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