis and treatment of OSA during pregnancy may have potential benefits on pregnancy outcomes.
- Current conventional OSA screening questionnaires perform poorly in pregnancy. And there is no standard guideline for screening and treatment of OSA during pregnancy.

**Research agenda**

- There is an urgent need for a simple and accurate screening tool for OSA during pregnancy.
- As pregnancy progresses, there are dynamic changes of OSA and pregnancy. Hence a specific screening tool that can successively screen for OSA throughout pregnancy should be developed and validated.

---

**Table 3**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Conditions</th>
<th>No. of studies</th>
<th>$I^2$</th>
<th>Sensitivity (95%CI) p-value</th>
<th>Specificity (95%CI) p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies</td>
<td></td>
<td>6</td>
<td>96.5</td>
<td>0.67 (0.45, 0.83) &lt;0.001</td>
<td>0.62 (0.48, 0.74) 0.35</td>
</tr>
<tr>
<td>Risk of pregnancy</td>
<td>High</td>
<td>3</td>
<td>83%</td>
<td>0.44 (0.32, 0.56)</td>
<td>0.74 (0.62, 0.85)</td>
</tr>
<tr>
<td></td>
<td>General</td>
<td>3</td>
<td>81%</td>
<td>0.81 (0.71, 0.91)</td>
<td>0.51 (0.35, 0.67)</td>
</tr>
<tr>
<td>Reference sleep test</td>
<td>PSG</td>
<td>2</td>
<td>79%</td>
<td>0.90 (0.78, 1.00) 0.03</td>
<td>0.42 (0.23, 0.60) 0.03</td>
</tr>
<tr>
<td></td>
<td>Non-PSG</td>
<td>4</td>
<td></td>
<td>0.52 (0.37, 0.67)</td>
<td>0.70 (0.60, 0.79)</td>
</tr>
<tr>
<td>Respiratory index</td>
<td>AHI</td>
<td>4</td>
<td>79%</td>
<td>0.52 (0.37, 0.67)</td>
<td>0.70 (0.60, 0.79)</td>
</tr>
<tr>
<td></td>
<td>RDI</td>
<td>2</td>
<td></td>
<td>0.90 (0.78, 1.00) 0.03</td>
<td>0.42 (0.23, 0.60)</td>
</tr>
<tr>
<td>Early pregnancy</td>
<td>Yes</td>
<td>2</td>
<td>58%</td>
<td>0.47 (0.17, 0.78) 0.15</td>
<td>0.78 (0.65, 0.90) 0.38</td>
</tr>
<tr>
<td>(GA &lt;20 weeks)</td>
<td>No</td>
<td>4</td>
<td></td>
<td>0.74 (0.56, 0.93)</td>
<td>0.54 (0.42, 0.66)</td>
</tr>
</tbody>
</table>

Abbreviations: AHI = apnea-hypopnea index; CI = confidence interval; GA = gestational age; OSA = obstructive sleep apnea; PSG = polysomnography; RDI = respiratory disturbance index.
Conflicts of interest

None of the authors had a conflict of interest or financial support.

Authors’ contribution

Study concept and design: VT, PN, CG, WK, AT. Study selection and risk of bias assessment: VT, PN, WK. Data extraction: VT, PN. Data analysis: VT, AT. Interpretation of data: VT, PN, CG, MM, PP, WK, JA, AT. Drafting the manuscript: VT, PN, CG, MM, PP, JA, AT. Critical revision of the manuscript for important intellectual content: VT, PN, CG, MM, PP, JA, AT. Final approval of the version to be published: all authors read and approved the final manuscript.

Acknowledgments

This study is a part of Dr. Visasiri Tantrakul’s training in Ph.D. program for Clinical Epidemiology, Faculty of Medicine, Ramathibodi Hospital and Faculty of Graduate Studies, Mahidol University, Bangkok, Thailand. All authors declare that no funding source or any sponsor was involved in this study and manuscript preparation for publication. We would like to thank Mr. Stephen Pinder for checking and improving the English of our manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.smrv.2016.11.003.

References


* The most important references are denoted by an asterisk.


