Sleep Medicine Reviews 30 (2016) 25–33

Contents lists available at ScienceDirect

Sleep Medicine Reviews

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

CLINICAL REVIEW

Prediction of oral appliance treatment outcomes in obstructive sleep apnea: A systematic review

Kentaro Okuno*, Benjamin T. Pliska, Mona Hamoda, Alan A. Lowe, Fernanda R. Almeida

Department of Oral Health Sciences, Faculty of Dentistry, The University of British Columbia, Canada

ARTICLE INFO

Article history:
Received 3 July 2015
Received in revised form
14 October 2015
Accepted 25 November 2015
Available online 8 December 2015

Keywords:
Sleep apnea
Prediction
Oral appliance
Systematic review
QUADAS

SUMMARY

While oral appliances (OA) have demonstrated good efficacy in patients ranging from mild to severe levels of obstructive sleep apnea (OSA), this form of treatment is not completely effective in all patients. As a successful treatment response is not dependent solely on apnea hypopnea index severity, the prediction of OA treatment efficacy is of key importance for efficient disease management. This systematic review aims to investigate the accuracy of a variety of clinical and experimental tests for predicting OA treatment outcomes in OSA. A systematic literature review was conducted and the quality of the selected studies was assessed using the quality assessment of diagnostic accuracy studies (QUADAS-2) tool. Some 17 studies involving various prediction methods were included in this review. The predictive accuracy varied depending on the definitions of treatment success used as well as the type of index test. The studies with the best predictive accuracy and lowest risk of bias and concerns of applicability used a multisensor catheter. While a remotely controlled mandibular positioner study showed high accuracy, there was a high risk of bias. The available information on the validity of predictive index tests is very useful in clinical practice and allows for greater disease management efficiency.

Introduction

Obstructive sleep apnea (OSA) is a common syndrome that is characterized by recurrent episodes of partial or complete upper airway obstruction during sleep, resulting in sleep fragmentation and oxygen desaturation. OSA is associated with reduced quality of life, decreased cardiovascular health, and increased healthcare utilization and mortality [1,2]. Continuous positive airway pressure (CPAP) is an efficient treatment for OSA and has been demonstrated to improve daytime symptoms and to reduce cardiovascular disease [3]. Although CPAP is highly efficacious in preventing upper airway collapse, patient acceptance, tolerance, and adherence are often low, consequently reducing effectiveness [4].

Treatment with an oral appliance (OA) is an alternative to CPAP for OSA and although less efficacious, it is more acceptable by patients. An American Academy of Sleep Medicine and American Academy of Dental Sleep Medicine clinical practice guideline recommends OA treatment for adult patients with OSA who prefer OA therapy or are intolerant of CPAP therapy [5]. A recent comprehensive review of OA treatment showed that a complete response occurred in around 48% of patients, with a range of 29%–71% among studies [6]. At present, patient selection for OA therapy is largely based on the apnea hypopnea index (AHI) severity alone. However, patients with severe OSA who successfully respond to OA therapy have also been reported [7–9]. Treatment recommendations based solely on AHI restrict a potentially preferred treatment option to a small portion of OSA patients. As the efficacy of OAs varies greatly in patients with OSA, the prediction of OA treatment response is of key importance for efficient disease management.

A number of studies have reported predictors of OA treatment outcomes using polysomnographic parameters [10–13], cephalogram [14,15], CPAP pressure [16,17], spirometer [18], drug-induced sleep endoscopy [19], remotely controlled mandibular positioner [20,21], and multisensor catheter parameters [22]. However all these studies are derivation studies rather than validation studies, which are lacking in the existing literature. While these methods

Abbreviations: AHI, apnea hypopnea index; BMI, body mass index; CPAP, continuous positive airway pressure; OA, oral appliance; ODI, oxygen desaturation index; OSA, obstructive sleep apnea; PSG, polysomnography; QUADAS, quality assessment of diagnostic accuracy studies.

* Corresponding author. Department of Oral Health Sciences, Faculty of Dentistry, The University of British Columbia, 2199 Wesbrook Mall, Vancouver, B.C., V6T 1Z3, Canada. Tel.: +1 604 822 3750; fax: +1 604 822 3562.
E-mail address: k-okuno@dent.osaka-u.ac.jp (K. Okuno).

http://dx.doi.org/10.1016/j.smrv.2015.11.007
1087-0792/

still have some clinical importance, they vary greatly in terms of technical complexity, prediction accuracy, and clinical applicability and have not been systematically reviewed, which makes comparisons difficult.

This systematic review aims to investigate the accuracy of a variety of clinical and experimental tests in predicting OA treatment outcomes in OSA using the quality assessment of diagnostic accuracy studies (QUADAS-2) tool.

Methods

Eligibility criteria

This review includes studies that evaluate the accuracy of clinical tests for the prediction of OA treatment outcomes. Participants in each study have been diagnosed with OSA by polysomnography (PSG) and have been treated with an OA that functions to protrude the mandible. Studies of appliances that hold the tongue forward by suction (tongue retaining devices) have been specifically excluded from this review as they have been shown to be poorly tolerated and display inadequate retention in some patients and this could reduce effectiveness [23]. The study intervention included the index test predicting OA treatment response, which was compared to the reference PSG test of evaluating OA treatment outcomes.

Literature search

The electronic databases of Medline, EMBASE, Web of Science, cumulative index to nursing and allied health literature (CINAHL), and the Cochrane Library were independently searched by two authors (K.O., F.A.) on 20 November 2014. A search strategy was developed and executed with the following target population keywords used for the literature search: (((((((“Sleep Apnea, Obstructive”[Mesh]) OR (obstructive sleep AND (apnoea OR apnea)) OR (sleep AND (breathing disorder* OR respiratory disorder*)))))) AND ((((“Orthodontic Appliances”[Mesh]) OR ((oral OR dental OR (mandib* AND (advancement* OR repositioning))) AND (device* OR appliance* OR splint))))) AND (predict*).

Study selection

The included studies assessed the predictive accuracy of OA treatment outcomes in patients with OSA. Two authors (K.O., F.A.) independently screened the titles and abstracts, followed by a screening of the possibly relevant full-text articles. No restrictions were applied to the year of publication or language.

Data extraction

Data extraction was independently completed by two authors (K.O., F.A.) and included author, year, type of study, characteristics of the study population, level of evidence, type of index test, definition of a successful treatment outcome, and reference standard. Study outcomes were sensitivity, specificity, positive predictive value, and negative predictive value. Sensitivity refers to the test's capacity to identify individuals who responded to treatment; the higher the value, the higher the test's capacity to identify responsive individuals. Specificity indicates the test's capacity to identify individuals who did not respond to the treatment in question; the higher the value, the higher the chance that the test will identify individuals who are not responsive to the treatment. Positive predictive value refers to the proportion of responsive individuals with positive results, and negative predictive value refers to the proportion of non-responsive individuals with negative results.

Quality assessment

The methodological quality of the included studies was evaluated with the QUADAS-2 tool [24]. This tool is designed to assess the quality of primary diagnostic accuracy studies to rate the risk of bias and concerns regarding applicability [25].

The tool comprises four key domains that discuss bias associated with patient selection, index test, reference standard, flow of patients through the study, timing of the index test, and reference standard. The first three domains are also assessed in terms of concerns regarding applicability. Reviewers are thus able to judge each domain in terms of risk of bias and concerns regarding applicability as ‘Low,’ ‘High,’ or ‘Unclear.’ Specifically, two categories (risk of bias and applicability concerns) were assessed and studies with two or more domains of high risk would be designated as high risk. Those with only one domain of high risk would be designated as medium risk while those with no domains of high risk would be designated as low risk. The validity and reliability of QUADAS-2 has been established previously [24]. In this investigation, QUADAS-2 ratings were conducted independently and in duplicate by two authors (K.O., F.A.).

Results

Description of studies

The search identified 155 articles from the database and by hand-searching relevant reviews [5,26–28]. Fig. 1 presents the flowchart of the study selection process. After excluding irrelevant articles based on title and abstract, 66 studies were retrieved for full-text assessment. Of these, 25 articles were excluded as irrelevant articles, and seven review articles were excluded. There were 17 studies [9,29–44] focused on prediction OA treatment success. However, because these studies did not provide the required outcomes of sensitivity, specificity, positive predictive value, and negative predictive value, they were excluded from our analysis, and only described in Appendix A. The remaining 17 publications [10–22,45–48] were included for detailed analysis.

Table 1 presents the characteristic of the included studies. Out of the 17, 15 studies [10–12,14,15,17–22,45–48] were prospective, 14 studies [10,12–19,22,45–48] used PSG as the reference standard, and three studies [11,20,21] used a level III monitor instead of PSG for the follow-up assessment.

A variety of predictors were used: PSG as the predictive index test in four studies [10–13], cephalographs in two studies [14,15], CPAP pressure in two studies [16,17], overnight PSG with remotely controlled mandibular positioner in two studies [20,21], multisensor catheter in two studies [22,46], nasopharyngeal fiberscope [45], drug-induced sleep endoscopy [19], spirometry [18], and posterior rhinomanometry [47] in one study each, and one study used both body mass index (BMI) and Mallampati score [48]. In addition, two [10,11] out of the 17 studies used the same index test, methodology, and cut-off values. However, different methodologies and cut-off values of the index test were used in all of the other studies.

Quality assessment

According to QUADAS-2, the quality assessment was composed of two categories: risk of bias and applicability concerns, and was described as Low to High (Table 2). In three studies [12,18,46] based on a multisensor catheter, spirometer, and PSG variables, both the risk of bias and concerns with

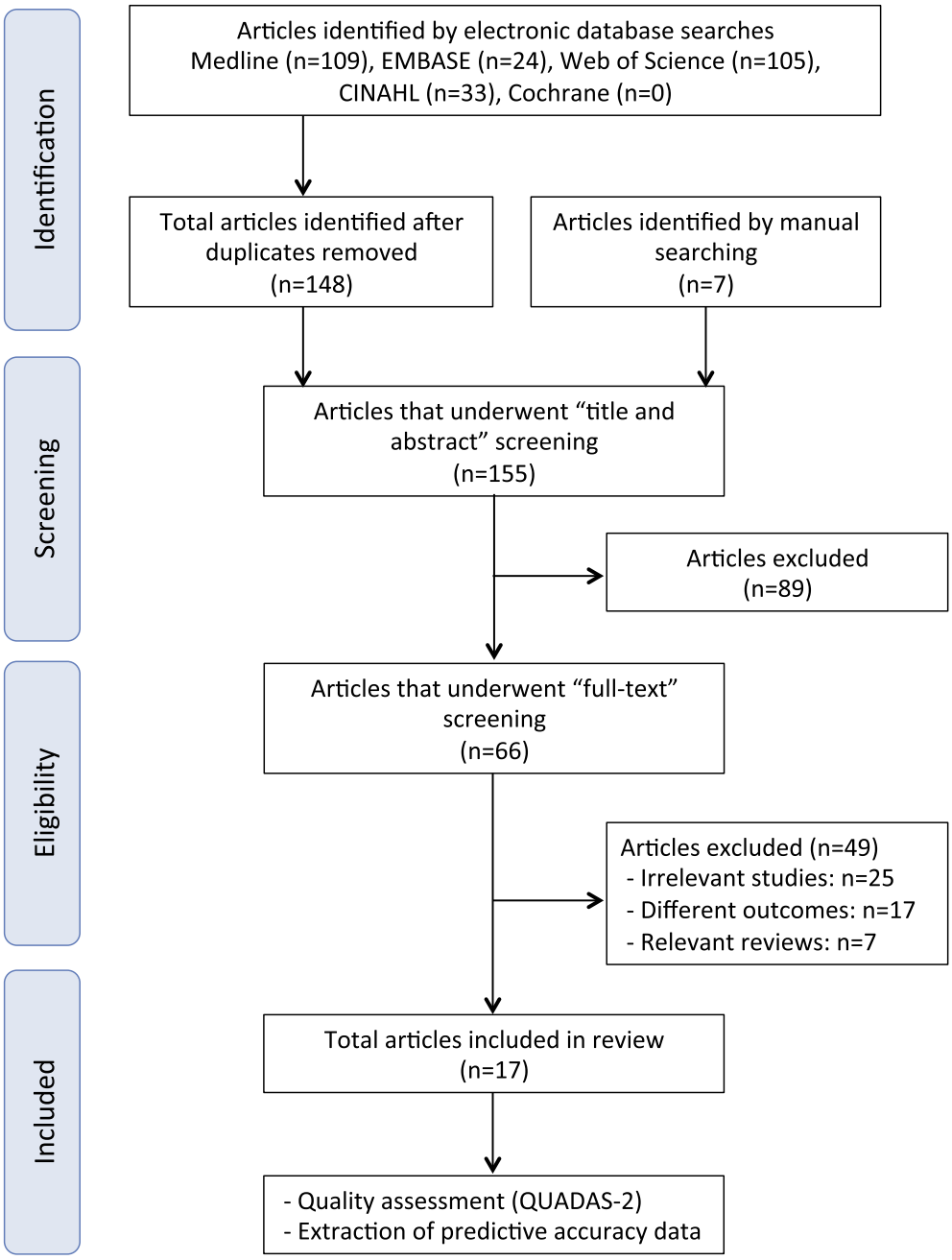

Fig. 1. Flow diagram of article management. CINAHL = cumulative index to nursing and allied health literature; QUADAS = quality assessment of diagnostic accuracy studies; PSG = polysomnography.

applicability were low. In three studies [11,20,21] involving a remotely controlled mandibular positioner and PSG variables, the risk of bias and concerns with applicability were both high. All of the other studies showed a combination of high, medium, and low for bias and applicability.

Predictive accuracy

Table 3 presents the accuracy analysis of the 17 studies included in this review. Because definitions of treatment success in the included studies varied, Table 3 presents the various outcomes (sensitivity, specificity, positive predictive value, and negative predictive value) in the respective success definitions. The most commonly used definition of success was an AHI < 10/h in addition to a > 50% reduction in baseline AHI, which was used in eight [10,12–17,22] out of the 17 studies. Multiple success criteria were used in six studies [10,13,15,17,22,46], and two [15,17] of these six studies presented more than three success criteria as their predictor values. Oxygen desaturation index (ODI) was used by two studies [20,21] instead of AHI as the predictor values because these studies used a level III monitor instead of PSG for the follow-up assessment. Sensitivity ranged from 36 [18] to 100 [16], specificity ranged from 25 [16] to 100 [46], positive predictive value ranged from 38 [48] to 100 [46], and negative predictive value ranged from 33 [13,18] to 100 [16].

Fig. 2 presents relationships between sensitivity and specificity, and between positive predictive value and negative predictive

Table 1
Characteristics of included studies

Authors (year)	Type of study	N	Index test			Type of reference standard
			Type	Methodology	Cut-off value	
Yoshida et al. (2001)[12]	Prospective	72 (AHI >5)	PSG variables	The patients were classified into three groups: supine, lateral, and prone, checked by PSG, according to the position in which the apneas and hypopneas were most frequently observed	Supine dependent OSA	PSG
Tsai et al. (2004)[21]	Prospective	19 (ODI >15)	Overnight PSG with RCMP	During PSG, the mandible was advanced in 1 mm increments until optimal advancement (i.e., elimination of majority of obstructive apneas, hypopneas, and nocturnal oxygen desaturation) was achieved	Optimal advancement <15 mm	Level III monitor (baseline and follow-up)
Ng et al. (2006)[46]	Prospective	12 (AHI >10)	Multisensor catheter	During PSG, the use of multisensor catheters within the upper airway and esophagus were used to identify the primary and the secondary site or sites of airway obstruction in sleep	Primary oropharyngeal closure	PSG
Zeng et al. (2008)[47]	Prospective	38 (AHI >5)	Posterior rhinomanometry	While awake in sitting position in the afternoon, measuring nasal airway resistance by posterior rhinomanometry	Nasal airway resistance ≦6.4 cmH_2O/L/sec	PSG
Chan et al. (2010)[45]	Prospective	35 (AHI >10)	NPF	While awake, assessing changes in the cross-sectional areas of velopharynx, oropharynx, and hypopharynx when the Müller Manoeuvre was performed with mandiblar advancement	An increase of ≧5% in relative terms in the cross-sectional area of the velopharynx	PSG
Chung et al. (2010)[11]	Prospective	72 (AHI >5)	PSG variables	The AHI of supine or lateral sleep position was assessed by PSG	Supine AHI : lateral AHI ≧2	Level III monitor
Tsuiki et al. (2010)[17]	Prospective	35 (AHI>5)	CPAP pressure	Optimal pressure of CPAP was manually determined by registered PSG technologists and targeted to abolish respiratory events such as apnoea, hypoapnoea, and flow limitation	CPAP pressure ≦10.5 cmH_2O	PSG
Bosshard et al. (2011)[22]	Prospective	33 (AHI >5)	Multisensor catheter	While awake, the site of upper airway collapse was identified by examining the flow-pressure relationships of flow-limited twitches when measuring velopharygeal and oropharyngeal pressure with a pressure-tipped catheter	The presence of oropharyngeal collapse	PSG
Chan et al. (2011)[18]	Prospective	35 (AHI >5)	Spirometer	While awake, calculating MIF_{50} and the MEF_{50} : MIF_{50} ratio when the subjects inspired maximally, then exhaled forcibly and continuously into the spirometer until residual volume with maximal effort	MIF_{50} <6 L/s and MEF_{50} : MIF_{50} >0.7	PSG
Holley et al. (2011)[13]	Retrospective	497 (AHI >5)	PSG variables	Assessing the AHI on baseline by PSG	5< AHI <15	PSG
Lee et al. (2012)[10]	Prospective	100 (AHI >5)	PSG variables	The AHI of supine or lateral sleep position was assessed by PSG	Supine AHI : lateral AHI ≧2	PSG
Ng et al. (2012)[15]	Prospective	72 (AHI >10)	Age, gender, cephalogram	While awake in upright position, lateral cephalometric radiographs were taken according to a standardized methodology and a formula was calculated as follows: The probability (P) of OA treatment success = [exp(-2.594-0102×age-0.180×PmP-0.118×BaSN)]/[1+exp(-2.594-0102×age-0.180×PmP-0.118×BaSN)]	The probability (P) of formula > 0.50	PSG
Shen et al. (2012)[14]	Prospective	52 (AHI >10)	Cephalogram	While awake in upright position, lateral cephalometric radiographs were performed according to a standardized methodology and a formula was calculated as follows: The probability (P) of OA treatment success = [exp(56.143-0.180×AFH-0.379×SNB-0.329×MinRGA)]/[1+exp(56.143-0.180×AFH-0.379×SNB-0.329×MinRGA)]	The probability (P) of formula > 0.50	PSG
Remmers et al. (2013)[20]	Prospective	67 (ODI >10)	Overnight PSG with RCMP	During PSG, stepwise protrusion of the mandible was continued over the range of mandiblar protrusion until respiratory events were unequivocally eliminated in REM and NREM sleep in both the supine and lateral decubitus positions, or until maximum protrusion was reached	AHI ≦1 per 5 min of REM	Level III monitor (baseline and follow-up)
Tsuiki et al. (2013)[48]	Prospective	95 (15< AHI <30)	BMI, Mallampati score	Assessment of body mass index and Mallampati score	BMI >24 kg/m_2 with Mallampati score class 1 to 3	PSG
Vroegop et al. (2013)[19]	Prospective	103 (AHI >5)	DISE	During artificial sleep with sedative drugs, assessing residual collapse at upper airway, including palate, oropharynx, tongue base and hypopharynx level, using the simulation bite	No residual collapse at any upper airway level	PSG
Sutherland et al. (2014)[16]	Retrospective	78 (AHI >10)	CPAP pressure	Therapeutic CPAP pressure was determined by the 95^{th} percentile pressure from usage exceeding 4 hours by overnight PSG on CPAP treatment	CPAP pressure <13 cmH_2O	PSG

AFH = anterior face height; AHI = apnea hypopnea index; BaSN = cranial base angulation; BMI = body mass index; CPAP = continuous positive airway pressure; DISE = drug-induced sleep endoscopy; MEF = maximum expiratory force; MIF = maximum inspiratory force; MinRGA = minimal retroglossal airway; NPF = nasopharyngeal fiberscope; NREM = non rapid eye movement; ODI = oxygen desaturation index; OSA = obstructive sleep apnea; PmP = soft palate length measured from pterygomaxillare to the tip of the soft palate; PSG = polysomnograhy; RCMP = remotely controlled mandibular positioner; REM = rapid eye movement; SNB = angle of sella, nasion, B point.

Table 2
Quality assessment score (QUADAS-2)

	Risk of bias					Applicability concerns			
Author (year)	Patient selection	Index test	Reference standard	Flow and timing	Total	Patient selection	Index test	Reference standard	Total
Yoshida et al. (2001)[12]	Unclear	Low	Unclear	Low	Low	Low	Low	Low	Low
Tsai et al. (2004)[21]	Unclear	Low	High[d)]	High[e)]	High	High[d)]	Low	High[d)]	High
Ng et al. (2006)[46]	Unclear	Low	Low	Low	Low	Low	Low	Low	Low
Zeng et al. (2008)[47]	Unclear	High[c)]	Low	Low	Medium	Low	Low	Low	Low
Chan et al. (2010)[45]	Unclear	High[c)]	Low	Low	Medium	Low	Low	Low	Low
Chung et al. (2010)[11]	Low	Low	High[d)]	High[d)]	High	High[d)]	Low	High[d)]	High
Tsuiki et al. (2010)[17]	Low	High[c)]	Unclear	High[e)]	High	Low	Low	Low	Low
Bosshard et al. (2011)[22]	Low	Low	Low	High[e)]	Medium	Low	Low	Low	Low
Chan et al. (2011)[18]	Unclear	Low	Unclear	Low	Low	Low	Low	Low	Low
Holley et al. (2011)[13]	High[a)]	Low	Unclear	High[e)]	High	Low	Low	Low	Low
Lee et al. (2012)[10]	High[b)]	Low	Unclear	Low	Medium	Low	Low	Low	Low
Ng et al. (2012)[15]	Low	High[c)]	Low	Low	Medium	Low	Low	Low	Low
Shen et al. (2012)[14]	Low	High[c)]	Unclear	Low	Medium	Low	Low	Low	Low
Remmers et al. (2013)[20]	Low	Low	High[d)]	High[d)]	High	High[d)]	Low	High[d)]	High
Tsuiki et al. (2013)[48]	Low	High[c)]	Unclear	Low	Medium	Low	Low	Low	Low
Vroegop et al. (2013)[19]	Unclear	Low	Low	High[f)]	Medium	High[f)]	Low	Low	Medium
Sutherland et al. (2014)[16]	High[a)]	High[c)]	Low	Low	High	Low	Low	Low	Low

PSG = polysomnograhy; QUADAS = quality assessment of diagnostic accuracy studies.
[a)] Retrospective study
[b)] Excluded patients who did not undergo follow-up PSG because of economic reasons
[c)] A threshold was not pre-specified
[d)] Reference standard was level III monitor
[e)] All patients did not receive follow-up PSG
[f)] Patients who were assessed as not suitable by index test did not start oral appliance treatment

value in each study. If there were multiple success criteria reported by the study, we used AHI < 10/h in addition to a > 50% reduction in baseline AHI, the most commonly adopted definition, as the representative value.

Sensitivity and specificity

With regard to the relationships between sensitivity and specificity, the remotely controlled mandibular positioner study [20], the nasopharyngeal fiberscope study [45], and the multisensor catheter study [46] showed higher combinations of sensitivity and specificity. In terms of sensitivity alone, the CPAP pressure study [16] and the PSG variable study [10] showed high sensitivity but low specificity, thus being the best methods to identify good OA responders. However, in terms of specificity alone, the PSG variable study [13] and the spirometer study [18] reported high specificity but low sensitivity, being the best methods to identify non-responders.

Positive predictive value and negative predictive value

In terms of positive predictive value and negative predictive value, the remotely controlled mandibular positioner study [20], nasopharyngeal fiberscope study [45], and multisensor catheter study [46] showed higher combinations of positive and negative predictive values. The BMI and Mallampati study [47] showed a high negative predictive value but a low positive predictive value and the PSG variable study [13] and the spirometer study [18] showed a high positive predictive value but a low negative predictive value.

Discussion

In this review, 17 studies were evaluated for outcomes to assess methods of predicting OA treatment success. The sensitivity, specificity, positive predictive value, and negative predictive value showed a wide variability and ranged from 36 to 100%, 25 to 100%, 38 to 100%, and 33 to 100%, respectively, and varied depending on definitions of treatment success and specific methods of prediction used. The wide variability of the results makes it difficult to delineate their usefulness in routine clinical practice. Importantly, when results are described using different criteria of treatment success, it allows clinicians to utilize different criteria depending on their patient-specific treatment goal. For some patients, the treatment goal could be set to an AHI < 5/h in addition to a > 50% reduction in baseline AHI, which is considered to be the most strict standard. For others, the target value could be set at an AHI < 10/h in addition to a > 50% reduction in baseline AHI, reflecting what usually happens in clinical practice. Therefore the clinical relevance is increased when prediction values for multiple criteria of success are reported, as in the studies by Ng et al. [15] and Tsuiki et al. [17]. Similarly, the predictive index tests employed in the included studies varied widely in terms of both simplicity and impact on the patient. Cephalometrics [14,15] and nasopharyngeal fiberscopy [45] are invasive prediction methods. Remotely controlled mandibular positioners [20,21], multisensor catheters [22,46], and drug-induced sleep endoscopy [19] are complex, costly, and laborious approaches. Meanwhile, spirometry [18] and rhinomanometry [47] are simple and less invasive methods. For patients diagnosed by PSG and prescribed CPAP, PSG variables [10–13] and CPAP pressure [16,17] are already available at the time of treatment decision-making and could be used to predict treatment success. As CPAP pressure is related to upper airway collapsibility, it may also allow for improved understanding of how this influences OA treatment response. CPAP pressure may represent a simple predictor and be clinically useful in the many patients who have failed or are non-adherent to CPAP. However this method is of course restricted to patients who have used CPAP. Additionally, predictive index tests obtained during baseline clinical evaluation such as BMI and Mallampati scores [48] are very useful as predictors of treatment success in clinical practice.

The relationship between sensitivity and specificity, and between positive predictive value and negative predictive value in

Table 3
Summary of findings

Authors (year)	N	Type of index test	Findings according to the definition of success											
			AHI <5 and 50% reduction		AHI <10 and 50% reduction		AHI <20 and 50% reduction		AHI 50% reduction		ODI <10 and 50% reduction		ODI <15 and 50% reduction	
			Se/Sp	PPV/ NPV	Se/Sp	PPV/ NPV	Se/Sp	PPV/ NPV	Se/Sp	PPV/ NPV	Se/Sp	PPV/ NPV	Se/Sp	PPV/ NPV
Yoshida et al. (2001)[12]	72	PSG variables			69/49	61/57								
Tsai et al. (2004)[21]	19	Overnight PSG with RCMP											60/89	86/67
Ng et al. (2006)[46]	12	Multisensor catheter	80/100	100/80					57/100	100/63				
Zeng et al. (2008)[47]	38	Posterior rhinomanometry							65/75	85/50				
Chan et al. (2010)[45]	35	NPF							83/77	79/81				
Chung et al. (2010)[11]	72	PSG variables	69/57	69/57										
Tsuiki et al. (2010)[17]	35	CPAP pressure	90/56	45/93	86/62	60/87			73/69	80/60				
Bosshard et al. (2011)[22]	33	Multisensor catheter			79/60	65/75			78/73	83/67				
Chan et al. (2011)[18]	35	Spirometer							36/80	82/33				
Holley et al. (2011)[13]	497	PSG variables	48/75	70/55	43/80	86/33								
Lee et al. (2012)[10]	100	PSG variables			89/29	58/70	84/31	78/40						
Ng et al. (2012)[15]	72	Age, gender, cephalogram	61/82	68/77	74/70	74/70	71/55	66/64	78/56	69/67				
Shen et al. (2012)[14]	52	Cephalogram			96/72	73/96								
Remmers et al. (2013)[20]	67	Overnight PSG with RCMP									86/92	94/83		
Tsuiki et al. (2013)[48]	53	BMI, Mallampati score	85/55	38/92										
Vroegop et al. (2013)[19]	103	DISE							49/78	83/40				
Sutherland et al. (2014)[16]	78	CPAP pressure			100/25	75/100								

AHI = apnea hypopnea index; BMI = body mass index; CPAP = continuous positive airway pressure; DISE = drug-induced sleep endoscopy; NPF = nasopharyngeal fiberscope; NPV = negative predictive value; ODI = oxygen desaturation index; PPV = positive predictive value; PSG = polysomnography; RCMP = remotely controlled mandibular positioner; Se = sensitivity; Sp = specificity.

each study, helps in OA treatment decision-making in clinical practice. The remotely controlled mandibular positioner study [20], the nasopharyngeal fiberscope study [45], and the multisensor catheter study [46] showed high sensitivity, specificity, and positive and negative predictive values, but these index tests are invasive or require a highly technical method and laborious approach, and are costly. While a high positive predictive value could be a predictor of OA treatment success, equally important is that a high negative predictive value could be an indicator of OA treatment failure. The BMI and Mallampati study [47], the cephalometric study [14], and the CPAP pressure study [16] showed that a high negative predictive value may be a predictor of OA treatment failure. Depending on the data available to the clinicians and the treatment goals, ideally a combination of different prediction methods may better help to predict OA treatment success.

Several studies concerned with the prediction of OA treatment success have been excluded from this review for not presenting one or more of the required outcomes of sensitivity, specificity, positive predictive value, or negative predictive value, as seen in Appendix A. Some studies compared OA treatment responders and non-responders in terms of age, AHI, neck circumference, and BMI and showed that a younger age [31,32,42], lower AHI [38], smaller neck circumference [9], and lower BMI [32,37] were identified as indicators of treatment success. Some cephalometric characteristics, including a shorter soft palate, larger retropalatal airway space, lower hyoid bone position, narrow angle of sella-nasion-B point, and higher angle of sella-nasion-A point have also been associated with a favorable OA treatment response [31,33,35,36,39,40,42]. However, although these studies showed statistically significant differences between responders and non-responders, they did not present their cut-off value and outcomes including sensitivity, specificity, and positive and negative predictive values. Therefore, these studies were not included in the detailed analysis and comparisons.

In some studies [9,29,30], multiple regression analysis allowed the formulation of an equation for predicting the AHI after OA treatment. Neck circumference, baseline AHI, retropalatal airway space, and some cephalometric characteristics (angle of sella-nasion-A point, hyoid-to-mandibular plane distance, posterior facial height) were identified as predictors of follow-up AHI. In another study [41], the data obtained from computational fluid dynamics by magnetic resonance was found to have the strongest relationship with the AHI change (ΔAHI%) after OA treatment. However, these studies were concerned with the prediction of the AHI after OA treatment or AHI changes after OA treatment and not with the prediction of OA treatment success (e.g., an AHI < 10/h in addition to a > 50% reduction in baseline AHI). Two studies [34,37] evaluated the predictive value of variables for OA treatment success with logistic regression analysis. Baseline AHI, female gender, angle of sella-nasion-A point, angle of A point-nasion-B point, and maximum mandibular advancement were identified as predictors for OA treatment success. Although these studies showed odds ratios associated with each predictor value, they did not present outcomes, including sensitivity, specificity, and positive and negative predictive values. Therefore, they were not included in this current review.

The validity and reliability of QUADAS-2 has been described as high, and is commonly used in systematic reviews of diagnostic tests. As per our protocol, the QUADAS-2 tool was applied to all included articles independently by two authors. When later compared, no discrepancies between the scoring were found. According to QUADAS-2, the quality of included studies for prediction falls into two categories: risk of bias and concerns of applicability. The results showed that both the concerns of applicability and the

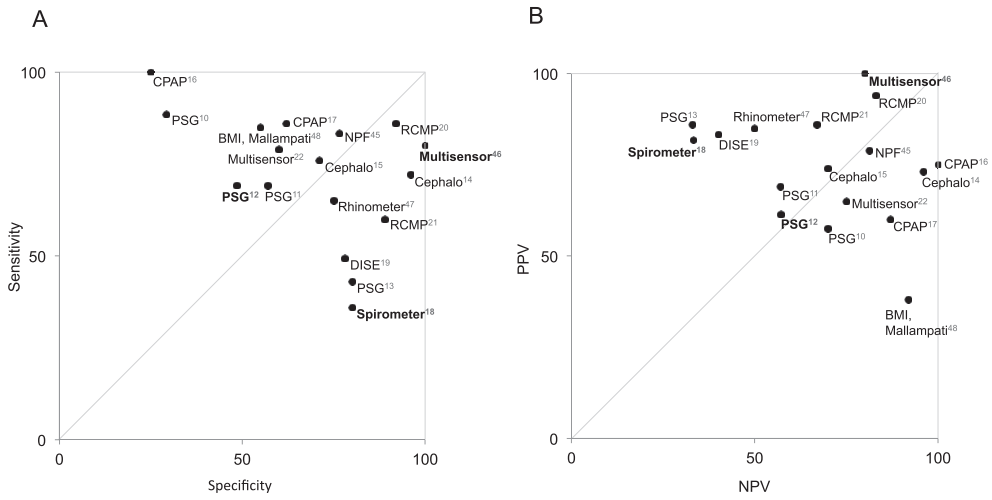

Fig. 2. A: The relationships between sensitivity and specificity in each study. B: The relationships between positive predictive value and negative predictive value in each study. Bold typeface indicates studies with low risk of bias and low applicability concerns. Responder criteria were defined as follows: AHI < 5, 50% reduction; BMI and Mallampati,[48] multisensor,[46] PSG.[11] AHI < 10, 50% reduction; CPAP,[16] Cephalo,[14] PSG,[10] Cephalo,[15] multisensor,[22] CPAP,[17] PSG,[12] PSG.[13] 50% AHI reduction; DISE,[19] Spirometer,[18] NPF,[45] Rhinometer.[47] ODI < 10, 50% reduction; RCMP.[20] ODI < 15, 50% reduction; RCMP[21]. AHI = apnea hypopnea index; BMI = body mass index; Cephalo = cephalogram; CPAP = continuous positive airway pressure; DISE = drug-induced sleep endoscopy; NPF = nasopharyngeal fiberscope; ODI = oxygen desaturation index; PSG = polysomnography; RCMP = remotely controlled mandibular positioner.

risk of bias were low to high in the studies assessed by this review. Remmers et al. [18] and Tsai et al. [19] used the remotely controlled mandibular positioner as a predictive index test and were considered to have a high risk of bias and high applicability concerns. The remotely controlled mandibular positioner method consists of stepwise protrusion of the mandible until improvement of respiratory events is achieved or maximum protrusion is reached while the patient is undergoing PSG. However, in this review, these studies were considered to be at high risk since level III monitors were used to assess OSA with the OA instead of PSG. Chung et al. [11] also used level III monitors, while all other studies used PSG evaluations. It is important to note that all included studies were derivation studies and as such lack prospective validation of their methods. It has been shown that a derivation study hypothesis is often not confirmed in validation studies and therefore the results of this review should be interpreted with some reservation. Future trials should prospectively validate methods for OA treatment prediction.

There were some limitations in this systematic review. Meta-analyses were not undertaken because of heterogeneity in terms of the different index tests used for prediction, which varied in methodology and cut-off values. Additionally, inconsistency in the definition of treatment responses described as successful made it difficult to undertake a meta-analysis. The lack of consensus around the definition of a successful treatment outcome has led to this limitation. Based on this limitation, we would like to emphasize that there is an important need for the field to develop uniform reporting standards in order to allow more sophisticated meta-analysis. To this end, AHI as well as ODI 4% and ODI 3% are widely accepted outcome variables that could easily be reported in future studies. Future work should also explore the use of a combination of prediction methods given that the results of single assessment tools are not overly predictive.

Conclusions

Although many clinicians use PSG data as their main assessment tool to recommend an OA, studies using PSG variables have shown lower predictive accuracy. The studies using a remotely controlled mandibular positioner and multisensor catheters showed high predictive accuracy but required a highly technical method and a laborious approach. In terms of clinical techniques, the nasopharyngeal fiberscope has shown the best combination of predictive accuracy and quality.

Based on this systematic review of the literature on the prediction of OA treatment responses, it can be concluded that predictive index tests may be very useful in clinical practice for physicians and dentists concerned with whether or not OA therapy has high chances of success, especially in moderate to severe OSA patients. Using the available information, the chosen method to predict OA treatment response can be based on the clinician's goals and requirements for treatment success.

Practice points

1. The predictive accuracy showed a wide variability that depends on the definitions of treatment success and specific method of prediction. Although many clinicians use PSG data as their main assessment to recommend an OA, such variables have shown lower predictive accuracy.
2. The use of a variety of definitions of OA treatment success was very useful in clinical practice depending on the clinician's goals and requirements for treatment success.
3. As for clinical techniques, the nasopharyngeal fiberscope study, which has shown high predictive accuracy, medium risk of bias, and low concerns of applicability, was the best combination of predictive accuracy and quality.
4. Studies using a remotely controlled mandibular positioner (high risk of bias and high concerns of applicability) and multisensor catheters (low risk of bias and low concerns of applicability) showed high predictive accuracy, but required a highly technical method and detailed protocol.

Research agenda

Future studies evaluating the prediction of OA treatment outcomes in OSA should focus on:

1. Providing a variety of definitions for OA treatment success and using PSG as the reference standard.
2. Identifying an index test with high predictive accuracy, that is simple, non-invasive, and useful in clinical practice.

Conflict of interests

The authors indicated no potential conflicts of interest.

Acknowledgements

We thank Mrs. Ingrid Ellis for her editorial assistance.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.smrv.2015.11.007.

References

[1] Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 1993;328(17):1230–5.
[2] Flemons WW. Clinical practice. Obstructive sleep apnea. N Engl J Med 2002;347(7):498–504.
[3] Lévy P, Pépin JL, McNicholas WT. Should all sleep apnoea patients be treated? Yes Sleep Med Rev 2002;6(1):17–26. discussion 27.
[4] Kribbs NB, Pack AI, Kline LR, Smith PL, Schwartz AR, Schubert NM, et al. Objective measurement of patterns of nasal CPAP use by patients with obstructive sleep apnea. Am Rev Respir Dis 1993;147(4):887–95.
[5] Ramar K, Dort LC, Katz SG, Lettieri CJ, Harrod CG, Thomas SM, et al. Clinical practice guideline for the treatment of obstructive sleep apnea and snoring with oral appliance therapy: an update for 2015. J Clin Sleep Med 2015;11(7):773–827.
[6] Sutherland K, Vanderveken OM, Tsuda H, Marklund M, Gagnadoux F, Kushida CA, et al. Oral appliance treatment for obstructive sleep apnea: an update. J Clin Sleep Med 2014;10(2):215–27.
[7] Johal A, Battagel JM, Kotecha BT. Sleep nasendoscopy: a diagnostic tool for predicting treatment success with mandibular advancement splints in obstructive sleep apnoea. Eur J Orthod 2005;27(6):607–14.
[8] Gotsopoulos H, Chen C, Qian J, Cistulli PA. Oral appliance therapy improves symptoms in obstructive sleep apnea: a randomized, controlled trial. Am J Respir Crit Care Med 2002;166(5):743–8.
[9] Mehta A, Qian J, Petocz P, Darendeliler MA, Cistulli PA. A randomized, controlled study of a mandibular advancement splint for obstructive sleep apnea. Am J Respir Crit Care Med 2001;163(6):1457–61.
*[10] Lee CH, Jung HJ, Lee WH, Rhee CS, Yoon IY, Yun PY, et al. The effect of positional dependency on outcomes of treatment with a mandibular advancement device. Arch Otolaryngol Head Neck Surg 2012;138(5):479–83.
[11] Chung JW, Enciso R, Levendowski DJ, Morgan TD, Westbrook PR, Clark GT. Treatment outcomes of mandibular advancement devices in positional and nonpositional OSA patients. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2010;109(5):724–31.
*[12] Yoshida K. Influence of sleep posture on response to oral appliance therapy for sleep apnea syndrome. Sleep 2001;24(5):538–44.
*[13] Holley AB, Lettieri CJ, Shah AA. Efficacy of an adjustable oral appliance and comparison with continuous positive airway pressure for the treatment of obstructive sleep apnea syndrome. Chest 2011;140(6):1511–6.
[14] Shen HL, Wen YW, Chen NH, Liao YF. Craniofacial morphologic predictors of oral appliance outcomes in patients with obstructive sleep apnea. J Am Dent Assoc 2012;143(11):1209–17.
*[15] Ng AT, Darendeliler MA, Petocz P, Cistulli PA. Cephalometry and prediction of oral appliance treatment outcome. Sleep Breath 2012;16(1):47–58.
[16] Sutherland K, Phillips CL, Davies A, Srinivasan VK, Dalci O, Yee BJ, et al. CPAP pressure for prediction of oral appliance treatment response in obstructive sleep apnea. J Clin Sleep Med 2014;10(9):943–9.
*[17] Tsuiki S, Kobayashi M, Namba K, Oka Y, Komada Y, Kagimura T, et al. Optimal positive airway pressure predicts oral appliance response to sleep apnoea. Eur Respir J 2010;35(5):1098–105.
*[18] Chan AS, Lee RW, Srinivasan VK, Darendeliler MA, Cistulli PA. Use of flow-volume curves to predict oral appliance treatment outcome in obstructive sleep apnea: a prospective validation study. Sleep Breath 2011;15(2):157–62.
[19] Vroegop AV, Vanderveken OM, Dieltjens M, Wouters K, Saldien V, Braem MJ, et al. Sleep endoscopy with simulation bite for prediction of oral appliance treatment outcome. J Sleep Res 2013;22(3):348–55.
*[20] Remmers J, Charkhandeh S, Grosse J, Topor Z, Brant R, Santosham P, et al. Remotely controlled mandibular protrusion during sleep predicts therapeutic success with oral appliances in patients with obstructive sleep apnea. Sleep 2013;36(10):1517–25. 1525A.
[21] Tsai WH, Vazquez JC, Oshima T, Dort L, Roycroft B, Lowe AA, et al. Remotely controlled mandibular positioner predicts efficacy of oral appliances in sleep apnea. Am J Respir Crit Care Med 2004;170(4):366–70.
*[22] Bosshard V, Masse JF, Sériès F. Prediction of oral appliance efficiency in patients with apnoea using phrenic nerve stimulation while awake. Thorax 2011;66(3):220–5.
[23] Deane SA, Cistulli PA, Ng AT, Zeng B, Petocz P, Darendeliler MA. Comparison of mandibular advancement splint and tongue stabilizing device in obstructive sleep apnea: a randomized controlled trial. Sleep 2009;32(5):648–53.
[24] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2 group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529–36.
[25] Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.
[26] Sutherland K, Cistulli P. Mandibular advancement splints for the treatment of sleep apnea syndrome. Swiss Med Wkly 2011;141:w13276.
[27] Pliska BT, Almeida F. Effectiveness and outcome of oral appliance therapy. Dent Clin North Am 2012;56(2):433–44.
[28] Almeida FR, Lowe AA. Principles of oral appliance therapy for the management of snoring and sleep disordered breathing. Oral Maxillofac Surg Clin North Am 2009;21(4):413–20.
[29] Eveloff SE, Rosenberg CL, Carlisle CC, Millman RP. Efficacy of a herbst mandibular advancement device in obstructive sleep apnea. Am J Respir Crit Care Med 1994;149(4 Pt 1):905–9.
[30] Mayer G, Meier-Ewert K. Cephalometric predictors for orthopaedic mandibular advancement in obstructive sleep apnoea. Eur J Orthod 1995;17(1):35–43.
[31] Liu Y, Park YC, Lowe AA, Fleetham JA. Supine cephalometric analyses of an adjustable oral appliance used in the treatment of obstructive sleep apnea. Sleep Breath 2000;4(2):59–66.
[32] Liu Y, Lowe AA, Fleetham JA, Park YC. Cephalometric and physiologic predictors of the efficacy of an adjustable oral appliance for treating obstructive sleep apnea. Am J Orthod Dentofacial Orthop 2001;120(6):639–47.
[33] Skinner MA, Robertson CJ, Kingshott RN, Jones DR, Taylor DR. The efficacy of a mandibular advancement splint in relation to cephalometric variables. Sleep Breath 2002 Sep;6(3):115–24.
[34] Marklund M, Stenlund H, Franklin KA. Mandibular advancement devices in 630 men and women with obstructive sleep apnea and snoring: tolerability and predictors of treatment success. Chest 2004;125(4):1270–8.
[35] Horiuchi A, Suzuki M, Ookubo M, Ikeda K, Mitani H, Sugawara J. Measurement techniques predicting the effectiveness of an oral appliance for obstructive sleep apnea hypopnea syndrome. Angle Orthod 2005;75(6):1003–11.
[36] Otsuka R, Almeida FR, Lowe AA, Ryan F. A comparison of responders and nonresponders to oral appliance therapy for the treatment of obstructive sleep apnea. Am J Orthod Dentofac Orthop 2006;129(2):222–9.
[37] Hoekema A, Doff MH, de Bont LG, van der Hoeven JH, Wijkstra PJ, Pasma HR, et al. Predictors of obstructive sleep apnea-hypopnea treatment outcome. J Dent Res 2007;86(12):1181–6.
[38] Petri N, Svanholt P, Solow B, Wildschiodtz G, Winkel P. Mandibular advancement appliance for obstructive sleep apnoea: results of a randomised placebo controlled trial using parallel group design. J Sleep Res 2008;17:221–9.
[39] Doff MH, Hoekema A, Pruim GJ, van der Hoeven JH, de Bont LG, Stegenga B. Effects of a mandibular advancement device on the upper airway morphology: a cephalometric analysis. J Oral Rehabil 2009;36(5):330–7.
[40] Lee CH, Kim JW, Lee HJ, Seo BS, Yun PY, Kim DY, et al. Determinants of treatment outcome after use of the mandibular advancement device in patients with obstructive sleep apnea. Arch Otolaryngol Head Neck Surg 2010;136(7):677–81.
[41] Zhao M, Barber T, Cistulli P, Sutherland K, Rosengarten G. Computational fluid dynamics for the assessment of upper airway response to oral appliance treatment in obstructive sleep apnea. J Biomech 2013;46:142–50.

* The most important references are denoted by an asterisk.

[42] Milano F, Billi MC, Marra F, Sorrenti G, Gracco A, Bonetti GA. Factors associated with the efficacy of mandibular advancing device treatment in adult OSA patients. Int Orthod 2013;11(3):278–89.
[43] Sasao Y, Nohara K, Okuno K, Nakamura Y, Sakai T. Videoendoscopic diagnosis for predicting the response to oral appliance therapy in severe obstructive sleep apnea. Sleep Breath 2014;18(4):809–15.
[44] Hein H. CPAP as a predictor of the efficacy of mandibular advancement devices. Pneumologie 2015;69(1):13–6.
*[45] Chan AS, Lee RW, Srinivasan VK, Darendeliler MA, Grunstein RR, Cistulli PA. Nasopharyngoscopic evaluation of oral appliance therapy for obstructive sleep apnoea. Eur Respir J 2010;35(4):836–42.
*[46] Ng AT, Qian J, Cistulli PA. Oropharyngeal collapse predicts treatment response with oral appliance therapy in obstructive sleep apnea. Sleep 2006;29(5):666–71.
[47] Zeng B, Ng AT, Qian J, Petocz P, Darendeliler MA, Cistulli PA. Influence of nasal resistance on oral appliance treatment outcome in obstructive sleep apnea. Sleep 2008;31(4):543–7.
[48] Tsuiki S, Ito E, Isono S, Ryan CF, Komada Y, Matsuura M, et al. Oropharyngeal crowding and obesity as predictors of oral appliance treatment response to moderate obstructive sleep apnea. Chest 2013;144(2):558–63.