OBES SURG (2017) 27:716–729
DOI 10.1007/s11695-016-2352-4

ORIGINAL CONTRIBUTIONS

Pre-Operative Ability of Clinical Scores to Predict Obstructive Sleep Apnea (OSA) Severity in Susceptible Surgical Patients

E. Deflandre[1,2,3] · S. Degey[2] · J-F. Brichant[4] · A-F. Donneau[5] · R. Frognier[6] · R. Poirrier[7] · V. Bonhomme[8]

Published online: 6 September 2016

Abstract

Background Severe obstructive sleep apnea (OSA) is an independent risk factor for perioperative complications. Clinical scores such as Snoring, Tiredness, Observed apnea, high blood Pressure, Body Mass Index (BMI) higher than 35 kg m^{-2}, Age older than 50 years, Neck circumference larger than 40 cm, and male gender (STOP-Bang), perioperative sleep apnea prediction (P-SAP), and OSA50 have been proposed for detecting OSA. We recently proposed a new score based on morphological metrics only, the DES-OSA score. This study compared the DES-OSA score to the three other ones with regard to their ability to detect OSA. Obese patients are particularly at risk of OSA.

Methods Following informed consent and institutional review board (IRB) approval, 1584 consecutive adults were. Should the STOP-Bang be indicative of increased risk of severe OSA, the patient was referred to complementary polysomnography (PSG). Eventual already existing recent PSG data were also collected. The abilities of the four scores to predict OSA severity were compared using sensitivity, specificity, Cohen's kappa coefficient (CKC), and area under ROC curve (AUROC) analysis.

Results PSG was performed in 150 patients. For detecting severe OSA, OSA50 had the highest sensitivity [value (95 % CI) 0.98 (0.90–1)]. STOP-Bang was significantly less sensitive than P-SAP and OSA50. In that respect, DES-OSA was significantly more specific than the three other ones [0.75 (0.65–0.83)]. The AUROC of DES-OSA was significantly the largest [0.9 (0.84–0.95)]. The highest CKC at detecting severe OSA was 0.62 (0.49–0.74) for DES-OSA. Similar results were obtained for moderate to severe OSA prediction.

Conclusions DES-OSA, which is the only exclusively morphological score available, appears to surpass the three other scores in their ability to predict moderate to severe and severe OSA, at least in our setting and in our screened population.

Clinical Trial Registration ClinicalTrial.gov NCT02051829

Keywords Sleep apnea · Obstructive · Postoperative complications · Obesity · Perioperative period · Mass screening

This report was previously presented, in part at the ASA Annual Meeting (San Diego, 2015) and SASM Annual Meeting (San Diego, 2015). It won the third best clinical poster award at the SASM meeting.

✉ E. Deflandre
eric.deflandre@gmail.com

1 Department of Anesthesia, Clinique Saint-Luc of Bouge, Namur, Belgium

2 Cabinet Medical ASTES, Jambes, Belgium

3 University of Liege, Chaussee de Tongres, 29, 4000 Liege, Belgium

4 Department of Anesthesia, University Hospital of Liege, University of Liege, Liege, Belgium

5 Department of Health Science, University of Liege, Liege, Belgium

6 Department of Pneumology, Clinique Saint-Luc of Bouge, Namur, Belgium

7 Department of Neurology, University of Liege, Liege, Belgium

8 University Department of Anesthesia and ICM, CHR Citadelle and CHU, Liege, Belgium

Introduction

The obstructive sleep apnea (OSA) syndrome represents an independent risk factor for perioperative complications [1–9]. OSA affects at least 7 and 8 % of the surgical population [10, 11]. Singh et al. recently demonstrated that, during the pre-

operative visit, anesthesiologists and surgeons fail to diagnose more than 60 and 92 % of mild and severe OSA patients, respectively [12]. Hence, there is a need for a systematic screening of OSA during the pre-operative consultation [13]. This screening will have two goals. First, it will prompt practitioners planning a precise postoperative management for monitoring suspected (or confirmed) OSA patients that do not routinely use a continuous positive airway pressure (CPAP) machine during the night and planning a perioperative opioid sparing strategy. Second, it will help refine criteria for ambulatory procedure eligibility [14–16]. Since the work of Mutter et al., it is widely accepted that practitioners should focus their attention to detect undiagnosed severe OSA patients [1]. These patients are at a high risk of postoperative complications. Obese patients are particularly at risk of exhibiting an OSA.

Polysomnography (PSG) still represents the gold standard of OSA diagnosis [17]. Evidently, submitting all surgical patients to PSG before their intervention is economically not affordable. In order to facilitate OSA detection, several predictive clinical scores have been proposed. They aim at detecting patients with increased risk of at least mild, moderate to severe, or severe OSA. The ideal score should be easy to use and performed within a reasonable amount of time. It should also have a favorable sensitivity-specificity ratio at detecting OSA. For example, the Sleep Disorders Questionnaire (SDQ), developed by Douglass et al. in 1994 [18], is one of the most sensitive-specific score [18, 19]. However, the 175 items composing the SDQ do not make it applicable during a routine pre-operative screening. In that setting, three clinical scores are commonly proposed. The STOP-Bang questionnaire, developed by Chung et al. [20, 21], is certainly the most frequently used. It includes the following eight criteria: Snoring, Tiredness, Observed apnea, high blood Pressure, Body Mass Index (BMI) higher than 35 kg m^{-2}, Age older than 50 years, Neck circumference larger than 40 cm, and male gender. The perioperative sleep apnea prediction (P-SAP) score, proposed by Ramachandran et al. [10], takes account of snoring, reduced tyro-mental distance (<6 cm), presence of type 2 diabetes, presence of high blood pressure, Mallampati class III or IV, BMI higher than 30 kg m^{-2}, age equal to or over 43 years, neck circumference larger than 40 cm, and male gender. The OSA50, developed by Chai-Coetzer et al., has four weighted parameters, including Obesity and waist circumference larger than 88 cm for females and 102 cm for males, Snoring, observed Sleep Apneas, and age over 50 years [22]. Currently, considering the number of STOP-Bang-related publications, this score is the gold standard of OSA screening scores [23].

The main drawback of clinical scores involving anamnestic criteria relies in the need for trusting patient responses to questioning regarding usual snoring, sensation of tiredness, and witnessed apnea, for example. A language barrier, the ignorance of the patient on his/her health status, and the need for a daily life witness may impede the reliability of obtained responses. To overcome those potential bias, we recently developed a new diagnostic score for OSA, the DES-OSA score [24]. This score compiles morphological criteria only, including the Mallampati score, the distance between thyroid and chin, the BMI, the neck circumference, and gender. Each criterion is weighted 1, 2, or 3 points, according to their observed value (Table 5 of the Appendix). A cutoff value of 7 corresponds to significantly increased probability of severe OSA [24].

Ideally, the reliability of those scores to predict OSA and its severity should be as high as possible. We here tested the hypothesis that the DES-OSA score performs better in this indication. Indeed, DES-OSA has been specifically designed to detect severe OSA [24], and the exclusive morphological nature of its criteria makes it probably less sensible to patients' subjectivity. The aim of this study was to compare the DES-OSA score to the three other ones, namely, the STOP-Bang score considered as the gold standard and used as a pre-screening tool, the P-SAP score, and the OSA50 score, in a homogeneous sample drawn from a pre-surgical population. We compared their ability to detect at least mild, moderate to severe, and severe OSA in terms of sensibility, specificity, Cohen's kappa coefficient, and area under ROC curves.

Material and Methods

Institutional Review Board/Consent

The study was approved by our Institutional Ethics Committee (Clinique Saint-Luc of Bouge, Dr. P. van der Rest, 20 April 2013, Namur, Belgium) with the reference number CE SLBO 2014/01. Due to the fact that this study is limited to an analysis of data that are routinely collected during the pre-anesthetic visit in our institution, the institutional review board (IRB) waived the written informed consent. However, oral informed consent was always obtained.

Clinical Trial Registration

This study was registered prior to patient enrolment at ClinicalTrial.gov NCT02051829.

Consecutive consenting patients (n = 1584), aged 18 years old or more, were screened. Those patients were initially admitted for a pre-anesthetic visit. In our institution, all patients undergo such a visit in the facilities of a pre-anesthesia clinic. Patient's height and weight were objectively measured, and age, gender, and ASA physical status were recorded. In addition, all items needed to calculate the four studied scores were compiled, including the following morphological criteria [24]. The distance

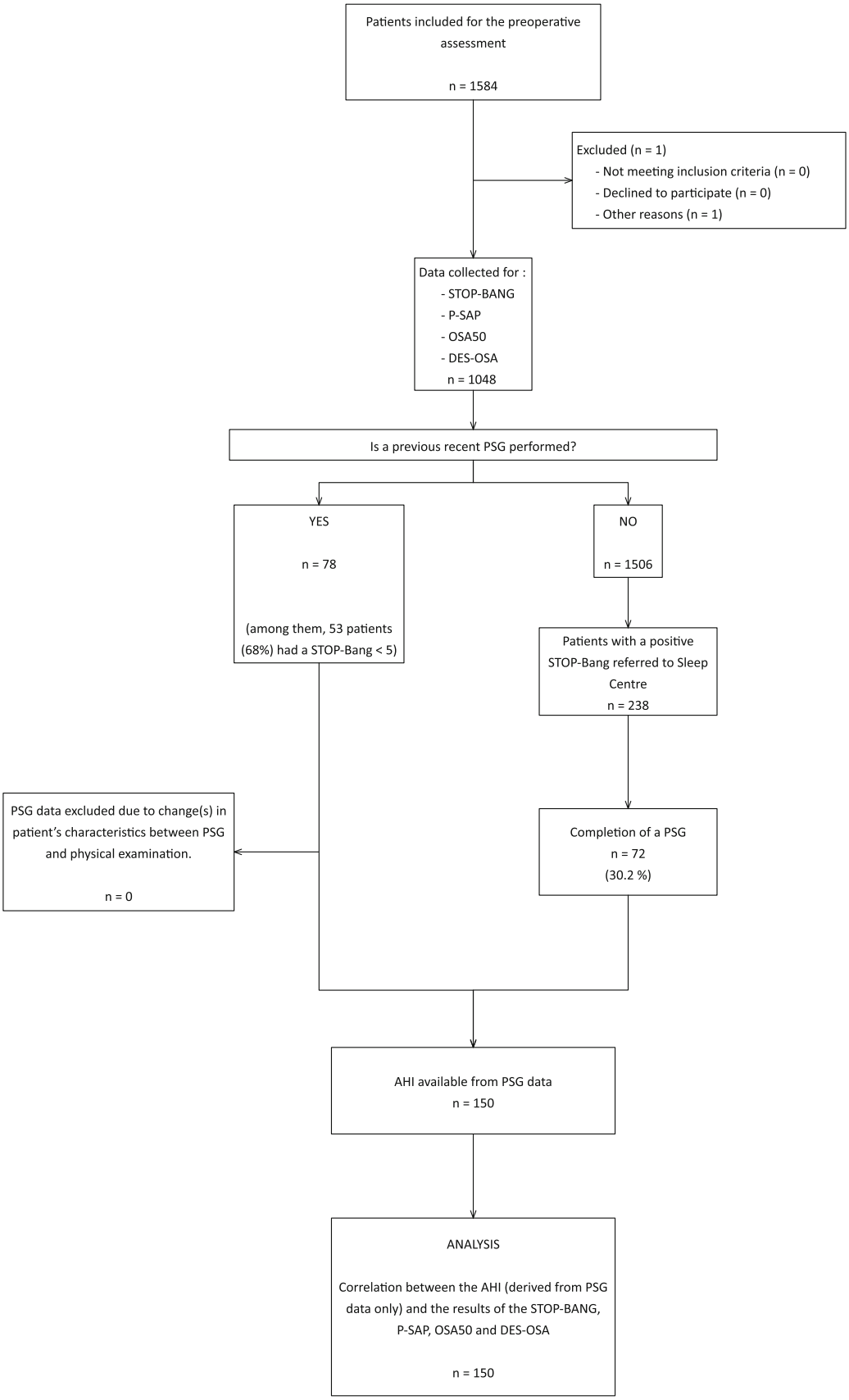

Fig. 1 Flowchart illustrating patient's recruitment. *PSG* polysomnography, *AHI* apnea-hypopnea index

Table 1 Demographic and other characteristics of patients in the screened population and in patients who achieved PSG

	Screened population	PSG	Statistics
Demographic data			
Number	1583	150	
Age (years)[b]	57.24 ± 17.05	59.66 ± 12.41	$t_{(1731)} = -1.697$
			$P = 0.090$
Male/female (%)[a]	49.10/50.90	70.00/30.00	$\chi^2_{(1)} = 9.067$
			$P = 0.003$*
ASA physical status[a]			$\chi^2_{(2)} = 6.092$
			$P = 0.048$*
ASA I (%)	11.37	2.67	
ASA II (%)	81.13	90.00	
ASA III (%)	5.5	7.33	
Distribution of patients according to their apnea and hypopnea index (AHI)			–
AHI >5 events h^{-1} (%)	–	89.33	
AHI >15 events h^{-1} (%)	–	62.00	
AHI >30 events h^{-1} (%)	–	42.00	
AHI (events h^{-1})	–	22.00 (10.25 to 41.27)	
Distribution of patients according to their Mallampati (MP) score[a]			$\chi^2_{(2)} = 5.991$
			$P < 0.001$*
MP I (%)	34.36	18.00	
MP II (%)	53.88	50.00	
MP III and IV (%)	11.75	32.00	
Distribution of patients according to their distance between thyroid and chin (DTC)[a]			$\chi^2_{(4)} = 1.010$
			$P = 0.603$
DTC >6 cm (%)	85.72	81.33	
DTC <6 and >5 cm (%)	13.14	16.00	
DTC <5 cm (%)	1.14	2.67	
Distribution of patients according to their body mass index (BMI)[a]			$\chi^2_{(4)} = 1.925$
			$P = 0.750$
BMI >28 kg/m (%)	42.39	78.00	
BMI >30 kg/m (%)	30.07	62.00	
BMI >35 kg/m (%)	10.93	30.67	
BMI >39 kg/m (%)	4.67	14.00	
BMI >41 kg/m (%)	3.09	9.33	
BMI (kg/m)[b]	27.79 ± 5.82	32.36 ± 2.26	$t_{(1731)} = -9.120$
			$P < 0.001$*
Distribution of patients according to their neck circumference (NC)[a]			$\chi^2_{(4)} = 9.847$
			$P = 0.043$*
NC >37 cm (%)	52.62	83.33	
NC >40 cm (%)	26.47	63.67	
NC >42 cm (%)	14.53	50.00	
NC >45 cm (%)	4.80	22.67	
NC >=48 cm (%)	1.39	8.67	
NC (cm)[b]	37.82 ± 4.26	41.97 ± 4.64	$t_{(1731)} = -11.304$
			$P < 0.001$*
Distribution of patients according to other characteristics			
Snoring (%)[a]	50.09	82.00	$\chi^2_{(1)} = 22.703$
			$P < 0.001$*
Tired (%)[a]	33.73	64.67	$\chi^2_{(1)} = 19.151$
			$P < 0.001$*
Observed apneas (%)[a]	14.34	54.67	$\chi^2_{(1)} = 35.986$
			$P < 0.001$*
Hypertension (%)[a]	36.89	56.00	$\chi^2_{(1)} = 7.341$
			$P = 0.007$*
Diabetes (%)[a]	10.30	22.67	$\chi^2_{(1)} = 5.557$
			$P = 0.018$*

Degrees of freedom are indicated between brackets

*A two-tailed P value <0.05 was considered significant

[a] Frequency data, between-group statistical comparisons made using a χ^2 test

[b] Normally distributed data presented as mean ± standard deviation, between-group comparisons made using two-tailed independent Student's t tests

[c] Non-normally distributed data presented as median (25th–75th percentile), between-group comparisons made using a Mann-Whitney U test

Table 2 Ability of the four scores to predict at least mild, moderate to severe, and severe OSA (AHI >5, 15, and 30 events h^{-1}, respectively) in terms of sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and negative likelihood ratio (−LR)

		STOP-Bang		P-SAP		OSA50		DES-OSA	
		Value	95 % CI	Value	95 % CI	Value	95 % CI	Value	95 % CI
AHI >5 Events h^{-1}	Se	0.679[2,3]	0.596 to 0.572	0.821[2,4,6]	0.746 to 0.877	0.933[3,5,6]	0.875 to 0.965	0.567[4,5]	0.483 to 0.648
	Sp	0.625	0.385 to 0.815	0.438[4]	0.232 to 0.668	0.313[5]	0.141 to 0.559	0.875[4,5]	0.625 to 0.975
	PPV	0.938	0.899 to 0.977	0.924	0.882 to 0.966	0.919	0.875 to 0.963	0.974	0.949 to 0.999
	NPV	0.189	0.126 to 0.252	0.226	0.159 to 0.293	0.357	0.280 to 0.434	0.194	0.131 to 0.257
	+LR	1.811	1.617 to 2.005	1.459	1.328 to 1.590	1.357	1.246 to 1.468	4.537	3.896 to 5.178
	−LR	0.513	0.433 to 0.593	0.409	0.330 to 0.488	0.215	0.149 to 0.281	0.495	0.415 to 0.575
AHI >15 Events h^{-1}	Se	0.753[2,3]	0.655 to 0.829	0.882[2,4]	0.798 to 0.934	0.946[3,5]	0.876 to 0.979	0.710[4,5]	0.610 to 0.792
	Sp	0.526[1,2,3]	0.399 to 0.650	0.351[2,4,6]	0.240 to 0.481	0.158[3,5,6]	0.084 to 0.277	0.789[1,4,5]	0.665 to 0.876
	PPV	0.722	0.650 to 0.794	0.689	0.615 to 0.763	0.647	0.571 to 0.723	0.846	0.788 to 0.904
	NPV	0.566	0.487 to 0.645	0.645	0.568 to 0.722	0.643	0.566 to 0.720	0.625	0.548 to 0.702
	+LR	1.589	1.434 to 1.744	1.358	1.246 to 1.470	1.124	1.064 to 1.184	3.371	2.919 to 3.823
	−LR	0.470	0.390 to 0.550	0.337	0.261 to 0.413	0.341	0.265 to 0.417	0.368	0.191 to 0.445
AHI >30 Events h^{-1}	Se	0.810[2,3]	0.694 to 0.888	0.968[2]	0.884 to 0.997	0.984[3]	0.906 to 1.000	0.889	0.784 to 0.947
	Sp	0.471[1,2,3]	0.370 to 0.575	0.333[2,4,6]	0.243 to 0.438	0.149[3,5,6]	0.089 to 0.241	0.747[1,4,5]	0.646 to 0.827
	PPV	0.526	0.446 to 0.606	0.513	0.433 to 0.593	0.456	0.376 to 0.536	0.718	0.646 to 0.790
	NPV	0.774	0.707 to 0.841	0.935	0.896 to 0.974	0.929	0.888 to 0.970	0.903	0.856 to 0.950
	+LR	1.531	1.387 to 1.675	1.452	1.322 to 1.582	1.147	1.089 to 1.225	3.515	3.039 to 3.991
	−LR	0.404	0.325 to 0.483	0.095	0.048 to 0.142	0.106	0.057 to 0.155	0.149	0.092 to 0.206

Results are given with threshold value as defined by their authors (5 for the STOP-Bang score, 4 for the P-SAP score, 5 for the OSA50 score, and 7 for the DES-OSA score). Results are given with 95 % confidence interval (95 % CI). Mc Nemar test was applied to compare sensitivities and specificities. To correct for multiple comparisons and to avoid type I errors, the level of statistical significance was set at $P = 0.0083$ (0.05/6). Significant results ($P < 0.0083$) were indicated as follows: "1" between STOP-Bang and DES-OSA, "2" between STOP-Bang and P-SAP, "3" between STOP-Bang and OSA50, "4" between DES-OSA and P-SAP, "5" between DES-OSA and OSA50, and "6" between OSA50 and P-SAP. The highest values for each analysis (Se, Sp, PPV, NPV, and +LR) and the lowest values for −LR are indicated in hatched cells

between thyroid and chin (DTC), neck circumference, and modified Mallampati score were measured using the methods described in our previous paper [24], also reported by other authors [25, 26]. If the STOP-Bang, considered as the standard, was indicative of increased risk of severe OSA according to previously reported thresholds (≥5), the patient was referred by the anesthesiologist to a pre-operative screening of OSA through PSG. Once PSG had been performed, the apnea-hypopnea index (AHI) was collected. AHI is defined as the number of apnea and hypopnea per hour and corresponds to OSA syndrome severity. An AHI value higher than 5, 15, and 30 events h^{-1} reflects at least mild, moderate to severe, and severe OSA, respectively [27].

Monitoring during PSG was achieved using the DREAM® device (Medatec®, Brussels, Belgium), which allows continuous recording at a 200-Hz sampling rate of a five-channel electroencephalogram (EEG), left and

Table 3 Comparison of area under ROC curves (95 % confidence intervals) between the four scores according to their ability to detect at least mild, moderate to severe, or severe OSA (AHI >5, >15, and >30 events h^{-1}, respectively)

AHI		STOP-Bang	P-SAP	OSA50	DES-OSA
>5 events h^{-1}	Areas under ROC curves (95 % CI)	0.712 (0.595 to 0.828)	0.739 (0.630 to 0.849)	0.626 (0.492 to 0.760)	0.809 (0.719 to 0.899)
>15 events h^{-1}	Areas under ROC curves (95 % CI)	0.713 (0.631 to 0.794)	0.719[2] (0.638 to 0.799)	0.666[3] (0.580 to 0.753)	0.827[2,3] (0.763 to 0.891)
>30 events h^{-1}	Areas under ROC curves (95 % CI)	0.732[1] (0.649 to 0.815)	0.761[2] (0.681 to 0.841)	0.659[3] (0.570 to 0.749)	0.896[1,2,3] (0.841 to 0.952)

To correct for multiple comparisons and to avoid type I errors, the level of statistical significance was set at $P = 0.0083$ (0.05/6). Significant results ($P < 0.0083$) are indicated as follows: "1" between STOP-Bang and DES-OSA, "2" between DES-OSA and P-SAP, and "3" between DES-OSA and OSA50

right electrooculogram (EOG), electrocardiogram (ECG), submental electromyogram (EMG), left and right tibial EMG, thoracic-abdominal movements, nasal air flow, plethysmographic pulse wave, peripheral blood oxygen saturation, and snoring sound. Signals were recorded on a dedicated computer using the Brainnet® for Windows software (Medatec®, Brussels, Belgium).

The analysis of PSG-recorded data included the determination of sleep stages according to the Rechtschaffen and Kales scoring rules [28] on 30-s adjacent epochs. Periods of arousal were also considered, according to the American Sleep Disorders Association (ASDA) [29]. They were defined as longer than 3-s bursts of α waves occurring in one of the EEG traces (C4-A1) during non-rapid eye movement (NREM) sleep or occurring concomitantly to increased muscle activity during rapid eye movement (REM) sleep. Criteria for the definition of apnea and hypopnea were in agreement with the American Academy of Sleep Medicine (AASM) rules [30].

Data of all 1584 patients were compiled into a Microsoft Excel spreadsheet. As indicated above, patients with a STOP-Bang score higher than 5 underwent a PSG pre-operative screening, and the AHI recorded during that screening was collected. If patients had undergone a PSG during the year before the pre-anesthetic visit, the AHI derived from that PSG was collected and patients were not requested to undergo a new PSG. We ensured that the patient's BMI remained stable between the time of PSG completion and the pre-anesthetic visit. If this was not the case, those PSG data were excluded from the analysis. The same was applied when patients had modified their "age" criterion in between, changing from less than to more than 43 years, which impacts on P-SAP calculation, or changing from less than to more than 50 years, which impacts on STOP-BANG and OSA50 calculation. Figure 1 illustrates the patient recruitment.

The ability of the four scores to detect at least mild, moderate to severe, or severe OSA was compared using sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), and area under ROC curve (AUROC) analysis. We also used the Cohen's kappa coefficient to assess the degree of concordance between clinical scores and AHI. The kappa value can be interpreted as follows: <0.20 corresponds to poor agreement, 0.21 to 0.40 to fair agreement, 0.41 to 0.60 to moderate agreement, 0.61 to 0.80 to good agreement, and 0.81 to 1.00 to very good agreement [31]. The Mc Nemar test was applied to compare the sensitivities and specificities derived from each OSA-screening score and a z test to compare AUROC's [32]. Comparisons of Cohen's kappa coefficients were performed using their 95 % CIs. A two-tailed P value <0.05 was considered statistically significant. Statistics were performed using XLSTAT for Mac® (version 2016.02) and MedCalc® (version 15.6.1). Results of calculated statistics are always presented as value (95 % CI), unless otherwise indicated.

Results

One patient was excluded from data analysis for the reason that he had benefitted from an uvulo-palato-pharyngoplasty a few years before and had therefore an unreliable Mallampati score. PSG was performed in 150 patients. Out of these 150 PSGs, 78 had already been performed shortly before the pre-anesthetic visit and inclusion into the study, and 72 were performed following the visit, according to the results of the STOP-Bang. Among the 78 patients who had already a PSG, 53 % had a STOP-Bang <5. Demographic and other characteristics of the whole sample are reported in Table 1.

Table 2 summarizes the sensitivity-specificity analysis results. In our pre-surgical patients who underwent PSG, the highest Se at detecting severe OSA (AHI >30 events h^{-1}) was obtained for the OSA50 [0.98 (0.9 to 1.00)], and the highest Sp for the same ability was obtained for the DES-OSA [0.75 (0.65 to 0.83)]. Mc Nemar testing revealed that DES-OSA was significantly more specific than the three other scores at detecting an AHI >30 events h^{-1} (Table Table 3) and that STOP-Bang was significantly more specific but less sensitive than P-SAP and OSA50 in that respect. For the same ability, DES-OSA had the highest PPV [0.72 (0.65 to 0.79)], and P-SAP had the highest NPV [0.93 (0.87 to 0.97) and 0.90 (0.71 to 1.00), respectively]. Corresponding highest positive likelihood ratio was obtained for the DES-OSA [3.51 (3.04 to 3.99)], and lowest negative likelihood ratio was obtained for P-SAP [0.09 (0.05 to 0.14].

The area under ROC curves (AUROC) for the four scores are summarized in Table Table 3 and displayed in Fig. 2. The largest AUROC at detecting severe OSA was the one of DES-OSA [0.9 (0.84 to 0.95)]. z tests revealed that DES-OSA had a significantly larger AUROC than the three other scores.

The results of the Cohen's kappa coefficient to assess the degree of concordance between clinical scores and AHI are provided in Table 4. Concordance was significantly better between DES-OSA and an AHI value

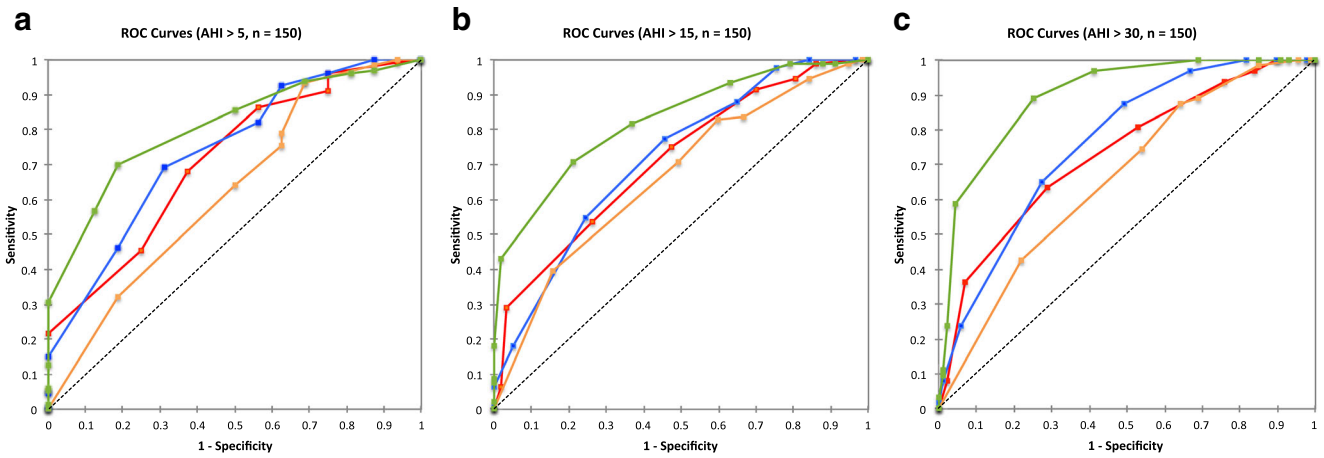

Fig. 2 Illustrations of ROC curves for the four scores (*red lines* STOP-Bang, *blue lines* P-SAP, *orange lines* OSA50, *green lines* DES-OSA). Results concern the ability of the scores to detect at least mild (**a**), moderate to severe (**b**), and severe (**c**) OSA. *AHI* apnea-hypopnea index (events h^{-1})

indicating moderate to severe or severe OSA (AHI >15 and >30 events h^{-1}, respectively) than the other scores.

STOP-Bang, P-SAP, OSA50, and/or DES-OSA scores were indicating increased risk of OSA in 16.6, 35.9, 62.7, and 12.6 % of screened patients (n = 1583), respectively.

Complete results are presented in the digital Appendix.

Discussion and Conclusion

This study compared DES-OSA, STOP-Bang, P-SAP, and OSA50 scores with regard to their ability to detect OSA during a pre-operative anesthetic visit in patients with an a priori increased risk of OSA according to a classically used clinical score, the main objective being the detection of severe OSA patients. Our results demonstrate superiority of DES-OSA at detecting moderate to severe and severe OSA patients.

The fact that DES-OSA performs best at detecting severe OSA is not surprising, insofar as its development was intended to detect such [24]. In that respect, DES-OSA may be considered as a useful tool for detecting severe OSA patients, that is those with the highest perioperative risk of respiratory complications and those where ambulatory surgery is particularly questionable [1, 14, 33]. Our study presents the advantage of having been performed in patients originating from a general population and presenting for a pre-anesthetic visit. It was not a selected population with a history of sleep disorders [34] and is therefore representative of the population anesthesiologists encounter most.

Our results indicate that the four studied score (STOP-Bang, P-SAP, OSA50, and DES-OSA) are efficient at

Table 4 Comparison of Cohen's kappa coefficients (95 % confidence intervals) to assess the degree of agreement in their ability to detect at least mild, moderate to severe, or severe OSA (AHI >5, >15, and >30 events h^{-1}, respectively). *Red cells* poor agreement, *orange cells* fair agreement, *blue cells* moderate agreement, *green cells* good agreement

AHI		STOP-Bang	P-SAP	OSA50	DES-OSA
> 5 events h^{-1}	Cohen Kappa Coefficient (95% CI)	0.151 (0.015 to 0.286)	0.183 (0.001 to 0.365)	0.260 (0.034 to 0.598)	0.174 (0.071 to 0.277)
> 15 events h^{-1}	Cohen Kappa Coefficient (95% CI)	0.283 (0.125 to 0.440)	0.255 (0.104 to 0.406)	0.122 (-0.001 to 0.244)	0.475 (0.336 to 0.614)
> 30 events h^{-1}	Cohen Kappa Coefficient (95% CI)	0.261 (0.125 to 0.398)	0.269 (0.163 to 0.374)	0.115 (0.042 to 0.189)	0.616 (0.493 to 0.739)

detecting OSA. However, there are significant between-score differences. All of them have the advantage of being fast to perform and easy to implement for the practitioner. The DES-OSA score has the added benefit of using morphological characteristics only; most of which are routinely collected during the pre-operative visit in some centers. DES-OSA has the best predictive ability for severe OSA because it has the highest specificity combined with a good sensitivity. DES-OSA has also the highest positive predictive value and positive likelihood ratio in that respect, while P-SAP and OSA50 have the highest negative predictive value and the lowest negative likelihood ratio. The AUROC was significantly larger for DES-OSA than the three other scores at detecting severe OSA. The Cohen's kappa coefficient analyses demonstrate that DES-OSA has moderate agreement at detecting at least moderate OSA and good agreement at detecting severe OSA, while the other scores have poor to fair agreement.

Our study corroborates our initial hypothesis that DES-OSA performs better than the three other scores in the detection of severe OSA. The reasons for better performance may reside in the fact that DES-OSA was initially designed to detect such and that it does not depend on patients' response to questioning. This is in accordance with the findings of Nunes et al., who reported that the clinical characteristics of OSA and the responses given to items of screening questionnaires can be affected by the underlying cause of hospitalization (coronary artery bypass grafting or abdominal surgery) [35]. Although this has not been specifically studied in this study, such bias is probably not a concern when using an exclusively morphological score.

This study also validates the DES-OSA score in a different population than the one from which it was built. Despite populations of different origin, sensitivities and specificities of the DES-OSA reveal similarities in both [24]. It could be used for mass screening, in particular in susceptible population such as obese patients.

Our study has a few limitations. The first lies in the fact that our morphological DES-OSA score was developed in a European population and is here validated again in a European population. Ethnic factors may influence apnea mechanisms and incidence [36]. A validation of DES-OSA in another ethnic population could be necessary before employing it worldwide. The second corresponds to the low proportion of patients having completed the PSG (30.2 %). However, this percentage is in accordance with previous studies dealing with pre-operative OSA screening [21]. Finally, except for patients with previously available PSG data, PSG was performed only in patients with a positive STOP-Bang score, omitting the information given by the other scores. This may have biased our results in favor of the STOP-Bang score. Indeed, according to recommended thresholds, P-SAP, OSA50, and/or DES-OSA scores were indicating increased risk of OSA in 35.9, 62.7, and 12.6 % of screened patients, respectively. Sending all those patients to a PSG screening would have overloaded our sleep center and was not affordable. The only way of overcoming this limitation would be to perform a prospective screening in an entire sample of consecutive patients admitted for a pre-anesthesia evaluation, irrespective of their calculated predictive scores. However, this would limit possible recruitment, for evident practical reasons. Other authors had the same limitations in their study [37].

In conclusion, this study allowed comparing the ability of four OSA screening scores in a surgical population and validate a new score against a largely used one. Although the ideal score does not exist, DES-OSA, which is the only morphological and multiple level-weighted score, appears to be highly reproducible and to surpass the three other scores in their ability to predict moderate to severe and severe OSA.

Details of Authors' Contributions E.D. Study design, data analysis, and writing up of the first draft of the paper.

S.D. Patient recruitment, data collection, and writing up the first draft of the paper.

J-F.B. Writing up of the first draft of the paper.

A-F.D. Data analysis.

R.F. Patient recruitment and data collection.

R.P. Writing up of the first draft of the paper.

V.B. Study design and writing up the first draft of the paper.

Compliance with Ethical Standards

Funding None.

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.

Statement of Informed Consent Informed consent was obtained orally from all individual participants included in the study.

Appendix 1

Table 5 The DES-OSA score

	1 point	2 points	3 points	Patient's Value
MC		II	III et IV	
DTC (cm)	> 6	5 – 6	< 5	
BMI (kg/m)	> 28	> 39	> 41	
NC (cm)	> 37	> 42	> 48	
Gender	Male			
SUM OF THE ABOVE 5 PARAMETERS = DES-OSA SCORE:				

MC Mallampati class, *DTC* distance between thyroid and chin, *BMI* body mass index, *NC* neck circumference

Table 6 Statistics observed for multiple threshold values of the STOP-Bang score in terms of number of true positives (TPs), true negatives (TNs), false positives (FPs), false negatives (FNs), sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and negative likelihood ratio (−LR) at detecting an AHI >5, >15, and >30 events/h

		Th	TP (*n*)	TN (*n*)	FP (*n*)	FN (*n*)	Sensitivity			Specificity			PPV	NPV	+LR	−LR
							Values	95 % CI		Values	95 % CI					
								Inf	Sup		Sup	Inf				
PSG (*n* = 150)	AHI >5 events h^{-1}	0	134	0	16	0	1.000	0.966	1.000	0.000	0.000	0.231	0.893		1.000	
		1	133	1	15	1	0.993	0.954	1.000	0.063	0.000	0.306	0.899	0.500	1.059	0.119
		2	129	4	12	5	0.963	0.913	0.986	0.250	0.099	0.501	0.915	0.444	1.284	0.149
		3	122	4	12	12	0.910	0.848	0.949	0.250	0.099	0.501	0.910	0.250	1.214	0.358
		4	116	7	9	18	0.866	0.796	0.914	0.438	0.232	0.668	0.928	0.280	1.539	0.307
		5	91	10	6	43	0.679	0.596	0.752	0.625	0.385	0.815	0.938	0.189	1.811	0.513
		6	61	12	4	73	0.455	0.373	0.540	0.750	0.499	0.901	0.938	0.141	1.821	0.726
		7	29	16	0	105	0.216	0.155	0.294	1.000	0.769	1.000	1.000	0.132	+Inf	0.784
		8	7	16	0	127	0.052	0.024	0.106	1.000	0.769	1.000	1.000	0.112	+Inf	0.948
	AHI >15 events h^{-1}	0	93	0	57	0	1.000	0.951	1.000	0.000	0.000	0.077	0.620		1.000	
		1	93	2	55	0	1.000	0.951	1.000	0.035	0.003	0.128	0.628	1.000	1.036	0.000
		2	92	8	49	1	0.989	0.935	1.000	0.140	0.071	0.257	0.652	0.889	1.151	0.077

Table 6 (continued)

	Th	TP (*n*)	TN (*n*)	FP (*n*)	FN (*n*)	Sensitivity			Specificity			PPV	NPV	+LR	−LR
						Values	95 % CI		Values	95 % CI					
							Inf	Sup		Sup	Inf				
	3	88	11	46	5	0.946	0.876	0.979	0.193	0.110	0.316	0.657	0.688	1.173	0.279
	4	85	17	40	8	0.914	0.836	0.957	0.298	0.195	0.428	0.680	0.680	1.302	0.288
	5	70	30	27	23	0.753	0.655	0.829	0.526	0.399	0.650	0.722	0.566	1.589	0.470
	6	50	42	15	43	0.538	0.437	0.635	0.737	0.609	0.834	0.769	0.494	2.043	0.627
	7	27	55	2	66	0.290	0.208	0.390	0.965	0.872	0.997	0.931	0.455	8.274	0.735
	8	6	56	1	87	0.065	0.028	0.137	0.982	0.897	1.000	0.857	0.392	3.677	0.952
AHI >30 events h^{-1}	0	63	0	87	0	1.000	0.929	1.000	0.000	0.000	0.052	0.420		1.000	
	1	63	2	85	0	1.000	0.929	1.000	0.023	0.002	0.086	0.426	1.000	1.024	0.000
	2	63	9	78	0	1.000	0.929	1.000	0.103	0.054	0.188	0.447	1.000	1.115	0.000
	3	61	14	73	2	0.968	0.884	0.997	0.161	0.098	0.254	0.455	0.875	1.154	0.197
	4	59	21	66	4	0.937	0.842	0.979	0.241	0.163	0.342	0.472	0.840	1.234	0.263
	5	51	41	46	12	0.810	0.694	0.888	0.471	0.370	0.575	0.526	0.774	1.531	0.404
	6	40	62	25	23	0.635	0.511	0.743	0.713	0.609	0.797	0.615	0.729	2.210	0.512
	7	23	81	6	40	0.365	0.257	0.489	0.931	0.854	0.970	0.793	0.669	5.294	0.682
	8	5	85	2	58	0.079	0.031	0.178	0.977	0.914	0.998	0.714	0.594	3.452	0.942

Table 7 Statistics observed for multiple threshold values of the P-SAP score in terms of number of true positives (TPs), true negatives (TNs), false positives (FPs), false negatives (FNs), sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and negative likelihood ratio (−LR) at detecting an AHI >5, >15, and >30 events/h

		Th	TP (*n*)	TN (*n*)	FP (*n*)	FN (*n*)	Sensitivity			Specificity			PPV	NPV	+LR	−LR
							Values	95 % CI		Values	95 % CI					
								Inf	Sup		Sup	Inf				
PSG (*n* = 150)	AHI >5 events h^{-1}	0	134	0	16	0	1.000	0.966	1.000	0.000	0.000	0.231	0.893		1.000	
		1	134	2	14	0	1.000	0.966	1.000	0.125	0.025	0.375	0.905	1.000	1.143	0.000
		2	129	4	12	5	0.963	0.913	0.986	0.250	0.099	0.501	0.915	0.444	1.284	0.149
		3	124	6	10	10	0.925	0.866	0.960	0.375	0.185	0.615	0.925	0.375	1.481	0.199
		4	110	7	9	24	0.821	0.746	0.877	0.438	0.232	0.668	0.924	0.226	1.459	0.409
		5	93	11	5	41	0.694	0.611	0.766	0.688	0.441	0.859	0.949	0.212	2.221	0.445
		6	62	13	3	72	0.463	0.381	0.547	0.813	0.560	0.940	0.954	0.153	2.468	0.661
		7	20	16	0	114	0.149	0.098	0.220	1.000	0.769	1.000	1.000	0.123	+Inf	0.851
		8	6	16	0	128	0.045	0.019	0.097	1.000	0.769	1.000	1.000	0.111	+Inf	0.955
		9	1	16	0	133	0.007	0.000	0.046	1.000	0.769	1.000	1.000	0.107	+Inf	0.993
	AHI >15 events h^{-1}	0	93	0	57	0	1.000	0.951	1.000	0.000	0.000	0.077	0.620		1.000	
		1	93	2	55	0	1.000	0.951	1.000	0.035	0.003	0.128	0.628	1.000	1.036	0.000
		2	93	9	48	0	1.000	0.951	1.000	0.158	0.084	0.277	0.660	1.000	1.188	0.000
		3	91	14	43	2	0.978	0.919	0.998	0.246	0.152	0.373	0.679	0.875	1.297	0.088
		4	82	20	37	11	0.882	0.798	0.934	0.351	0.240	0.481	0.689	0.645	1.358	0.337
		5	72	31	26	21	0.774	0.678	0.848	0.544	0.416	0.666	0.735	0.596	1.697	0.415
		6	51	43	14	42	0.548	0.447	0.645	0.754	0.627	0.848	0.785	0.506	2.233	0.599
		7	17	54	3	76	0.183	0.117	0.275	0.947	0.849	0.987	0.850	0.415	3.473	0.863
		8	6	57	0	87	0.065	0.028	0.137	1.000	0.923	1.000	1.000	0.396	+Inf	0.935
		9	1	57	0	92	0.011	0.000	0.065	1.000	0.923	1.000	1.000	0.383	+Inf	0.989
	AHI >30 events h^{-1}	0	63	0	87	0	1.000	0.929	1.000	0.000	0.000	0.052	0.420		1.000	
		1	63	2	85	0	1.000	0.929	1.000	0.023	0.002	0.086	0.426	1.000	1.024	0.000
		2	63	9	78	0	1.000	0.929	1.000	0.103	0.054	0.188	0.447	1.000	1.115	0.000
		3	63	16	71	0	1.000	0.929	1.000	0.184	0.116	0.280	0.470	1.000	1.225	0.000
		4	61	29	58	2	0.968	0.884	0.997	0.333	0.243	0.438	0.513	0.935	1.452	0.095
		5	55	44	43	8	0.873	0.765	0.936	0.506	0.403	0.608	0.561	0.846	1.766	0.251
		6	41	63	24	22	0.651	0.527	0.757	0.724	0.621	0.807	0.631	0.741	2.359	0.482
		7	15	82	5	48	0.238	0.150	0.358	0.943	0.868	0.978	0.750	0.631	4.143	0.808
		8	5	86	1	58	0.079	0.031	0.178	0.989	0.930	1.000	0.833	0.597	6.905	0.931
		9	1	87	0	62	0.016	0.000	0.094	1.000	0.948	1.000	1.000	0.584	+Inf	0.984

Table 8 Statistics observed for multiple threshold values of the OSA50 score in terms of number of true positives (TPs), true negatives (TNs), false positives (FPs), false negatives (FNs), sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and negative likelihood ratio (−LR) at detecting an AHI >5, >15, and >30 events/h

		Th	TP (*n*)	TN (*n*)	FP (*n*)	FN (*n*)	Sensitivity			Specificity			PPV	NPV	+LR	−LR
							Values	95 % CI		Values	95 % CI					
								Inf	Sup		Sup	Inf				
PSG (*n* = 150)	AHI >5 events h^{-1}	0	134	0	16	0	1.000	0.966	1.000	0.000	0.000	0.231	0.893		1.000	
		2	134	1	15	0	1.000	0.966	1.000	0.063	0.000	0.306	0.899	1.000	1.067	0.000
		3	132	2	14	2	0.985	0.943	0.999	0.125	0.025	0.375	0.904	0.500	1.126	0.119
		5	125	5	11	9	0.933	0.875	0.965	0.313	0.141	0.559	0.919	0.357	1.357	0.215
		6	106	6	10	28	0.791	0.714	0.851	0.375	0.185	0.615	0.914	0.176	1.266	0.557
		7	101	6	10	33	0.754	0.674	0.819	0.375	0.185	0.615	0.910	0.154	1.206	0.657
		8	86	8	8	48	0.642	0.557	0.718	0.500	0.281	0.719	0.915	0.143	1.284	0.716
		10	43	13	3	91	0.321	0.248	0.404	0.813	0.560	0.940	0.935	0.125	1.711	0.836
	AHI >15 events h^{-1}	0	93	0	57	0	1.000	0.951	1.000	0.000	0.000	0.077	0.620		1.000	
		2	93	1	56	0	1.000	0.951	1.000	0.018	0.000	0.103	0.624	1.000	1.018	0.000
		3	92	3	54	1	0.989	0.935	1.000	0.053	0.013	0.151	0.630	0.750	1.044	0.204
		5	88	9	48	5	0.946	0.876	0.979	0.158	0.084	0.277	0.647	0.643	1.124	0.341
		6	78	19	38	15	0.839	0.749	0.900	0.333	0.225	0.463	0.672	0.559	1.258	0.484
		7	77	23	34	16	0.828	0.737	0.892	0.404	0.286	0.533	0.694	0.590	1.388	0.426
		8	66	29	28	27	0.710	0.610	0.792	0.509	0.383	0.634	0.702	0.518	1.445	0.571
		10	37	48	9	56	0.398	0.304	0.500	0.842	0.723	0.916	0.804	0.462	2.520	0.715
	AHI >30 events h^{-1}	0	63	0	87	0	1.000	0.929	1.000	0.000	0.000	0.052	0.420		1.000	
		2	63	1	86	0	1.000	0.929	1.000	0.011	0.000	0.070	0.423	1.000	1.012	0.000
		3	63	4	83	0	1.000	0.929	1.000	0.046	0.015	0.117	0.432	1.000	1.048	0.000
		5	62	13	74	1	0.984	0.906	1.000	0.149	0.089	0.241	0.456	0.929	1.157	0.106
		6	56	27	60	7	0.889	0.784	0.947	0.310	0.223	0.414	0.483	0.794	1.289	0.358
		7	55	31	56	8	0.873	0.765	0.936	0.356	0.264	0.461	0.495	0.795	1.356	0.356
		8	47	40	47	16	0.746	0.625	0.837	0.460	0.359	0.564	0.500	0.714	1.381	0.552
		10	27	68	19	36	0.429	0.314	0.551	0.782	0.683	0.856	0.587	0.654	1.962	0.731

Table 9 Statistics observed for multiple threshold values of the DES-OSA score in terms of number of true positives (TPs), true negatives (TNs), false positives (FPs), false negatives (FNs), sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and negative likelihood ratio (−LR) at detecting an AHI >5, >15, and >30 events/h

		Th	TP (*n*)	TN (*n*)	FP (*n*)	FN (*n*)	Sensitivity			Specificity			PPV	NPV	+LR	−LR
							Values	95 % CI		Values	95 % CI					
								Inf	Sup		Sup	Inf				
PSG (*n* = 150)	AHI >5 events h^{-1}	1	134	0	16	0	1.000	0.966	1.000	0.000	0.000	0.231	0.893		1.000	
		2	130	2	14	4	0.970	0.922	0.991	0.125	0.025	0.375	0.903	0.333	1.109	0.239
		3	129	3	13	5	0.963	0.913	0.986	0.188	0.060	0.440	0.908	0.375	1.185	0.199
		4	126	5	11	8	0.940	0.884	0.971	0.313	0.141	0.559	0.920	0.385	1.368	0.191
		5	115	8	8	19	0.858	0.788	0.908	0.500	0.281	0.719	0.935	0.296	1.716	0.284
		6	94	13	3	40	0.701	0.619	0.772	0.813	0.560	0.940	0.969	0.245	3.741	0.367
		7	76	14	2	58	0.567	0.483	0.648	0.875	0.625	0.975	0.974	0.194	4.537	0.495
		8	41	16	0	93	0.306	0.234	0.389	1.000	0.769	1.000	1.000	0.147	+Inf	0.694
		9	17	16	0	117	0.127	0.080	0.195	1.000	0.769	1.000	1.000	0.120	+Inf	0.873
		10	8	16	0	126	0.060	0.029	0.116	1.000	0.769	1.000	1.000	0.113	+Inf	0.940

Table 9 (continued)

	Th	TP (*n*)	TN (*n*)	FP (*n*)	FN (*n*)	Sensitivity			Specificity			PPV	NPV	+LR	−LR
						Values	95 % CI		Values	95 % CI					
							Inf	Sup		Sup	Inf				
	11	7	16	0	127	0.052	0.024	0.106	1.000	0.769	1.000	1.000	0.112	+Inf	0.948
	12	2	16	0	132	0.015	0.001	0.057	1.000	0.769	1.000	1.000	0.108	+Inf	0.985
AHI >15 events h^{-1}	1	93	0	57	0	1.000	0.951	1.000	0.000	0.000	0.077	0.620		1.000	
	2	92	5	52	1	0.989	0.935	1.000	0.088	0.035	0.195	0.639	0.833	1.084	0.123
	3	92	7	50	1	0.989	0.935	1.000	0.123	0.059	0.237	0.648	0.875	1.128	0.088
	4	92	12	45	1	0.989	0.935	1.000	0.211	0.124	0.335	0.672	0.923	1.253	0.051
	5	87	21	36	6	0.935	0.863	0.972	0.368	0.255	0.499	0.707	0.778	1.481	0.175
	6	76	36	21	17	0.817	0.725	0.883	0.632	0.501	0.745	0.784	0.679	2.218	0.289
	7	66	45	12	27	0.710	0.610	0.792	0.789	0.665	0.876	0.846	0.625	3.371	0.368
	8	40	56	1	53	0.430	0.334	0.532	0.982	0.897	1.000	0.976	0.514	24.516	0.580
	9	17	57	0	76	0.183	0.117	0.275	1.000	0.923	1.000	1.000	0.429	+Inf	0.817
	10	8	57	0	85	0.086	0.043	0.164	1.000	0.923	1.000	1.000	0.401	+Inf	0.914
	11	7	57	0	86	0.075	0.035	0.151	1.000	0.923	1.000	1.000	0.399	+Inf	0.925
	12	2	57	0	91	0.022	0.002	0.081	1.000	0.923	1.000	1.000	0.385	+Inf	0.978
AHI >30 events h^{-1}	1	63	0	87	0	1.000	0.929	1.000	0.000	0.000	0.052	0.420		1.000	
	2	63	6	81	0	1.000	0.929	1.000	0.069	0.030	0.146	0.438	1.000	1.074	0.000
	3	63	8	79	0	1.000	0.929	1.000	0.092	0.046	0.174	0.444	1.000	1.101	0.000
	4	63	13	74	0	1.000	0.929	1.000	0.149	0.089	0.241	0.460	1.000	1.176	0.000
	5	63	27	60	0	1.000	0.929	1.000	0.310	0.223	0.414	0.512	1.000	1.450	0.000
	6	61	51	36	2	0.968	0.884	0.997	0.586	0.481	0.684	0.629	0.962	2.340	0.054
	7	56	65	22	7	0.889	0.784	0.947	0.747	0.646	0.827	0.718	0.903	3.515	0.149
	8	37	83	4	26	0.587	0.464	0.700	0.954	0.883	0.985	0.902	0.761	12.774	0.433
	9	15	85	2	48	0.238	0.150	0.358	0.977	0.914	0.998	0.882	0.639	10.357	0.780
	10	7	86	1	56	0.111	0.053	0.216	0.989	0.930	1.000	0.875	0.606	9.667	0.899
	11	6	86	1	57	0.095	0.042	0.197	0.989	0.930	1.000	0.857	0.601	8.286	0.915
	12	2	87	0	61	0.032	0.003	0.116	1.000	0.948	1.000	1.000	0.588	+Inf	0.968

Table 10 Comparison of sensibilities and specificities of the four scores at predicting an AHI value >5, >15, and >30 events/h (at least mild, moderate to severe, and severe OSA, respectively) using the Mc Nemar test

				STOP-Bang	P-SAP	OSA50	DES-OSA
PSG (*n* = 150)	Sensitivity (%)	AHI >5 events/h	STOP-Bang	–	<0.001*	<0.001*	0.022
			P-SAP	–	–	0.001*	<0.001*
			OSA50	–	–	–	<0.001*
			DES-OSA	–	–	–	–
		AHI >15 events/h	STOP-Bang	–	0.001*	<0.001*	0.433
			P-SAP	–	–	0.058	0.001*
			OSA50	–	–	–	<0.001*
			DES-OSA	–	–	–	–
		AHI >30 events/h	STOP-Bang	–	0.002*	0.001*	0.166
			P-SAP	–	–	0.564	0.059
			OSA50	–	–	–	0.034
			DES-OSA	–	–	–	–
	Specificity (%)	AHI >5 events/h	STOP-Bang	–	0.180	0.025	0.102

Table 10 (continued)

			STOP-Bang	P-SAP	OSA50	DES-OSA
		P-SAP	–	–	0.317	0.008*
		OSA50	–	–	–	0.003*
		DES-OSA	–	–	–	–
	AHI >15 events/h	STOP-Bang	–	0.007*	<0.001*	0.002*
		P-SAP	–	–	0.002*	<0.001*
		OSA50	–	–	–	<0.001*
		DES-OSA	–	–	–	–
	AHI >30 events/h	STOP-Bang	–	0.005*	<0.001*	<0.001*
		P-SAP	–	–	<0.001*	<0.001*
		OSA50	–	–	–	<0.001*
		DES-OSA	–	–	–	–

*To correct for multiple comparisons and avoid type I errors, the level of statistical significance was set at $P = 0.0083$ (0.05/6)

Table 11 Comparison of the areas under ROC curves describing the ability of the four scores at detecting an AHI value >5, >15, and >30 events/h (at least mild, moderate to severe, and severe OSA, respectively) using a *z* test

			STOP-Bang	P-SAP	OSA50	DES-OSA
PSG ($n = 150$)	AHI >5 events/h	STOP-Bang	–	0.587	0.096	0.172
		P-SAP	–	–	0.074	0.294
		OSA50	–	–	–	0.017
		DES-OSA	–	–	–	–
	AHI >15 events/h	STOP-Bang	–	0.844	0.186	0.009
		P-SAP	–	–	0.262	0.005*
		OSA50	–	–	–	0.002*
		DES-OSA	–	–	–	–
	AHI >30 events/h	STOP-Bang	–	0.397	0.050	<0.001*
		P-SAP	–	–	0.031	<0.001*
		OSA50	–	–	–	<0.001*
		DES-OSA	–	–	–	–

*To correct for multiple comparisons and to avoid type I errors, the level of statistical significance was set at $P = 0.0083$ (0.05/6)

Table 12 Percentage of patients with a positive result in each of the four analyzed score

	STOP-Bang ≥5	P-SAP ≥4	OSA50 ≥5	DES-OSA ≥7
Among whole population ($n = 1583$)	16.60	35.92	62.69	12.63
Among patients with a STOP-Bang ≥5 ($n = 263$)	–	94.68	98.10	53.23
Among patients with a P-SAP ≥4 ($n = 569$)	43.76	–	89.81	31.81
Among patients with a P-SAP ≥5 ($n = 993$)	25.98	51.46	–	18.73
Among patients with a DES-OSA ≥7 ($n = 200$)	70.00	90.50	93.00	–

References

1. Mutter TC, Chateau D, Moffatt M, Ramsey C, Roos LL, Kryger MA. Matched cohort study of postoperative outcomes in obstructive sleep apnea: could preoperative diagnosis and treatment prevent complications? Anesthesiology. 2014;121:707–18.
2. Dahan A, Aarts L, Smith TW. Incidence, reversal, and prevention of opioid-induced respiratory depression. Anesthesiology. 2010;112:226–38.
3. Eikermann M, Blobner M, Groeben H, et al. Postoperative upper airway obstruction after recovery of the train of four ratio of the adductor pollicis muscle from neuromuscular blockade. Anesth Analg. 2006;102:937–42.
4. Peppard PE, Young T, Palta M, Skatrud J. Prospective study of the association between sleep-disordered breathing and hypertension. N Engl J Med. 2000;342:1378–84.
5. Leung RS, Bradley TD. Sleep apnea and cardiovascular disease. Am J Respir Crit Care Med. 2001;164:2147–65.
6. Durgan DJ, Bryan RM. Cerebrovascular consequences of obstructive sleep apnea. J Am Heart Assoc. 2012;1:1–15.
7. Ehrhardt J, Schwab M, Finn S, et al. Sleep apnea and asymptomatic carotid stenosis: a complex interaction. Chest. 2015;147:1029–36.
8. Bazan V, Grau N, Valles E, et al. Obstructive sleep apnea in patients with typical atrial flutter: prevalence and impact on arrhythmia control outcome. Chest. 2013;143:1277–83.
9. Deflandre E, Degey S, Opsomer N, Brichant JF, Joris J. Obstructive sleep apnea and smoking as a risk factor for venous thromboembolism events: review of the literature on the common pathophysiological mechanisms. Obes Surg. 2016;26:640–8.
10. Ramachandran SK, Kheterpal S, Consens F, et al. Derivation and validation of a simple perioperative sleep apnea prediction score. Anesth Analg. 2010;110:1007–15.
11. Memtsoudis S, Liu S, Ma Y, et al. Perioperative pulmonary outcomes in patients with sleep apnea after noncardiac surgery. Anesth Analg. 2011;112:113–21.
12. Singh M, Liao P, Kobah S, Wijeysundera DN, Shapiro C, Chung F. Proportion of surgical patients with undiagnosed obstructive sleep apnoea. Br J Anaesth. 2013;110:629–36.
13. Mehta V, Subramanyam R, Shapiro C, Chung F. Health effects of identifying patients with undiagnosed obstructive sleep apnea in the preoperative clinic: a follow-up study. Can J Anesth. 2012;59:544–55.
14. Ankichetty S, Chung F. Considerations for patients with obstructive sleep apnea undergoing ambulatory surgery. Curr Opin Anaesthesiol. 2011;24:605–11.
15. Kurrek MM, Cobourn C, Wojtasik Z, Kiss A, Dain SL. Morbidity in patients with or at high risk for obstructive sleep apnea after ambulatory laparoscopic gastric banding. Obes Surg. 2011;21:1494–8.
16. Joshi GP, Ankichetty SP, Gan TJ, Chung F. Society for ambulatory anesthesia consensus statement on preoperative selection of adult patients with obstructive sleep apnea scheduled for ambulatory surgery. Anesth Analg. 2012;115:1060–8.
17. Flemons WW, Littner MR. Measuring agreement between diagnostic devices. Chest. 2003;124:1535–42.
18. Douglass AB, Bornstein R, Nino-Murcia G, et al. The sleep disorders questionnaire. I: creation and multivariate structure of SDQ. Sleep. 1994;17:160–7.
19. Ramachandran SK, Josephs LAA. Meta-analysis of clinical screening tests for obstructive sleep apnea. Anesthesiology. 2009;110:928–39.
20. Chung F, Yegneswaran B, Liao P, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. Anesthesiology. 2008;108:812–21.
21. Chung F, Subramanyam R, Liao P, Sasaki E, Shapiro C, Sun Y. High STOP-Bang score indicates a high probability of obstructive sleep apnoea. Br J Anaesth. 2012;108:768–75.
22. Chai-Coetzer CL, Antic NA, Rowland LS, et al. A simplified model of screening questionnaire and home monitoring for obstructive sleep apnoea in primary care. Thorax. 2011;66:213–9.
23. Nagappa M, Liao P, Wong J, 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. doi:10.1371/journal.pone .eCollection 2015
24. Deflandre E, Degey S, Brichant JF, Poirrier R, Bonhomme V. Development and validation of a morphologic obstructive sleep apnea prediction score: the DES-OSA score. Anesth Analg. 2016;122:363–72.
25. Lewis M, Keramati S, Benumof JL, Berry CC. What is the best way to determine oropharyngeal classification and mandibular space length to predict difficult laryngoscopy? Anesthesiology. 1994;81:69–75.
26. Samsoon GL, Young JR. Difficult tracheal intubation: a retrospective study. Anaesthesia. 1987;42:487–90.
27. Adesanya A, Lee W, Greilich N, Joshi G. Perioperative management of obstructive sleep apnea. Chest. 2010;138:1489–98.
28. Rechtshaffen A, Kales A. A manual of standardized terminology and scoring system for sleep stages of human subjects. Washington, DC: U.S. Government Printing Office ; 1968.NIH Publication No. 204
29. ASDA. EEG arousals: scoring rules and examples: a preliminary report from the sleep disorders atlas task force of the American Sleep Disorders Association. Sleep. 1992;15:173–84.
30. Berry RB, Budhiraja R, Gottlieb DJ, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. Deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine. J Clin Sleep Med. 2012;8:597–619.
31. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. 1977;33:363–74.
32. Hanley JA, McNeil BJA. Method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148:839–43.
33. Practice guidelines for the perioperative management of patients with obstructive sleep apnea: an updated report by the American Society of Anesthesiologists Task Force on perioperative management of patients with obstructive sleep apnea. *Anesthesiology* 2014;120: 268–86
34. Abrishami A, Khajehdehi A, Chung FA. Systematic review of screening questionnaires for obstructive sleep apnea. Can J Anaesth. 2010;57:423–38.
35. Nunes F, Danzi-Soares N, Genta P, Drager L, Cesar LM, Lorenzi-Filho G. Critical evaluation of screening questionnaires for obstructive sleep apnea in patients undergoing coronary artery bypass grafting and abdominal surgery. Sleep Breath. 2015;19:115–22.
36. Cakirer B, Hans MG, Graham G, Aylor J, Tishler PV, Redline S. The relationship between craniofacial morphology and obstructive sleep apnea in whites and in African-Americans. Am J Respir Crit Care Med. 2001;163:947–50.
37. Chung F, Abdullah HR, Liao P. STOP-Bang questionnaire: a practical approach to screen for obstructive sleep apnea. Chest. 2016;149:631–8.