Sleep and Breathing (2019) 23:969–977
https://doi.org/10.1007/s11325-018-1748-x

EPIDEMIOLOGY • ORIGINAL ARTICLE

A validation study on three screening questionnaires for obstructive sleep apnea in a Korean community sample

Hong Jun Jeon[1] · Young Rong Bang[2] · In-Young Yoon[3,4]

Received: 1 July 2018 / Revised: 23 October 2018 / Accepted: 30 October 2018 / Published online: 17 November 2018

Abstract

Purpose Obstructive sleep apnea (OSA) is highly prevalent and causes serious cardiovascular complications. Several screening questionnaires for OSA have been introduced, but only few validation studies have been conducted in general population. The aim of the present study was to assess the diagnostic value of three OSA screening questionnaires (Berlin Questionnaire, BQ; STOP-Bang Questionnaire, STOP-B; Four-Variable Screening Tool, Four-V) in a Korean community sample.

Methods A total of 1148 community-dwelling participants completed the BQ, STOP-B, and Four-V. An overnight in-laboratory polysomnography (PSG) was conducted in randomly selected 116 participants. Sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, and area under the curve (AUC) were calculated.

Results The Four-V with cutoff ≥ 8 showed high sensitivity for overall OSA (69.4%), and the Four-V with cutoff ≥ 9 showed high specificity for both overall OSA (81.5%) and moderate to severe OSA (69.0%). On the other hand, the STOP-B showed acceptable sensitivity and specificity for both overall OSA (61.3 and 79.6%, respectively) and moderate to severe OSA (72.4 and 67.8%, respectively). The STOP-Bang also showed the largest area under the receiver-operator characteristic curve for both overall OSA (0.752) and moderate to severe OSA (0.750). The BQ showed the lowest performance in predicting OSA.

Conclusions Among the three questionnaires, the STOP-B was revealed as the most useful screening tool for OSA in terms of sensitivity, specificity, and area under the receiver-operator characteristic curve in the population of South Korea.

Keywords Berlin questionnaire · Four-variable screening tool · Obstructive sleep apnea · Screening · STOP-Bang questionnaire

Introduction

Untreated OSA is associated with serious medical complications, including hypertension, glucose intolerance, myocardial infarction, and cerebrovascular diseases [1–6], but only a small portion of moderate to severe OSA patients are diagnosed and receive treatment [7]. In-laboratory overnight polysomnography (PSG) has been regarded as a gold standard diagnostic tool [8], but it is expensive, time-consuming, and burdensome. Therefore, a screening tool for OSA is necessary in order to screen individuals at high risks of OSA in general population. Screening the subjects who are in urgent need for PSG and providing them treatment in time would prevent further complications and decrease social costs. The screening tool should be easy to perform, inexpensive, and reliable to have adequate predictive performance.

Several screening questionnaires for OSA have been introduced in the last two decades, and the reported screening methods have mostly been validated in clinical samples [9–14]; therefore, it is possible that sensitivity of the studied screening tools has been overestimated because of the high prevalence of OSA. There exist only a few reports on community-dwelling populations; hence, only limited results are available for interpretation since in-home PSG has been used or only middle-aged or older subjects were included in the study [15, 16]. Besides, validation studies for OSA screening questionnaires have been reported mostly in Western countries.

✉ In-Young Yoon
iyoon@snu.ac.kr

[1] Department of Psychiatry, Konkuk University Medical Center, Seoul, South Korea

[2] Department of Psychiatry, Dong-A University Hospital, Busan, South Korea

[3] Department of Psychiatry, Seoul National University Bundang Hospital, 463-707, 166 Gumi-ro, Bundang-gu, Seongnam-si, Gyeonggi-do, South Korea

[4] Department of Psychiatry, School of Medicine, Seoul National University, Seoul, South Korea

In the present study, we aimed to compare the validity of three widely used screening questionnaires for OSA, namely, the Berlin Questionnaire (BQ), STOP-Bang (STOP-B), and Four-Variable Screening Tool (Four-V), in community-dwelling Korean adults.

Material and methods

Subjects

This study is a prospective study and carried out at Seoul National University Bundang Hospital. Community-dwelling adults aged above 20 years were enrolled as the representative sample via advertisements from May 2015 to July 2015 in the southern area of Gyeonggi province in South Korea. The sampling was conducted by a stratified random sampling method taking into account age and sex groups based on the 2014 Korean census. Exclusion criteria were as follows: (1) acute or unstable medical or neurological conditions; (2) major psychiatric disorders; (3) substance abuse; (4) pregnancy; and (5) severe cognitive impairment. The subjects who did not meet the exclusion criteria completed the following three screening questionnaires for OSA: the BQ, STOP-B, and Four-V. The three questionnaires were clipped together in a randomized order to avoid bias that comes from the order of questionnaires. Six types of questionnaire formations were made through this process and distributed to subjects by the block randomization method. A total of 1148 subjects completed the three questionnaires. To test the validity of the three screening questionnaires, subjects who agreed to undergo PSG were randomly selected by considering age, sex, and risk group classification of each screening questionnaire. Finally, an overnight in-laboratory PSG was performed on 116 subjects. The process of the subject enrollment is summarized in Fig. 1. This study was approved by the Institutional Review Board of the Seoul National University Bundang Hospital, and written informed consent was obtained from all the participants.

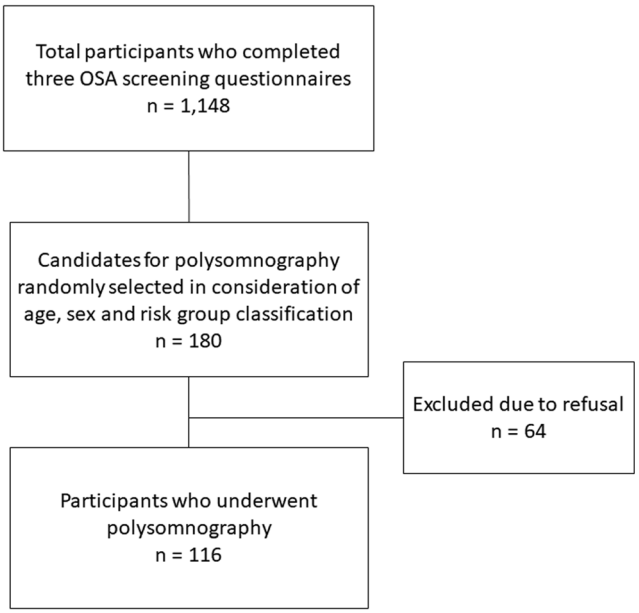

Fig. 1 Flow chart showing the subject enrollment. OSA, obstructive sleep apnea

Instruments

Berlin questionnaire

The Berlin questionnaire (BQ) is one of the most widely used screening tools for OSA. It consists of 11 items organized into three categories. The first category includes five questions on snoring and witnessed apnea, the second category includes three questions on daytime fatigue and sleepiness, and the third category includes one item on the history of hypertension. The first and second categories require the total score ≥ 2 to be positive. The third category requires a history of hypertension or a body mass index (BMI) ≥ 30 kg/m^2 to be positive. Subjects who had two or more positive categories were classified into the high-risk group. A validated Korean version of the BQ was used in this study [17], with a modified BMI cutoff value of 25 kg/m^2 compared to the cutoff value of 30 kg/m^2 in the original version according to the proposal of WHO Western Pacific Regional Office defining moderate obesity in Asian [18].

STOP-Bang questionnaire

The STOP questionnaire was developed as a screening tool for OSA in preoperative surgical patients. It consists of four items on snoring, tiredness, observed apnea, and blood pressure. A total score of two or more positive responses is considered as a high risk for OSA. The STOP-Bang (STOP-B) includes the STOP questions and incorporates the following four more questions: BMI ≥ 35 kg/m^2, age ≥ 50, neck circumference ≥ 40 cm, and male gender. The STOP-B has been validated both in clinical and non-clinical populations [19]. A total score of three or more positive responses is considered as a high risk. In this study, a BMI cutoff value of 30 kg/m^2 was used instead of 35 kg/m^2 based on the definition of severe obesity in Asian [18].

Four-variable screening tool

The Four-variable screening tool (Four-V) was developed to identify moderate to severe OSA [20]. It includes four questions, and each question assesses gender, blood pressure, BMI, and snoring. Males were scored with 4 and females with 0. Snoring almost every day or often was assigned 4, and snoring sometimes, almost never, or unknown was assigned

0. BMI was scored from 1 to 6 based on the following categorization: < 21.0, 21.0–22.9, 23.0–24.9, 25.0–26.9, 27–29.9, ≥ 30. Blood pressure (BP) was categorized as systolic BP (SBP) < 140 or diastolic BP (DBP) < 90, SBP 140–159 or DBP 90–99, SBP 160–179 or DBP 100–109, SBP ≥ 180 or DBP ≥ 110 and scored from 1 to 4. Originally, cutoff values from 9 to 14 were recommended by Takegami et al. However, we used values of 8 and 9 as indicating a high risk for OSA in the present study.

Polysomnography

An overnight in-laboratory overnight polysomnography (PSG) (Embla™ N7000, Embla, Reykjavik, Iceland) with standard electrodes and sensors was performed in 116 subjects. Electroencephalography electrodes were applied at O1/A2, O2/A1, C4/A1, and C3/A2. Two electrooculography electrodes were applied at the sides of both the eyes for recording vertical and horizontal eye movements. Electromyography electrodes were applied to the submentalis muscles and both anterior tibialis muscles. Strain gauges were applied to the chest and abdomen to record respiratory movements. Nasal pressure transducers were used to record airflow, and pulse oximeters were applied on the index finger to measure the arterial oxygen saturation. The recordings were scored according to the American Academy of Sleep Medicine guidelines [21]. Apnea was defined as an episode of complete air flow cessation for at least 10 s. Hypopnea was defined as 50% or more reduction of airflow for at least 10 s, moderate airflow reduction for at least 10 s accompanied by electroencephalographic arousal, or oxygen desaturation (≥ 4%) [22]. Subjects with apnea-hypopnea index (AHI) ≥ 5 were considered to have mild OSA, and those with AHI ≥ 15 were considered to have moderate to severe OSA [23].

Statistical analysis

The characteristics of the subjects are presented as means ± standard deviations for continuous variables and numbers and percentages for categorical variables. Comparisons between groups were conducted by the χ^2 test or the Student's t test as appropriate using SPSS version 22 (SPSS Inc., Chicago, IL, USA). The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio, and negative likelihood ratio were calculated for each questionnaire. The diagnostic odds ratio (DOR) was also calculated. The DOR is the ratio of the odds of the test being positive in disease relative to the odds of positive results in the non-disease, and higher DOR indicates better discriminatory test performance. Comparison between ROC curves were performed using MedCalc Statistical Software version 18.10 (MedCalc Software bvba, Ostend, Belgium). The significance level was set at a two-tailed value of $P < 0.05$.

Results

Descriptive characteristics of the studied subjects are shown in Table 1. Of the 1148 subjects, 49.3% were male with a mean age of 47.4 ± 15.9 years. The mean BMI of the subjects was 23.3 ± 3.0 kg/m^2, and the mean neck circumference was 34.6 ± 3.8 cm. The mean systolic and diastolic BP was 112.9 ± 15.9 and 79.8 ± 12.0 mmHg, respectively. Among the subjects, 29.7, 26.1, 30.8, 41.6, and 31.0% were classified as a high-risk group by the BQ, STOP, STOP-B, Four-V with cutoff ≥ 8, and Four-V with cutoff ≥ 9, respectively. In spite of different constructs of questions between questionnaires, each high-risk group determined based on the BQ, STOP, STOP-B, and Four-V showed significantly older age, a larger portion of the male gender, higher BMI, higher neck circumference, and higher BP than the paired low-risk group. However, the BMI difference between groups classified by the Four-V with cutoff ≥ 9 was not statistically significant.

Table 2 shows the demographic and clinical characteristic of the subjects who underwent PSG. Of the 116 subjects, 62 subjects turned out to have OSA and 54 subjects were normal. The high-risk subjects classified by the BQ, STOP, STOP-Bang, Four-V with cutoff ≥ 8, or Four-V with cutoff ≥ 9 accounted for 50.0, 56.9, 43.1, 51.7, and 37.9%, respectively. Subjects in the OSA group were older compared to the subjects in the non-OSA group, and the male percentage was higher in the OSA group. In addition, subjects in the OSA group had higher BMI, neck circumference, systolic/diastolic BP, AHI, supine AHI, and the lowest O_2 saturation during sleep. The OSA group had a higher percentage of high-risk subjects classified by any three questionnaires than the non-OSA group. There were no significant differences between the two groups in terms of the Pittsburgh sleep quality index (PSQI), insomnia severity index (ISI), and Beck depression inventory (BDI).

Predictive parameters of the studied questionnaires for detecting overall OSA are shown in Table 3(A). The Four-V with cutoff ≥ 8 (69.4%) showed better sensitivity than the BQ (61.3%), STOP (58.1%), STOP-B (61.3%), and Four-V with cutoff ≥ 9. However, the specificity of the Four-V with cutoff ≥ 9 (81.5%) was the highest, followed by the STOP-B (79.6%), STOP (74.1%), Four-V with cutoff ≥ 8 (68.5%), and BQ (63.0%). For detecting moderate to severe OSA (Table 3(B)), the STOP-B showed the highest sensitivity (72.4%) while the Four-V with cutoff ≥ 9 showed the highest specificity (69.0%). ROC curves for the three questionnaires are shown in Fig. 2. The STOP-B had the largest AUC (0.752) for overall OSA and moderate to severe OSA (0.750) among the studied screening questionnaires, even though a statistical significance was found only between the STOP-B and the BQ for detecting moderate to severe OSA ($P = 0.0013$).

Table 1 Demographic data of the studied subjects

	BQ		STOP		STOP-B		Four-V (cutoff ≥ 8)		Four-V (cutoff ≥ 9)		Total
	High risk	Low risk	High risk	Low risk	High risk	Low risk	High risk	Low risk	High risk	Low risk	
N (%)	341 (29.7)	807 (70.3)	300 (26.1)	848 (73.9)	354 (30.8)	794 (69.2)	477 (41.6)	671 (59.4)	356 (31.0)	792 (69.0)	1148
Gender, male (%)	221 (64.8)**	345 (42.8)	209 (69.7)**	357 (42.1)	287 (81.1)**	279 (35.1)	432 (90.6)**	134 (20.0)	326 (91.6)**	240 (30.3)	566 (49.3)
Age (yr)	48.8 ± 15.1*	46.8 ± 16.2	51.8 ± 15.6**	45.9 ± 15.7	54.6 ± 15.6**	44.2 ± 15.0	48.9 ± 15.1*	46.4 ± 16.4	48.7 ± 14.7*	46.8 ± 16.4	47.4 ± 15.9
BMI (kg/m^2)	25.0 ± 3.0**	22.5 ± 2.6	24.8 ± 3.1**	22.7 ± 2.8	24.9 ± 3.0**	22.5 ± 2.7	25.3 ± 2.5**	21.8 ± 2.4	25.8 ± 2.6	22.1 ± 2.4	23.3 ± 3.0
Neck circumference (cm)	36.5 ± 3.6**	33.8 ± 3.6	36.6 ± 3.6**	33.9 ± 3.6	37.4 ± 3.3**	33.3 ± 3.3	37.7 ± 2.7**	32.3 ± 2.7	38.1 ± 2.7**	33.0 ± 3.1	34.6 ± 3.8
SBP (mmHg)	127.5 ± 16.1**	120.9 ± 15.4	129.4 ± 16.6**	120.6 ± 15.0	130.7 ± 16.2**	119.4 ± 14.4	130.4 ± 15.6**	117.5 ± 13.7	132.9 ± 15.8*	118.3 ± 13.7	112.9 ± 15.9
DBP (mmHg)	83.5 ± 11.8**	78.3 ± 11.8	84.1 ± 12.7**	78.4 ± 11.4	84.7 ± 12.5**	77.7 ± 11.2	85.0 ± 12.2**	76.2 ± 10.5	86.7 ± 12.3*	76.8 ± 10.6	79.8 ± 12.0

*$P < 0.05$ vs. low risk

**$P < 0.001$ vs. low risk

BQ, Berlin questionnaire; STOP, STOP questionnaire; STOP-B, STOP-Bang questionnaire; Four-V, Four-variable screening tool; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure

Table 2 Demographic and clinical data of the subjects who underwent polysomnography

	OSA group (AHI ≥ 5)	Non-OSA group (AHI < 5)	Total	*P* value
N (%)	62 (53.4)	54 (46.6)	116	
Gender, male (%)	37 (59.7)	21 (38.9)	58 (50.0)	0.040
Age (yr)	55.4 ± 16.5	44.0 ± 14.0	50.1 ± 16.4	< 0.001
BMI (kg/m^2)	24.8 ± 2.7	22.8 ± 2.5	23.9 ± 16.4	0.001
Neck circumference (cm)	36.0 ± 3.5	33.4 ± 2.8	34.8 ± 3.4	0.001
SBP (mmHg)	127.5 ± 15.7	120.2 ± 16.7	124.1 ± 16.5	0.018
DBP (mmHg)	81.9 ± 12.5	77.4 ± 10.4	79.8 ± 11.7	0.039
AHI (event/h)	16.5 ± 11.1	1.5 ± 1.3	9.5 ± 11.1	< 0.001
Supine AHI (event/h)	23.6 ± 18.1	2.1 ± 2.1	13.6 ± 17.0	< 0.001
Lowest SpO_2 (%)	84.6 ± 4.7	90.9 ± 3.2	87.6 ± 5.2	< 0.001
High risk (BQ)	38 (61.3)	20 (37.0)	58 (50.0)	0.008
High risk (STOP)	36 (58.1)	14 (25.9)	66 (56.9)	< 0.001
High risk (STOP-Bang)	38 (61.3)	12 (22.2)	50 (43.1)	< 0.001
High risk (Four-V, cutoff = 8)	43 (69.4)	17 (31.5)	60 (51.7)	< 0.001
High risk (Four-V, cutoff = 9)	34 (54.8)	10 (18.5)	44 (37.9)	< 0.001
PSQI (≥ 5, %)	6.3 ± 2.5 (69.2)	5.3 ± 2.3 (60.5)	5.9 ± 2.4 (65.3)	0.430
ISI (≥ 8, %)	6.8 ± 5.3 (33.9)	6.1 ± 5.2 (33.3)	6.5 ± 5.2 (33.6)	0.490
ESS (> 10, %)	7.5 ± 4.3 (17.7)	8.3 ± 4.9 (29.6)	7.9 ± 4.6 (23.3)	0.357
BDI (≥ 10, %)	8.4 ± 7.0 (40.3)	8.7 ± 7.2 (38.9)	8.5 ± 7.1 (39.7)	0.853

OSA, obstructive sleep apnea; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; AHI, apnea-hypopnea index; SpO_2, oxygen saturation; BQ, Berlin questionnaire; STOP, STOP questionnaire; STOP-B, STOP-Bang questionnaire; Four-V, Four-variable screening tool; PSQI, Pittsburgh sleep quality index; ISI, insomnia severity index; ESS, Epworth sleepiness scale; BDI, Beck depression inventory

Discussion

In this prospective study, we performed in-laboratory PSG in 116 subjects out of 1148 community-dwelling subjects, who completed the BQ, STOP-Bang, and Four-V, and evaluated the diagnostic performances of the three screening questionnaires for OSA. In terms of sensitivity, the Four-V with cutoff ≥ 9 showed the highest value in detecting OSA, but the STOP-Bang showed the highest result in detecting moderate to severe OSA. The STOP-B showed better performance in terms of specificity diagnosing both OSA and moderate to severe OSA than the BQ and the Four-V with cutoff ≥ 8. The Four-V with cutoff ≥ 9 showed the highest specificity, but the sensitivity was too low to be used as a screening tool.

Previous studies on comparing screening questionnaires for OSA have shown a wide-range of predictive performances depending on the studied population. Pataka et al. investigated predictive parameters of five different screening questionnaires for assessing OSA in a relatively large number of patients, who visited a sleep clinic [9]. They reported that the sensitivity of the BQ, STOP-B, and Four-V was 84.4, 95, and 78% in detecting OSA, respectively. The specificity of the BQ, STOP-B, and Four-V was 35.3, 14, and 40.8%, respectively. A Korean study performed in sleep clinic patients reported similar results [10]. In their study, the sensitivity of the BQ for OSA was 71.5% and the specificity was 32.0%. The sensitivity and specificity of the STOP-B were 97.0 and 18.6%, respectively. Patients who visited a sleep clinic were more likely to carry risk factors for OSA, such as the presence of snoring, obesity, or sleepiness, than the general population, and this selection bias could explain the higher sensitivity and lower specificity in the study population. Another important issue in validation of screening tests is the study design. In retrospective studies carried out in clinical settings, a verification bias can occur since subjects were usually pre-screened during the decision for the gold standard test. Subjects who had more severe symptoms were more likely to be advised to have PSG test compared to others with milder symptoms. This would make the prevalence of OSA higher among subjects who underwent PSG compared to its original prevalence, which eventually increases the sensitivity and decreases the specificity. A study by Jinmei et al., for example, reported a high sensitivity of the STOP-B in diagnosing moderate to severe OSA (96.5%) among the patients who visited the sleep-disordered breathing center [11]. The prevalence of OSA in their sample was 92.5%, and there existed a possibility that their sensitivity was overestimated. Since we recruited study subjects in a general population prospectively, we report relatively lower sensitivity but higher specificity compared to previous studies that were conducted in clinical settings retrospectively.

Table 3 Predictive values for the Berlin questionnaire, STOP, STOP-Bang questionnaire, and Four-variable screening tool for overall OSA (A) and moderate to severe OSA (B)

$N = 116$	BQ	STOP	STOP-B	Four-V (cutoff ≥ 8)	Four-V (cutoff ≥ 9)
(A) AHI ≥ 5					
Sensitivity (%)	61.3 (52.0–69.8)	58.1 (48.9–65.9)	61.3 (52.4–68.3)	69.4 (60.2–77.3)	54.8 (46.0–61.8)
Specificity (%)	63.0 (52.3–72.7)	74.1 (63.6–83.0)	79.6 (69.4–87.8)	68.5 (58.0–77.6)	81.5 (71.4–89.5)
PPV (%)	65.5 (55.6–74.6)	72.0 (60.7–81.7)	77.6 (66.3–88.6)	71.7 (62.2–79.8)	77.3 (64.8–87.1)
NPV (%)	58.6 (48.7–67.7)	60.6 (52.0–67.9)	64.2 (55.9–70.8)	66.1 (55.9–74.8)	61.1 (53.5–67.1)
Positive LR	1.655 (1.090–2.555)	2.240 (1.345–3.887)	3.009 (1.711–5.630)	2.203 (1.434–3.448)	2.961 (1.606–5.863)
Negative LR	0.615 (0.416–0.918)	0.566 (0.411–0.803)	0.486 (0.359–0.686)	0.447 (0.293–0.686)	0.544 (0.427–0.757)
DOR	2.692 (1.188–6.145)	3.956 (1.675–9.461)	6.189 (2.493–15.672)	4.926 (2.089–11.766)	3.611 (1.395–9.555)
(B) AHI ≥ 15					
Sensitivity (%)	58.6 (41.3–74.5)	58.6 (41.3–74.3)	72.4 (55.1–85.7)	69.0 (51.5–83.2)	58.6 (41.5–74.1)
Specificity (%)	52.9 (47.1–58.2)	62.1 (56.3–67.3)	67.8 (62.0–72.3)	54.0 (48.2–58.8)	69.0 (63.2–74.1)
PPV (%)	29.3 (20.7–37.2)	34.0 (24.0–43.1)	42.9 (32.6–50.7)	33.3 (24.9–40.2)	38.6 (27.3–48.9)
NPV (%)	79.3 (70.7–87.2)	81.8 (74.2–88.7)	88.1 (80.6–93.8)	83.9 (74.9–91.3)	83.3 (76.4–89.6)
Positive LR	1.244 (0.781–1.780)	1.545 (0.946–2.274)	2.250 (1.452–3.091)	1.500 (0.994–2.016)	1.889 (1.128–2.867)
Negative LR	0.783 (0.439–1.246)	0.667 (0.381–1.042)	0.407 (0.197–0.723)	0.574 (0.287–1.006)	0.600 (0.349–0.926)
DOR	1.589 (0.627–4.057)	2.318 (0.909–5.967)	5.531 (2.007–15.660)	2.611 (0.988–7.035)	3.148 (1.218–8.222)

BQ, Berlin questionnaire; STOP, STOP questionnaire; STOP-B, STOP-Bang questionnaire; Four-V, Four-variable screening tool; AHI, apnea-hypopnea index; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio; DOR, diagnostic odds ratio

Only a few studies have validated the BQ, STOP-Bang, or Four-V in a general population. Silva et al. compared the Four-V, STOP, STOP-Bang, and Epworth Sleepiness Scale (ESS) in a large community sample ($n = 4770$) [16] and reported that the Four-V showed the highest specificity (93.2%) in detecting moderate to severe OSA, but the sensitivity of the STOP-Bang was the highest (87.0%) among the studied questionnaires. Also, the BQ was validated in 643 community dwellers [15], showing a sensitivity of 77% and a specificity of 39% in detecting AHI ≥ 15. However, in those two studies, in-home PSG was performed to diagnose OSA instead of standard in-laboratory PSG. There is evidence that in-home PSG is associated with poor data quality and discordance in RDI with in-laboratory PSG, especially when it is performed unattended [24, 25]. In addition, the authors targeted older subjects (average age at 62.4 and 65.6 years, respectively) who generally carry more risk factors for OSA than younger people. This also could narrow the generalizability of their results.

An ideal screening test should be easy to perform, inexpensive, and have adequate diagnostic performance. Terms of ideal screening tests are different depending on the characteristics of the population that the tests are targeting. In a community-dwelling sample that shows a much lower prevalence of OSA than clinical samples, it is important that the screening test should have sufficient specificity, otherwise, it would lead healthy subjects to unnecessary testing and increase costs. Even though the Four-V with cutoff ≥ 9 showed the highest specificity in detecting OSA among questionnaires, the sensitivity was too low to be used as a screening tool. The STOP-B showed high specificity and acceptable sensitivity. In addition, mild OSA has not been proven in the literature to cause serious medical complications as much as moderate to severe OSA does [26]. Therefore, screening for OSA in a general population should aim to detect moderate to severe OSA. In the present study, the STOP-B showed the highest sensitivity for moderate to severe OSA among studied questionnaires and the specificity was also higher than the BQ and the Four-V with cutoff ≥ 8. Besides, the STOP-B consists of straightforward questions and has an easy scoring method. Previous studies also have shown that the STOP-B had good performance for the screening of OSA in various populations [19] and recommended the STOP-B as a proper screening tool for OSA due to its high methodological quality and user-friendly features [9, 10, 16, 27].

The Four-V was originally developed by Takegami et al. to screen moderate to severe OSA in Asian populations [20]. However, there is great variability in the predictive performance of the Four-V according to the reported studies. In the study by Takegami et al., the sensitivity and specificity for moderate to severe OSA were 73.9 and 66.1% with cut-offs ≥ 11, respectively. Using cut-off ≥ 9, the sensitivity increased to 91.3%, but the specificity decreased to 40.2%. In the study by Pataka et al., the sensitivity and specificity of the Four-V were 78.0 and 40.8% for OSA, respectively; the sensitivity and specificity for moderate to severe OSA were 79.0 and 36.0%, respectively [9]. In a large cohort study by Silva et al., the Four-V showed very low sensitivity (24.7%) and low specificity (41.5%) in detecting moderate to severe OSA

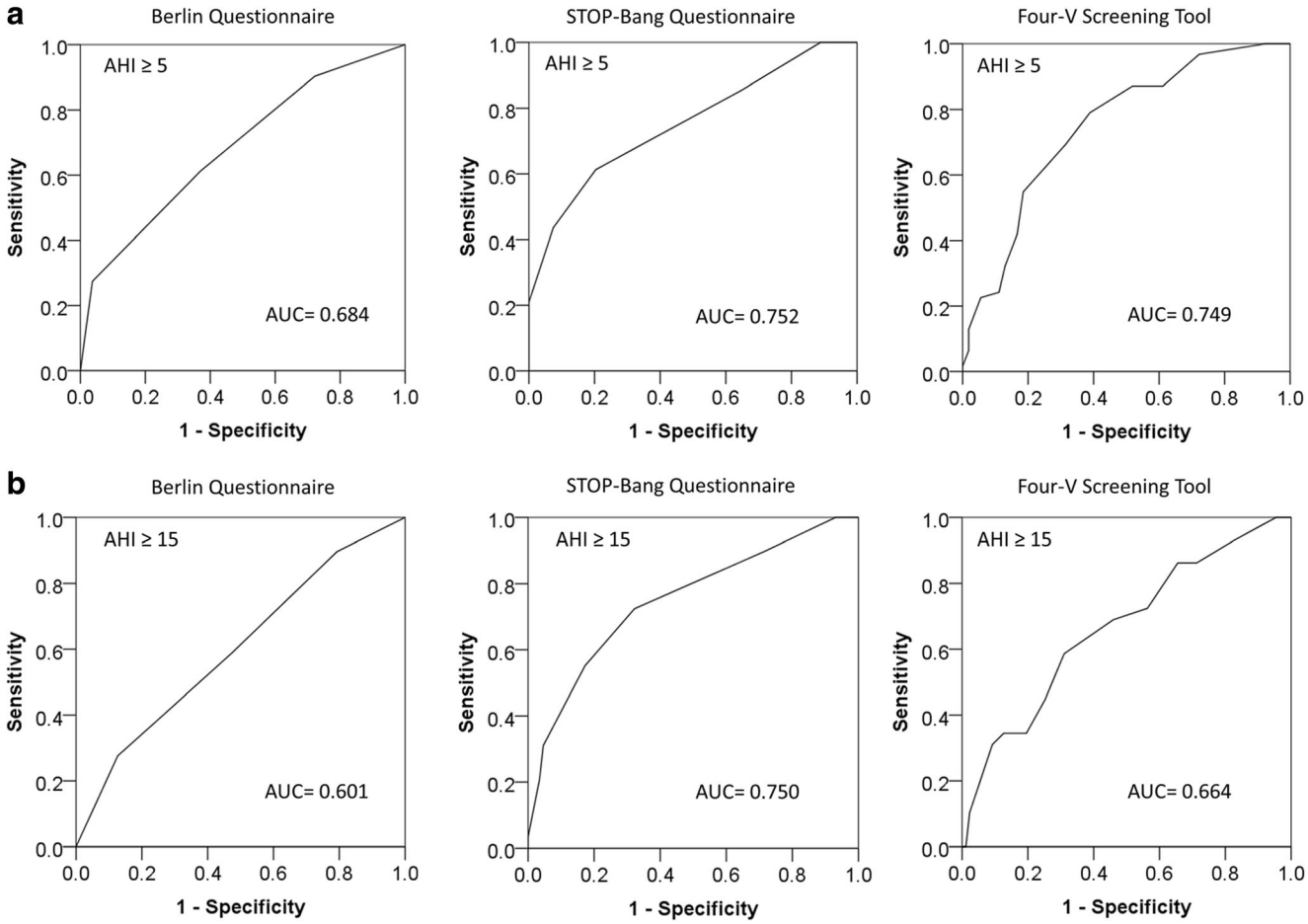

Fig. 2 Receiver-operator characteristic (ROC) curves for the Berlin questionnaire, STOP-Bang questionnaire, and Four-V screening tool with an AHI cutoff ≥ 5 (**a**) and ≥ 15 (**b**). The area under the ROC curve (AUC) closer to one indicates the higher diagnostic accuracy of the screening questionnaire. AHI, apnea-hypopnea index

with cut-off ≥ 14 [16]. In our study, the cut-offs ≥ 8 and 9 were evaluated in calculating predictive parameters since the prevalence of the high-risk with cut-off ≥ 11 was too low (16.2%) compared to those from the STOP-V (30.1%) and the BQ (32%). The predictive parameters shown in our study are close to the values reported by Takegami et al. in a community sample. The reason for observed variabilities could include the BP scoring system of the Four-V, which classifies BP into six grades regardless of taking anti-hypertensives. Therefore, patients with OSA who have well-controlled hypertension might not be properly screened by the Four-V. Besides, the Four-V weighs male gender greatly by scoring men with 4 points and women with only 1 point. This might contribute to the variability.

The BQ has been widely used internationally in screening OSA, and Ramachandran et al. concluded that the BQ is the most accurate screening questionnaires for diagnosis of OSA in a meta-analysis [28]. However, the BQ showed the lowest diagnostic performance among studied questionnaires in the present study. The predictive performance of the BQ varied depending on the studied populations in the literature [12, 15, 17]. Sharma et al. reported the sensitivity and the specificity of the BQ with AHI cut-off ≥ 5 as high as 85.5 and 95.2% [29], respectively; on the other hand, Ahmadi et al. argued that the BQ is not appropriate for identifying OSA reporting 61.8% sensitivity and 42.7% specificity [30]. The BQ includes several questions with multiple choices that need specific answers. Four out of five questions in the first category require detailed information on habitual snoring from a bed partner. For example, one item asks subjects how often other people witness them stop breathing. Besides, there is no specific guideline for subjects who do not drive regarding the question "Have you ever nodded off or fallen asleep while driving a vehicle?" In addition to this, a complexity of the scoring methods of the BQ limits its use in a large community sample.

There are several limitations in the present study. First, the samples in the present study were recruited based on advertisements. Therefore, it is possible that subjects who have symptoms or risk factors of OSA were willing to participate in the study and this would have influenced the results.

Second, some of the predictive parameters, such as PPV or NPV, might have been influenced by the high prevalence of OSA (53.5%) among the subjects who underwent PSG. Lastly, night-to-night variability has been suggested in sleep parameters on PSG [31]. Since the PSG was conducted only once for each subject, the effect of night-to-night variability cannot be excluded regarding the results.

Nevertheless, the key strength of the present study is that our study was a prospective study conducted in a community-dwelling sample with a wide age range. The sample consisted of both men and women that had an age structure close to the actual Korean age structure. We investigated the predictive performance of the three screening questionnaires for OSA comparing with overnight PSG that is a gold standard for diagnosing OSA. In conclusion, we recommend the STOP-B as a screening tool for OSA in the general population of Korea, as it showed high diagnostic performance for OSA and for moderate to severe OSA. In addition, the STOP-B consists of a small number of straightforward questions that are easy to answer and the scoring method is simple. The use of a screening questionnaire will be helpful in saving limited medical resources with respect to public health perspectives by screening OSA patients in general populations and providing them timely treatment.

Funding This study was funded by the Research Program funded by the Korea Centers for Disease Control and Prevention (2016-E34008-00).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Ancoli-Israel S, Kripke DF, Klauber MR, Fell R, Stepnowsky C, Estline E, Khazeni N, Chinn A (1996) Morbidity, mortality and sleep-disordered breathing in community dwelling elderly. Sleep 19(4):277–282
2. Shahar E, Whitney CW, Redline S, Lee ET, Newman AB, Nieto FJ, O'Connor GT, Boland LL, Schwartz JE, Samet JM (2001) Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the Sleep Heart Health Study. Am J Respir Crit Care Med 163(1):19–25. https://doi.org/10.1164/ajrccm.163.1.2001008
3. Kiely JL, McNicholas WT (2000) Cardiovascular risk factors in patients with obstructive sleep apnoea syndrome. Eur Respir J 16(1):128–133
4. Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, Mohsenin V (2005) Obstructive sleep apnea as a risk factor for stroke and death. N Engl J Med 353(19):2034–2041. https://doi.org/10.1056/NEJMoa043104
5. Dursunoglu N, Dursunoglu D, Kilic M (2005) Impact of obstructive sleep apnea on right ventricular global function: sleep apnea and myocardial performance index. Respiration 72(3):278–284. https://doi.org/10.1159/000085369
6. Lavie L, Polotsky V (2009) Cardiovascular aspects in obstructive sleep apnea syndrome—molecular issues, hypoxia and cytokine profiles. Respiration 78(4):361–370. https://doi.org/10.1159/000243552
7. Young T, Evans L, Finn L, Palta M (1997) Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. Sleep 20(9):705–706
8. Kushida CA, Littner MR, Morgenthaler T, Alessi CA, Bailey D, Coleman J Jr, Friedman L, Hirshkowitz M, Kapen S, Kramer M, Lee-Chiong T, Loube DL, Owens J, Pancer JP, Wise M (2005) Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 28(4):499–521
9. Pataka A, Daskalopoulou E, Kalamaras G, Fekete Passa K, Argyropoulou P (2014) Evaluation of five different questionnaires for assessing sleep apnea syndrome in a sleep clinic. Sleep Med 15(7):776–781. https://doi.org/10.1016/j.sleep.2014.03.012
10. Kim B, Lee EM, Chung YS, Kim WS, Lee SA (2015) The utility of three screening questionnaires for obstructive sleep apnea in a sleep clinic setting. Yonsei Med J 56(3):684–690. https://doi.org/10.3349/ymj.2015.56.3.684
11. Luo J, Huang R, Zhong X, Xiao Y, Zhou J (2013) 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 127(17):3065–3070
12. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, Khajehdehi A, Shapiro CM (2008) Validation of the Berlin questionnaire and American Society of Anesthesiologists checklist as screening tools for obstructive sleep apnea in surgical patients. Anesthesiology 108(5):822–830. https://doi.org/10.1097/ALN.0b013e31816d91b5
13. Chung F, Subramanyam R, Liao P, Sasaki E, Shapiro C, Sun Y (2012) High STOP-Bang score indicates a high probability of obstructive sleep apnoea. Br J Anaesth 108(5):768–775. https://doi.org/10.1093/bja/aes022
14. Devaraj U, Rajagopala S, Kumar A, Ramachandran P, Devereaux PJ, D'Souza GA (2017) Undiagnosed obstructive sleep apnea and postoperative outcomes: a prospective observational study. Respiration 94(1):18–25. https://doi.org/10.1159/000470914
15. Sforza E, Chouchou F, Pichot V, Herrmann F, Barthelemy JC, Roche F (2011) Is the Berlin questionnaire a useful tool to diagnose obstructive sleep apnea in the elderly? Sleep Med 12(2):142–146. https://doi.org/10.1016/j.sleep.2010.09.004
16. Silva GE, Vana KD, Goodwin JL, Sherrill DL, Quan SF (2011) Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth Sleepiness Scales. J Clin Sleep Med 7(5):467–472. https://doi.org/10.5664/JCSM.1308
17. Kang K, Park KS, Kim JE, Kim SW, Kim YT, Kim JS, Lee HW (2013) Usefulness of the Berlin questionnaire to identify patients at high risk for obstructive sleep apnea: a population-based door-to-door study. Sleep Breath 17(2):803–810. https://doi.org/10.1007/s11325-012-0767-2
18. World Health Organization (2000) The Asia-Pacific perspective: redefining obesity and its treatment. Health Communications Australia, Sydney
19. Nagappa M, Liao P, Wong J, Auckley D, Ramachandran SK, Memtsoudis S, Mokhlesi B, Chung F (2015) Validation of the

STOP-Bang questionnaire as a screening tool for obstructive sleep apnea among different populations: a systematic review and meta-analysis. PLoS One 10(12):e0143697. https://doi.org/10.1371/journal.pone.0143697
20. Takegami M, Hayashino Y, Chin K, Sokejima S, Kadotani H, Akashiba T, Kimura H, Ohi M, Fukuhara S (2009) Simple four-variable screening tool for identification of patients with sleep-disordered breathing. Sleep 32(7):939–948
21. Medicine AAoS (2007) Terminology and technical specifications. In: The AASM manual for the scoring of sleep and associated events: rules. AASM, Westchester
22. Quan S, Gillin JC, Littner M, Shepard J (1999) Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. Editorials. Sleep 22(5):662–689
23. Flemons WW, Buysse D, Redline S, Pack A, Strohl K, Wheatley J, Young T, Douglas N, Levy P, McNicholas W, Fleetham J, White D, Schmidt-Nowarra W, Carley D, Romaniuk J, Force AASMT (1999) Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. Sleep 22(5):667–689
24. Golpe R, Jimenez A, Carpizo R (2002) Home sleep studies in the assessment of sleep apnea/hypopnea syndrome. Chest 122(4): 1156–1161
25. Portier F, Portmann A, Czernichow P, Vascaut L, Devin E, Benhamou D, Cuvelier A, Muir JF (2000) Evaluation of home versus laboratory polysomnography in the diagnosis of sleep apnea syndrome. Am J Respir Crit Care Med 162(3 Pt 1):814–818. https://doi.org/10.1164/ajrccm.162.3.9908002
26. McNicholas WT, Bonsignore MR, Levy P, Ryan S (2016) Mild obstructive sleep apnoea: clinical relevance and approaches to management. Lancet Respir Med 4(10):826–834. https://doi.org/10.1016/S2213-2600(16)30146-1
27. Abrishami A, Khajehdehi A, Chung F (2010) A systematic review of screening questionnaires for obstructive sleep apnea. Can J Anaesth 57(5):423–438. https://doi.org/10.1007/s12630-010-9280-x
28. Ramachandran SK, Josephs LA (2009) A meta-analysis of clinical screening tests for obstructive sleep apnea. Anesthesiology 110(4): 928–939. https://doi.org/10.1097/ALN.0b013e31819c47b6
29. Sharma SK, Vasudev C, Sinha S, Banga A, Pandey RM, Handa KK (2006) Validation of the modified Berlin questionnaire to identify patients at risk for the obstructive sleep apnoea syndrome. Indian J Med Res 124(3):281–290
30. Ahmadi N, Chung SA, Gibbs A, Shapiro CM (2008) The Berlin questionnaire for sleep apnea in a sleep clinic population: relationship to polysomnographic measurement of respiratory disturbance. Sleep Breath 12(1):39–45. https://doi.org/10.1007/s11325-007-0125-y
31. Stepnowsky CJ Jr, Orr WC, Davidson TM (2004) Nightly variability of sleep-disordered breathing measured over 3 nights. Otolaryngol Head Neck Surg 131(6):837–843. https://doi.org/10.1016/j.otohns.2004.07.011