The Laryngoscope

Systematic Review

Outcome Measurements in Obstructive Sleep Apnea: Beyond the Apnea-Hypopnea Index

Samantha Tam, MD; B. Tucker Woodson, MD, FACS; Brian Rotenberg, MD, MPH

Objectives/Hypothesis: The apnea-hypopnea index (AHI) is overwhelmingly used as the main therapeutic metric in the assessment of obstructive sleep apnea (OSA) in surgical studies. However, using AHI as the sole measure is problematic. This study investigates the utility of other outcome measures for patients with OSA undergoing surgery.

Study Design: Systematic review of cohort and review studies.

Methods: A review was performed using the PubMed database. English articles focusing on outcome measures in adults with OSA were included. Studies in pediatric populations, those combining obstructing and central sleep apnea, and those without the use of outcome measures were excluded. Articles were categorized according to level of evidence. The Downs and Black scale and AMSTAR scale were used to assess quality.

Results: Of a total of 10,454 retrieved articles, 21 studies met inclusion and exclusion criteria. Most articles related to continuous positive airway pressure outcomes. Many categories of outcome measures were found: general quality of life, OSA-specific quality of life, measurements of sleepiness, performance, and physiological. Subjects with OSA scored differently in measurement tools in all categories compared to control populations or after treatment, and generally a poor correlation with AHI was seen.

Conclusions: The literature shows a range of tools based on symptoms and physiology of OSA that can assess effects of treatment. Assessment of surgical treatment for OSA should neither be limited to AHI as an outcome, nor should this be the only outcome stressed.

Key Words: Obstructive sleep apnea; outcome measures; quality of life; cognitive outcomes; physiological outcomes.

Level of Evidence: NA.

***Laryngoscope*, 124:337–343, 2014**

INTRODUCTION

Obstructive sleep apnea (OSA) is a common illness affecting 9% of middle-aged men in North America and 3% of women.[1] Untreated OSA can have numerous negative effects on sufferers including fatigue, somnolence, headaches, cardiovascular disease, decreased quality of life (QOL), and increased risk of motor vehicle accidents.[1–3] Multiple treatment options are available including weight loss, continuous positive airway pressure (CPAP), oral appliance therapy, and a wide variety of surgical procedures.[4]

Gold standard assessment for OSA is overwhelmingly cited as polysomnography (PSG).[5] During PSG, the frequency of obstructive events is reported as the apnea-hypopnea index (AHI), and the effectiveness of surgical treatments for OSA is almost exclusively based on reported changes in AHI. However, there is a disconnect between the levels of AHI used to denote outcomes of therapy and real-world clinical outcomes such as QOL, patient perception of disease, cardiovascular measures, or survival. Moreover the definitions of apnea, hypopnea, and AHI cutoff categories, although reflective of perturbations in the range of sleep physiology, are still essentially human constructs without objective physiological basis.[1–3] Thus, parameters with which OSA is measured are different from other chronic diseases, such as anemia or diabetes, where truly objective physiological measurements are made.

Given the multifaceted nature of OSA, therapeutic assessments should reflect variability in presentation.[6] Moreover, patient-centered outcome measures and QOL assessments are gaining substantial traction as priority items to assess when gauging effect of therapy for all manner of diseases. QOL, daytime sleepiness, personal performance, and various physiological measurements have been proposed as proxy markers for OSA, but have

From the Department of Otolaryngology–Head and Neck Surgery (S.T., B.R.), Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada; and the Department of Otolaryngology and Human Communication (B.T.W.), Medical College of Wisconsin, Milwaukee, Wisconsin, U.S.A.

Send correspondence to Brian Rotenberg, MD, St. Joseph's Healthcare Centre, 268 Grosvenor St., London, Ontario, Canada N6A 4V2.
E-mail: brian.rotenberg@sjhc.london.on.ca

This Manuscript was received March 28, 2013, revised May 23, 2013, and accepted for publication June 4, 2013.

Presented as a podium presentation at the 116th Triological Society Annual Meeting at COSM, Orlando, Florida, U.S.A., April 13, 2013.

This study was funded by a grant from the International Sleep Surgical Society.

The authors have no other funding, financial relationships, or conflicts of interest to disclose.

DOI: 10.1002/lary.24275

not been widely adopted. Outcomes research for surgical OSA therapy continues to depend heavily on AHI to validate treatment effectiveness. However, AHI is known to correlate poorly with both patient perception of their OSA as well as many other measures of disease burden.[1–3] Despite this, AHI remains paradoxically persistent as the main, and frequently only, outcome reported in the vast majority of surgical OSA studies. Our current study aimed to investigate the role of non-AHI measurements of OSA and report on their utility in the management of the disease process.

MATERIALS AND METHODS

This study was a systematic review of the literature from 1990 to 2012 using the PubMed database. Search terms were "sleep apnea," "sleep apnea outcomes," "outcomes for sleep apnea," "apnea hypopnea," "apnea hypopnea index," "Epworth Sleepiness Scale," "Epworth Sleepiness Score," "daytime sleepiness," "daytime sleepiness scale," "daytime sleepiness score," "blood pressure and sleep apnea," "hypertension and sleep apnea," "cortisol and sleep apnea," "hormones and sleep apnea," "adrenal and sleep apnea." All abstracts were reviewed for inclusion into the study. Review or cohort articles focusing on identified therapeutic outcome measures in adults with OSA were included. Exclusion criteria were: 1) nonreview or cohort articles, 2) non-English language publications, 3) studies involving pediatric populations, 4) articles combining assessment of obstructive and central sleep apnea or referring to central sleep apnea alone, and 5) studies without defined outcome measures. Studies that met the criteria were then fully retrieved, and their bibliographies were reviewed to identify any further articles not identified in the initial search.

Quality Assessment

Each article was assigned a quality ranking, using levels of evidence as described by Shin et al.[7] Two scales were selected for quality assessment: a modified Downs and Black (DB) Scale and the assessment of multiple systematic reviews (AMSTAR) scale.[8,9] The DB scale is a 26-item tool, chosen due to its robustness in evaluating nonrandomized methodologies, including cohort and case-control studies.[8] The AMSTAR scale is a validated and widely used tool to assess systematic reviews consisting of 11 items.[9] Clinical trials were evaluated with the DB scale, whereas reviews were evaluated using the AMSTAR scale. If neither scale was appropriate for a study, the authors met as a group to make a decision on a case by case basis. The authors completed evaluation of the studies while blinded to each other's ratings. Final scores were compared using the Spearman ρ coefficient of inter-rater reliability. Inter-rater reliability above 0.7 was considered acceptable a priori, with an α of $<.05$ set as statistically significant.

RESULTS

A total of 10,454 abstracts were retrieved and reviewed. A statistician performed the initial screening, with the authorship team then reviewing a short list and ultimately identifying 21 studies that met all inclusion and exclusion criteria. All included studies were published between 1997 and 2012. Overall, the inter-rater reliability from quality assessment was calculated to be 0.86 ($P < .005$), indicating good agreement between reviewers. A summary of these studies can be found in Table I. Most of the retrieved articles related to reviews of measurements of outcomes for CPAP therapy, because few surgical trials have been performed that reported on non-AHI outcomes.

TABLE I.
Summary of Included Articles.

Article	Level of Evidence	DB Scale (n/27)	AMSTAR (n/11)
Akpinar ME et al.[25]	4	19	—
American Thoracic Society and American Sleep Disorders Association[24]	5	—	—
Caples SM et al.[19]	2	—	10.5
Douglas NJ & Engleman HM[20]	5	—	—
Jenkinson C et al.[10]	4	17	—
Kezirian EJ et al.[17]	3	—	7.5
Kezirian EJ et al.[6]	5	—	—
McDaid C et al.[13]	3	—	10.5
Moore P et al.[11]	4	17	—
Moyer CA et al.[1]	3	—	4
Mulgrew AT et al.[18]	1	21.5	—
Piccirillo JF[4]	4	—	—
Stucki A et al.[16]	5	—	—
Sunwoo BY et al.[23]	3	—	—
Thong JF & Pang KP[3]	4	—	—
Tregear S et al.[22]	1	—	10
Weaver EM et al.[21]	1	21.5	—
Weaver TE et al.[12]	4	21	—
	5	—	—
Wright J et al.[2]	3	—	9
Ye L et al.[14]	4	20	—

AMSTAR = assessment of multiple systematic reviews; DB Scale = Downs and Black scale.

General QOL Measurements

Five reviews and cohort studies were identified that assessed general QOL measurements with respect to OSA outcomes after therapy (Table II). The most commonly used was the Short Form-36 (SF-36), but others were also identified including the Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), and others.[1,4,10–15] Most reports found either weak or poor correlations between AHI and SF-36 scores in the physical functioning, role physical, social functioning, mental health, vitality, and general health subscores.[1,10,13] The NHP was seen to show robust responses to OSA therapy, but no correlation with AHI was identified.[1,14,15] The SIP was also seen to show responses to therapy especially in regard to performance measures such as alertness, sleep, recreation, and work.[1,12,14]

Interpretation of general QOL measurements. Multiple well-validated scales have been used to assess both diagnostic and therapeutic outcomes in OSA. The measures are not OSA specific, but are patient centered and reflect changes in health status pertaining to OSA. No strong evidence exists linking general QOL measurements to AHI.

TABLE II.
Treatment Outcomes Using General Quality of Life Measurements.

Parameter	Study	CPAP Outcomes	Surgery Outcomes
SF-36	Weaver EM et al.[15]		No correlation between SF-36 and AHI comparing subjects undergoing RF tongue and palate reduction and sham surgery
	Jenkinson C et al.[10]	No difference after 3 months CPAP treatment	
	Weaver TE[12]	CPAP had large effect on vitality subscale in three studies, social functioning and mental subscales in one study	
	McDaid C et al.[13]	Significant differences in bodily pain, general health, physical function in CPAP vs. placebo in three parallel studies	
		No significant difference in three crossover trials	
		CPAP users scored significantly better in physical and mental component scores and the bodily pain subscale compared to dental devices	
NHP	Moyer CA et al.[1]	CPAP users significantly greater improvement in NHP compared to changes in sleeping position	
		No effect of NHP in subjects with mild OSA	
	Weaver TE[12]	Significantly improved in social inclusion and energy domains in two studies after CPAP	
	McDaid C et al.[13]	Significantly improved in social inclusion and energy domains in two studies after CPAP	
SIP	Weaver TE[12]	CPAP found to have large effect on total SIP scores in two studies	
		CPAP found to have large effect on sleep and rest, recreation, and pastime domains in one study	
		CPAP found to have moderate effect on psychological, household tasks, and work domains in one study	
FLP	Moyer CA et al.[1]	CPAP had positive effect on FLP	
	Weaver TE[12]	CPAP had positive effect on FLP	
		Large effect of CPAP on sleep and rest domain of FLP	
		Moderate effect of CPAP on social interaction, psychosocial domains, and total FLP score	
EQ-5D	Moyer CA et al.[1]	No change in scores after treatment of CPAP after 3 months	
	Jenkinson C et al.[10]	No change in scores after treatment of CPAP after 3 months	
	McDaid et al.[13]	No change in scores between subjects using CPAP and controls	
MLQDL	Moyer CA et al.[1]	Lower scores in untreated OSA than subjects using CPAP after 3 months	
Danoff scale	American Thoracic Society/American Sleep Disorders Association [24]	High life satisfaction after treatment for OSA	
PGI	Jenkinson C et al.[10]	Considerable change after 3 months of CPAP	
GRISS	McDaid C et al.[13]	No difference in quality of life comparing CPAP vs. dental devices	

AHI = apnea-hypopnea index; CPAP = continuous positive airway pressure; EQ-5D = EuroQoL; FLP = Functional Limitations Profile; GRISS = Golombok Rust Inventory of Sexual Satisfaction; MLQDL = Munich Life Quality Dimension List; NHP = Nottingham Health Profile; OSA = obstructive sleep apnea; PGI = Patient Generated Index; RF = radio frequency; SF-36 = Short Form-36; SIP = Sickness Impact Profile.

Disease-Specific QOL Measurements

Disease-specific QOL measurements have been developed to describe subjects with a known diagnosis of OSA and study outcomes to therapy. These include the Functional Outcomes of Sleep Questionnaire (FOSQ), Calgary Sleep Apnea Quality of Life Index (SAQLI), and the Obstructive Sleep Apnea Patient-Oriented Severity Index, all of which emphasize measurements on body

TABLE III.
Treatment Outcomes using Disease Specific Quality of Life Measures.

Parameter	Study	CPAP Outcomes	Surgery Outcomes
FOSQ	Weaver TE[12]	Large effect seen comparing CPAP and placebo	
	McDaid C et al.[13]	Significant improvement in scores comparing CPAP to placebo in four trials	
		Significantly more improvement in scores in subjects treated with CPAP compared to dental devices	
		No change in subjects treated with CPAP vs. sleep posture	
	Ye L et al.[14]	Women had significantly greater improvement after CPAP compared to men	
	Kezirian EJ et al.[17]	Scores improved after using tongue stabilization device	
SAQLI	Moyer CA et al.[1]	Significant improvement in all SAQLI domains after 4 weeks of CPAP treatment	
	Weaver TE[12]	Large effect in symptom subscale; moderate effect in daily functioning, social interactions, emotional functioning in one study	
	McDaid C et al.[13]	Improvement in daily function, emotional, and symptom subscales with CPAP vs. placebo	
		Improvement in SAQLI scores when "treatment symptoms" subscale omitted comparing CPAP and dental devices	
	Mulgrew AT et al.[18]	No difference in scores between formal PSG titration for CPAP and autotitrating CPAP	
OSAPOSI	Weaver TE[12]	Improvement in 50% subjects undergoing CPAP vs. 2.1% undergoing diet and exercise	Improvement in 24% undergoing surgical intervention

CPAP = continuous positive airway pressure; FOSQ = Functional Outcomes of Sleep Questionnaire; OSAPOSI = Obstructive Sleep Apnea Patient-Oriented Severity Index; PSG = polysomnography; SAQLI = Calgary Sleep Apnea Quality of Life Index.

functions, energy level, sustaining attention, memory, emotional functions, and in the case of the SAQLI, QOL as it pertains to therapy.[16–18] A summary of these results is seen in Table III. The most deeply studied is the FOSQ, and although this scale is highly responsive to OSA therapy via CPAP, to date it has not been sufficiently studied for surgical outcomes.[12–14,16,17] Although AHI and FOSQ have not been formally compared, little association exists between the two outcome measures after surgical interventions.[17] SAQLI has the advantage of containing a domain that can rate the subjects' QOL in reference to the therapy they are undergoing for OSA (domain E of the scale).[1,4,12,13,18]

Interpretation of disease-specific QOL measurements. Most of the disease-specific QOL measures of OSA are considered research tools, with their length in some cases limiting immediate clinical utility. However, good quality data exist linking their scores to diagnostic and therapeutic assessments of OSA. The reviewed studies did not show strong evidence linking disease-specific QOL measurements to AHI.

Measurements of Sleepiness

Assessment of sleepiness is a key subjective outcome in OSA therapy, and studies reviewing this as an outcome are seen in Table IV. The Epworth Sleepiness Scale (ESS) is one of the most commonly reported non-AHI measures in the OSA literature. Although it is a metric of pathological somnolence in general, rather than an OSA-specific measurement, it is commonly employed for OSA analysis in the CPAP literature. Eight studies included the ESS as an outcome measure, some of which were surgical articles.[3,12–14,17–20] The ESS does show a strong response to intervention for OSA in general, but when compared to AHI, Spearman correlation found no relation.[21] Other measures of sleepiness are not as well studied as the ESS.

Interpretation of sleepiness measurements. Sleepiness is a consistent complaint in OSA and is well measured via the ESS as well as other scales. The ESS in particular is robustly validated and easily employed in the clinical setting. The ESS correlates strongly with QOL measures.

Performance Measurements

Executive performance and decision making are higher cortical functions than are subjectively affected in OSA. The two most commonly used measures are an assessment of motor vehicle collisions (MVC), and the Psychomotor Vigilance Test (PVT); these results are summarized in Table V.[2,13–15,20,22,23] A systematic review by Tregear et al. clearly identified the significant role that motor vehicle accident rates can play in assessing severity of OSA from a population perspective, and also as a robust measure of response to therapy via CPAP.[22] MVC rate does seem to correlate with AHI, whereas

TABLE IV.
Treatment Outcomes Using Measurements of Sleepiness.

Parameter	Study	CPAP Outcomes	Surgical Outcomes
ESS			
	Weaver TE[12]	Large improvement in ESS when subjects treated with CPAP in 15 studies	
		Large improvement in ESS when subjects treated with autotitrating CPAP in three studies	
	McDaid C et al.[13]	Significant decrease in ESS comparing CPAP vs. placebo, especially in subjects with higher baseline ESS	
		No change in ESS when CPAP compared to dental devices or sleeping position	
	Kezirian EJ et al.[17]	Significant improvement in ESS in subjects treated with tongue stabilization devices in three of six studies	Improvement in ESS scores in hyoid suspension techniques in three of four studies
			Improvement of ESS from 16 to six in patients undergoing hyoepiglottoplasty
			Improvement of ESS in genioglossus advancement and hyoid suspension in three or seven studies
	Mulgrew AT et al.[18]	No significant change in ESS in subjects undergoing formal PSG titration of CPAP vs. autotitrating CPAP	
	Caples SM et al.[19]		Improvement in ESS after UPPP in two studies
			Improvement of ESS from 9.7 to 6.7 in subjects undergoing RF tongue base ablation, though no comparison with controls
			Improvement of ESS from 17.8 to 4.7 in subjects undergoing maxilla-mandibular advancement
	Douglas NJ & Engelman HM[20]	Significant decrease in ESS with CPAP	
SWAI	Weaver TE[12]	Moderate improvement after treatment with CPAP	
SSS	Weaver TE[12]	Large improvement after CPAP in two studies	
IDS	Weaver TE[12]	Large improvement after CPAP in one study	

CPAP = continuous positive airway pressure; ESS = Epworth Sleepiness Scale; IDS = Index of Daytime Sleepiness score; PSG = polysomnography; RF = radio frequency; SSS = Stanford Sleepiness Scale; SWAI = Sleep Wake Activity Inventory; UPPP = uvulopalatopharyngoplasty.

PVT, when assessed as an independent measure, is reliable when performed under controlled conditions by experienced raters, and is responsive to OSA treatment.[23]

Interpretation of performance measurements. Assessing MVC collision rate is not practical for small-scale or short-term studies examining OSA outcomes but is very useful as a population measure. PVT is a straightforward potential patient-centric avenue to explore regarding sleepiness response to therapy and is underemphasized in the literature. This particular test can be confounded with factors such as caffeine intake, shift work, and time of day that require stringent control before it is used, but it represents an area to explore further.

Biological Measurements

Despite the well-accepted concept that OSA is an independent risk factor for adverse cardiovascular function, and hypertension in particular, cardiovascular outcomes are under-reported in the OSA literature. Only five review articles reported the effect of OSA on hypertension (HTN).[2,3,13,19,24] Thong and Pang found that individuals with an elevated AHI were more likely to have severe hypertension.[3] Wright et al. describe 18 cross-sectional studies, of which four describe a significant association between OSA and morning blood pressure.[2] Evidence of improvement in HTN with treatment of OSA with CPAP is contradicting.[24] Only one study discussed other biologic outcomes in relation to OSA.[25] Myeloperoxidase (MPO) and C-reactive protein (CRP) are markers of inflammation and have been reported to be elevated in subject with OSA.[25] Salivary MPO and serum CRP were elevated in the OSA group compared to the control, but not serum MPO.[25] Using a regression model, salivary MPO had a positive correlation with AHI.[25]

Interpretation of biological measurements. Blood pressure is an easily assessed OSA outcome, and although influenced by many other factors, it remains a key outcome for OSA therapy. Blood pressure correlates well with OSA outcomes especially in the more severe disease state. Hypertensive outcomes are significantly

TABLE V.
Treatment Outcomes Using Performances Measures.

Parameter	Study	CPAP Outcomes
MVC	McDaid C et al.[13]	Sufferers of OSA performed as poorly as intoxicated subjects in stimulated driving tests
		No difference was found in the number of obstacles hit in a driving stimulator in CPAP vs. non-CPAP users
		Less deviation in position on the road when subjects treated with CPAP vs. sham-CPAP
	Douglas NJ & Engleman HM[20]	CPAP improved distances to drive and decreases near-miss accidents
	Tregear S et al.[22]	Rate of MVC decreases with treatment with CPAP
PVT	Ye L et al.[14]	Improvement in scores with CPAP, but no change in number of lapses in women

CPAP = continuous positive airway pressure; MVC = motor vehicle collisions; OSA = obstructive sleep apnea; PVT = performance vigilance test.

under-reported in the OSA literature. Work on other biological markers of disease is as yet sporadic and not ready for clinical application.

DISCUSSION

The management of OSA emphasizes AHI during initial diagnosis of OSA, severity stratification, and patient response to treatment.[5] OSA, however, is a multifactorial disease affecting patients in many dimensions including QOL, cognition, sleepiness, and cardiovascular physiology. These measures, far more patient-centered than AHI, have been severely underemphasized in the surgical OSA literature, although they are becoming more common in other areas of OSA work. The goal of our novel study was to demonstrate the use of other outcome dimensions as adjuncts in the measurement of OSA. We have found that a variety of QOL, sleepiness, performance, and biologic measures can be confidently used in the context of OSA therapy. Scores in both general and disease-specific QOL measures were different in patients with OSA compared to controls. Additionally, QOL scores are responsive to treatment of OSA. The ESS was increased in OSA sufferers and decreased with treatment. OSA resulted in poorer scores on PVT and driving-related measures that also improved with treatment. OSA can also affect physiology including morning blood pressure, ischemic changes in electrocardiogram, pulmonary hypertension, and incidence of stroke. Both objective and subjective non-AHI measures of OSA are sensitive to changes after treatment with CPAP and other modalities. Interestingly, correlation between AHI and ESS and SF-36 was variable, indicating that these two outcome measures may address different aspects of the disease not adequately represented by AHI. Additionally, some measures showed robust response with treatment at variance with AHI outcomes.

Patients with OSA present with a wide spectrum of complaints, ranging from asymptomatic disease to severe presentation of somnolence and poor sleep. Patient symptoms do not correlate well with severity of OSA as defined by AHI.[3] Despite the clinician's focus on AHI in diagnosis and treatment of OSA, other outcome measures tend to be of more importance to patients including QOL, subjective sleepiness, and level of performance. In addition, treatment aims to prevent the long-term deleterious effects of OSA such as high blood pressure and cardiac disease, yet such outcomes are substantially underutilized in evaluating surgical treatment outcomes.

Even when AHI is used, the interpretation of outcome after surgery can still be unclear. Hobson et al. recently showed in a creative study that differences even in the definition of AHI severity cutoff can greatly influence reported efficacy of surgery in patients with OSA.[26] When using the accepted Sher criteria of 50% reduction in AHI being considered success, the issues can be even further muddied.[26] For example, a patient with a baseline AHI of 80 who has a postoperative PSG showing an AHI of 41 would likely experience measurable symptomatic clinical improvement even though they are not defined as a successful surgical outcome by AHI criteria, whereas a patient with baseline AHI of 15, reduced postoperatively to under 5, is considered a successful AHI outcome even though the likelihood of clinical cardiovascular or QOL impact may be minimal. Postoperative AHI assessment is further complicated by the need for PSG, an expensive proposition for resource allocation and patient convenience at which both insurers and patients increasingly balk. Therefore, utilizing other outcome measures such as QOL questionnaires or cardiac assessments may be more practical to check effectiveness of interventions for OSA in some patient populations.

Certain advantages exist when using non-AHI outcomes measures in assessing OSA changes after therapy. Questionnaires such as the ESS, FOSQ, and SF-36, or physiological data such as blood pressure, are highly patient-centric outcomes and are thus well accepted by patients as valid and meaningful, almost personalized, metrics of assessment that can be done without inconvenience in the clinical setting. There are no problems in regard to scheduling of PSG or identifying the best time point after the intervention in which to do it; questionnaires can be administered at multiple time points, over the phone, or via virtual Internet means, all of which allow for greater flexibility in assessment and allow for multiple repeated measures. Two recent large-scale multi-institutional studies on CPAP, those being the APPLES (Apnea Positive Pressure Long Term Efficacy Study) and CATNAP (CPAP Apnea Trial North America Program) studies have recognized these issues and employed patient-reported QOL outcomes and functional status as central metrics in their assessments.[27,28] Although these studies focused on CPAP as opposed to surgical outcomes, their robust methodologies give credence to the utility of non-AHI outcomes measures as being of increasing importance in OSA assessment.

Our study has a number of limitations. Being a systematic review it is dependent on the quality of

published trials to show the direction of literature, but most of the literature reviewed were high-quality cohort or review studies. Therefore, the content of the reported findings is based on generally strong evidence. Effect size of the various outcomes could not be commented on because different studies used different gradations of measures of effect. Most of the retrieved articles were from the nonsurgical literature, calling into question the applicability of our findings to surgical trials. It is possible that if non-AHI measures identified by our work are adopted for surgical studies, different findings may be seen than were observed in the current study; this is a concept that would be useful to investigate. Finally, certain aspects of questionnaires, such as their ability to predict cardiovascular dysfunction or mortality risk, have yet to be clarified.

Future research directions may focus on the application of different outcome measures in the diagnosis and monitoring of OSA. It would be interesting to see if tools such as FOSQ or SAQLI can be combined with ESS to develop a more robust assessment measure that is OSA specific across studies. Likewise, measurements from different categories (e.g., physiologic vs. QOL) were rarely compared side by side, and this would be useful to explore. Finally, outcomes from QOL measures and ESS were much more uniform compared to performance and biologic measures, where studies used many different end points or measurement tools. These issues all merit further work to delineate.

CONCLUSION

OSA affects many aspects of biopsychosocial health, yet evaluation of surgical treatment has traditionally been focused almost exclusively on AHI outcomes. Other metrics of OSA measurement, such as QOL, subjective sleepiness, performance, and biological measures, are increasingly found in the CPAP literature and could play a larger role to detect differences with surgical treatment as well. Validated subjective measurements, such as QOL and subjective sleepiness scores, best reflect the patient experience. Objective, performance-based, or biologically based outcome measures (e.g., blood pressure) directly measure the important long-term goals that OSA interventions aim to treat. Such outcome measures can and should be utilized in concert with AHI to more fully display the impact of surgery for OSA on patients. Considering the wide variety of validated patient-centric outcomes that are readily available, it is not sufficient to limit reports of OSA surgical treatment to changes in AHI.

Acknowledgments

The authors gratefully acknowledge Dr. Kevin White (Biostatistics) for his contribution in the statistical portion of this work, and Dr. Charles George (Respirology) for his advice and review of the article.

BIBLIOGRAPHY

1. Moyer CA, Sonnad SS, Garetz SL, Helman JI, Chervin RD. Quality of life in obstructive sleep apnea: a systematic review of the literature. *Sleep Med* 2001;2:477–491.
2. Wright J, Johns R, Watt I, Melville A, Sheldon T. Health effects of obstructive sleep apnoea and the effectiveness of continuous positive airways pressure: a systematic review of the research evidence. *BMJ* 1997;314: 851–860.
3. Thong JF, Pang KP. Clinical parameters in obstructive sleep apnea: are there any correlations? *J Otolaryngol Head Neck Surg* 2008;37:894–900.
4. Piccirillo JF. Outcomes research and obstructive sleep apnea. *Laryngoscope* 2000;110(3 pt 3):16–20.
5. Epstein LJ, Kristo D, Strollo PJ Jr, et al. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med* 2009;5:263–276.
6. Kezirian EJ, Weaver EM, Criswell MA, De Vries N, Woodson BT, Piccirillo JF. Reporting results of obstructive sleep apnea syndrome surgery trials. *Otolaryngol Head Neck Surg* 2011;144:496–499.
7. Shin JJ, Rauch SD, Wasserman J, Coblens O, Randolph GW. Evidence-based medicine in otolaryngology, part 2: the current state of affairs. *Otolaryngol Head Neck Surg* 2011;144:331–337.
8. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377–384.
9. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
10. Jenkinson C, Stradling J, Petersen S. How should we evaluate health status? A comparison of three methods in patients presenting with obstructive sleep apnoea. *Qual Life Res* 1998;7:95–100.
11. Moore P, Bardwell WA, Ancoli-Israel S, Dimsdale JE. Association between polysomnographic sleep measures and health-related quality of life in obstructive sleep apnea. *J Sleep Res* 2001;10:303–308.
12. Weaver TE. Outcome measurement in sleep medicine practice and research. part 1: assessment of symptoms, subjective and objective daytime sleepiness, health-related quality of life and functional status. *Sleep Med Rev* 2001;5:103–128.
13. McDaid C, Griffin S, Weatherly H, et al. Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea-hypopnoea syndrome: a systematic review and economic analysis. *Health Technol Assess* 2009;13:iii–iv, xi–xiv, 1–119, 143–274.
14. Ye L, Pien GW, Ratcliffe SJ, Weaver TE. Gender differences in obstructive sleep apnea and treatment response to continuous positive airway pressure. *J Clin Sleep Med* 2009;5:512–518.
15. Weaver EM, Woodson BT, Steward DL. Polysomnography indexes are discordant with quality of life, symptoms, and reaction times in sleep apnea patients. *Otolaryngol Head Neck Surg* 2005;132:255–262.
16. Stucki A, Cieza A, Schuurmans MM, et al. Content comparison of health-related quality of life instruments for obstructive sleep apnea. *Sleep Med* 2008;9:199–206.
17. Kezirian EJ, Hussey HM, Brietzke SE, et al. Hypopharyngeal Surgery in Obstructive Sleep Apnea: Practice Patterns, Perceptions, and Attitudes. *Otolaryngol Head Neck Surg* 2012;147:964–971.
18. Mulgrew AT, Fox N, Ayas NT, Ryan CF. Diagnosis and initial management of obstructive sleep apnea without polysomnography: a randomized validation study. *Ann Intern Med* 2007;146:157–166.
19. Caples SM, Rowley JA, Prinsell JR, et al. Surgical modifications of the upper airway for obstructive sleep apnea in adults: a systematic review and meta-analysis. *Sleep* 2010;33:1396–1407.
20. Douglas NJ, Engleman HM. CPAP therapy: outcomes and patient use. *Thorax* 1998;53(suppl 3):S47–S48.
21. Weaver EM, Kapur V, Yueh B. Polysomnography vs self-reported measures in patients with sleep apnea. *Arch Otolaryngol Head Neck Surg* 2004; 130:453–458.
22. Tregear S, Reston J, Schoelles K, Phillips B. Continuous positive airway pressure reduces risk of motor vehicle crash among drivers with obstructive sleep apnea: systematic review and meta-analysis. *Sleep* 2010;33:1373–1380.
23. Sunwoo BY, Jackson N, Maislin G, Gurubhagavatula I, George CF, Pack AI. Reliability of a single objective measure in assessing sleepiness. *Sleep* 2012;35:149–158.
24. American Thoracic Society/American Sleep Disorders Association. Statement on health outcomes research in sleep apnea. *Am J Respir Crit Care Med* 1998;157:335–341.
25. Akpinar ME, Yigit O, Altundag A, Demirel GY, Kocak I. Salivary and Serum Myeloperoxidase in Obstructive Sleep Apnea. *J Otolaryngol Head Neck Surg* 2012;41:215–221.
26. Hobson JC, Robinson S, Antic NA, et al. What is "success" following surgery for obstructive sleep apnea? The effect of different polysomnographic scoring systems. *Laryngoscope* 2012;122:1878–1881.
27. Kushida CA, Nichols DA, Quan SF, et al. The Apnea Positive Pressure Long-term Efficacy Study (APPLES): rationale, design, methods, and procedures. *J Clin Sleep Med* 2006;2:288–300.
28. Weaver TE, Mancini C, Maislin G, et al. Continuous positive airway pressure treatment of sleepy patients with milder obstructive sleep apnea: results of the CPAP Apnea Trial North American Program (CANAP) randomized clinical trial. *Am J Respir Crit Care Med* 2012;186: 677–683.