



Published in final edited form as:

Otolaryngol Head Neck Surg. 2011 April ; 144(4): 496–499. doi:10.1177/0194599810396791.

Reporting Results of Obstructive Sleep Apnea Syndrome Surgery Trials

Eric J. Kezirian, MD, MPH,

Department of Otolaryngology—Head and Neck Surgery, University of California, San Francisco, San Francisco, California

Edward M. Weaver, MD, MPH,

Department of Otolaryngology—Head and Neck Surgery, University of Washington, Seattle, Washington

Mark A. Criswell, MD,

Otolaryngology — Head and Neck Surgery Service, Winn Army Comm. Hospital, Ft. Stewart, Georgia

Nico de Vries, MD, PhD,

Department of Otolaryngology/Head and Neck Surgery, Sint Lucas Andreas Ziekenhuis, Amsterdam, Netherlands

B. Tucker Woodson, MD, and

Department of Otolaryngology & Communication Sciences, Medical College of Wisconsin, Milwaukee, Wisconsin

Jay F. Piccirillo, MD

Department of Otolaryngology—Head and Neck Surgery, Washington University School of Medicine, St. Louis, Missouri

Abstract

Obstructive sleep apnea syndrome surgery studies largely evaluate single procedures or procedure combinations in case series designs, but it can be difficult to compare results across studies. We present a standardized format for presentation of surgical study results to facilitate pooled analyses and subgroup analyses. The format includes thorough characterization of baseline subject characteristics and the use of outcome measures that reflect the spectrum of obstructive sleep apnea and its consequences. As the apnea-hypopnea index is the most common, albeit controversial, primary outcome measure in obstructive sleep apnea syndrome surgery studies, we propose analysis and reporting standards to facilitate understanding its role as an outcome

Corresponding Author: Eric J Kezirian, MD, MPH, 2233 Post Street, 3rd Floor, Campus Box 1225, San Francisco, CA 94115, Phone: 415-353-2479, Fax: (415) 885-7546, ekezirian@ohns.ucsf.edu.

Financial Disclosures: Kezirian: Apnex Medical (medical advisory board, consultant), ArthroCare (consultant), Medtronic (consultant), Pavad Medical (consultant), ReVENT Medical (medical advisory board). Weaver: None. Criswell: None. De Vries: ReVent Medical (medical advisory board, consultant), Schering-Plough Netherlands (medical advisory board), Inspire Medical Systems (investigator). Woodson: Research support Inspire Medical; Consultant Resmed, Inspire Medical, Johnson and Johnson, Medtronic, Siesta Medical, Acceptent; royalty from hyoid suspension patent Medtronic ENT. Piccirillo: Apnex Medical (Chair, Data Safety & Monitoring Board).

measure. Because surgical outcomes vary according to subject characteristics, investigators should also evaluate the potential association between baseline subject characteristics and outcomes.

INTRODUCTION

The literature concerning the definitive (as opposed to adjunctive) surgical treatment of obstructive sleep apnea syndrome (OSA) consists primarily of case series studies, although randomized surgical trials and cohort studies exist.¹⁻³ OSA surgery studies largely evaluate single procedures or procedure combinations without a direct comparison of alternative treatments within a single study. Comparisons of results across case series studies have been performed, but these are limited without standardized formats for presentation of study results.^{1,2,4} While multi-institutional, randomized surgical trials can potentially provide generalizable, high-level evidence, there are numerous challenges to conducting these studies, including those related to ethics, feasibility, and study design. In light of these barriers, the available literature, including case series studies, is relevant. Nevertheless, there is potential to make the reporting of OSA surgery study results even more valuable.

A framework for reporting results from OSA surgical trials, regardless of study design, will enhance study interpretation, improve patient care, address fundamental unanswered questions, and direct future investigations. In this Commentary, we focus on reporting of baseline characteristics, outcome measures, and the analysis of results, including statistical analysis. Previous publications have described unique methodology considerations for OSA surgical trials,⁵ and this framework complements rather than replaces these discussions.

BASELINE CHARACTERISTICS

Subject factors are associated with outcomes of OSA surgical treatments

Individual procedures and procedure combinations vary widely in results for individual subjects and across studies.² Although some is due to differences in surgical technique and variability in outcome measurement, a substantial portion may be related to differences in baseline subject characteristics. Complete presentation of baseline characteristics and the evaluation of associations with outcomes may clarify the underlying mechanisms. Explicit characterization of subjects will establish the generalizability of results, allow for pooled data analyses, and enable subgroup analyses.

Reporting of the following baseline subject characteristics are recommended

- Demographic factors (e.g., age, gender, race, and ethnicity)
- Study eligibility criteria, including intolerance to positive airway pressure, and a full reporting of the number who did and did not meet eligibility criteria
- Baseline sleep study results when relevant, with sleep study technique and explicit criteria used for interpretation and scoring of disordered breathing events
- Clinical characteristics, including those used as selection criteria and those potentially associated with outcomes (body mass index, tonsil size, modified Mallampati position, Friedman Stage, nasal obstruction, and findings of other

airway evaluations such as awake endoscopy, lateral cephalogram, or drug-induced sleep endoscopy)

- Subject-based and/or objective measures of physical, functional, and emotional consequences

OUTCOME MEASURES

OSA is not solely a number

Whether the number in question is the change in the apnea-hypopnea index (AHI), the percentage of subjects who achieved an arbitrary response to treatment, or some other metric, the use of single numbers in reporting trial outcomes oversimplifies the complexity of OSA disease burden and treatment outcome.

As a surrogate outcome, the AHI is standard but not a sufficient examination of treatment outcomes, as it does not capture the spectrum of OSA and clinically meaningful endpoints completely

Whenever possible, clinical outcomes should be measured primarily in preference to surrogate outcomes.⁶ While the AHI defines OSA severity, it and other sleep study parameters are surrogate measures. The AHI is the most common primary study endpoint, but other sleep study parameters may be more reliable and physiologically more important (e.g., apnea index, desaturation index, percentage of sleep time with oxygen saturation below 90%). OSA treatment is based not only on improving breathing patterns during sleep but also on alleviating the health-related (primarily cardiovascular and endocrine), behavioral (daytime sleepiness, quality of life), functional (performance, reaction time, driving), and social (disruptive snoring) consequences of the disorder, all of which must be weighed against surgical complications and side effects. The AHI and other sleep study results are intermediate measures poorly associated with many clinical outcomes at baseline, and it is unclear whether treatment-related changes in the AHI are associated with changes in these other potential consequences. OSA surgical treatment studies commonly report the changes in AHI but often do not report other endpoints.

Although objective measures minimize the potential placebo effect that can occur with any treatment, subjective endpoints are valuable

All medical and surgical treatments may have a substantial placebo effect, particularly with regard to subjective outcomes. The placebo-controlled trial is one method of evaluating a potential placebo effect. Although there are placebo-controlled trials for minimally-invasive OSA surgical treatments,^{7,8} the successful completion of such surgical trials is more challenging methodologically than for medications and may be impossible for many invasive procedures. Other controlled trial designs are available that provide higher-level evidence than case series, including comparisons to alternative treatment or no treatment. Objective outcomes are possible for some endpoints but can be cumbersome, expensive, or unavailable. Valid, reliable measures can evaluate subjective outcomes, with specific statistical and analytical approaches to evaluate potential placebo effects.

ANALYSIS OF RESULTS

In longitudinal studies (e.g., case series or cohort studies), preoperative and postoperative values of all outcome measures should be reported along with 95% confidence intervals around the difference. The paired t-test will evaluate for statistically significant changes in continuous outcome measures, and the confidence interval will evaluate the possibility of clinically meaningful differences

Data collected at baseline and following surgery in the same group of subjects are considered paired. The Student's t-test compares the mean and standard deviation of the preoperative and postoperative data but does not account for the paired nature of the data. With paired data, the paired t-test is preferred because it has greater statistical power and is therefore (appropriately) more likely to show a statistically-significant difference than a Student's t-test.

In controlled studies, changes in outcomes should be tested between groups

The between-group comparisons are not paired, so the standard Student's t-test is appropriate, although the paired t-test can evaluate for changes within each group individually. Technically, when assumptions of normality are not met, comparable non-parametric tests should be used (i.e., Mann-Whitney test). The calculation of the 95% confidence interval around the difference, along with knowledge of the size of a clinically meaningful difference, helps interpret the clinical significance of results.

Changes in the AHI (or other primary outcome measure) should be reported alongside other outcome measures, with a comparison of changes in the former with the latter

Because OSA treatment studies commonly employ a primary outcome that is a surrogate measure, such as the AHI, it is essential to determine whether changes in the AHI mirror those for other (e.g., health-related, behavioral, and functional) outcomes. These comparisons can also define the extent to which the AHI *should* be the primary outcome measure in OSA surgical studies. To the extent that the AHI is a meaningful objective outcome, this comparison may be particularly important for subjective outcomes, where it can help differentiate treatment and placebo effects.

There is no evidence-based definition of effective surgical OSA treatment, and presentation of results in absolute terms and according to specific criteria will help define target outcomes

For positive airway pressure, one common definition of adequate compliance is use at least 4 hours per night on at least 5 nights per week, but this is based on very limited evidence and represents a relatively small fraction of recommended sleep time (20 vs. 49–56 hours per week). OSA surgical studies have used the AHI as the primary outcome, and commonly-used definitions of response are based on little or no evidence. Arbitrary thresholds such as a reduction in the AHI of at least 50% to an absolute level below 15 or 20 events/hour without oxygen desaturation have been severely criticized.^{1,2} Some have argued for a more-stringent threshold requiring a postoperative AHI below 5 events/hour,¹⁰ but *all of these proposed criteria are not evidence-based.*

For studies with the AHI as the primary outcome, we propose presentation of data in aggregate (absolute change in the AHI) as well as subgroup analyses based on distinct definitions of treatment response, including a reduction of at least 50% in the AHI to levels below 20, 15, 10, and 5 events/hour

Eventually these data may establish an evidence basis for defining effective treatment a threshold for “responders” if one truly exists, based on whether specific subgroups do or do not have changes in secondary outcome measures. For example, assuming use of the AHI as the primary outcome, if there is a reduction in blood pressure and/or improvement in sleep-related quality of life in subjects who achieve a reduction in the AHI below 5 but not 10, 15, or 20 events/hour, this would suggest that the former is a more appropriate definition of treatment response. Use of the AHI and arbitrary cutpoints may prove useless, but presentation of data in a standard format will enable an evaluation. Presentation of subject-level data will enable pooling of data across multiple studies. Potential alternatives for evaluating the postoperative AHI include a change in the median index, the percentage reduction with surgery, or mathematical calculations that equate the residual index to a degree of positive airway pressure compliance.⁹ All or some of these can complement the above cutpoints. The timing of postoperative sleep studies relative to the surgery date should be reported.

The examination of subject factors associated with outcomes is essential

Differences in baseline subject characteristics explain some variation in reported surgical outcomes. Although some OSA surgical studies have identified associations between outcomes and subject factors, many studies have not examined them. Statistical analysis should explore potential associations between outcomes and baseline characteristics, whether demographic factors or findings of specific preoperative evaluation techniques. For variable baseline characteristics such as body mass index or body weight, paired t-tests should consider changes following surgery, and statistical techniques (e.g., regression analysis, MANOVA, MANCOVA) enable adjustment. Subgroup analyses, while often only hypothesis-generating, will help identify predictors of surgical outcomes and surgical indications.

Statistical analysis requires a combination of tests and caveats, with an understanding of the distinction between statistical and clinical significance

This framework proposes what for many studies would be an expanded analysis plan, for which it is critical to utilize statistical tests appropriately and judiciously. With a greater number of statistical tests, patterns (such as the consistency of changes across outcome measures) are likely more important than individual results of statistical significance, and the Bonferroni correction or similar technique may be needed to avoid the problem of multiple comparisons. Most importantly, the distinction between statistical and clinical significance cannot be overestimated. Reporting of 95% confidence intervals and *a priori* establishment of clinically meaningful differences is recommended. The most valuable presentation of results uses raw data. Presenting raw data in tabular form for larger studies is most appropriate in the online supplemental appendix now widely available.

CONCLUSIONS

We recommend a standardized and comprehensive approach to the reporting of OSA surgery trials. These standards will help in the analysis of individual study outcomes and the determination of treatment effectiveness.

Acknowledgments

This Commentary is supported by the Sleep Disorders Committee of the American Academy of Otolaryngology--Head and Neck Surgery, and the authors wish to acknowledge their contributions.

Funding/Support: Dr. Kezirian is currently supported by a career development award from the National Center for Research Resources (NCRR) of the National Institutes of Health and a Triological Society Research Career Development Award of the American Laryngological, Rhinological, and Otolological Society. Dr. Weaver is supported by NIH/NHLBI R01 HL084139. The project was supported by NIH/NCRR/OD UCSF-CTSI Grant Number KL2 RR024130. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

1. Sher AE, Schechtman KB, Piccirillo JF. The efficacy of surgical modifications of the upper airway in adults with obstructive sleep apnea syndrome. *Sleep*. 1996; 19:156–77. [PubMed: 8855039]
2. Kezirian EJ, Goldberg AN. Hypopharyngeal surgery in obstructive sleep apnea: an evidence-based medicine review. *Arch Otolaryngol Head Neck Surg*. 2006; 132:1–8.
3. 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; 33:1396–407. [PubMed: 21061863]
4. Lin HC, Friedman M, Chang HW, Gurpinar B. The efficacy of multilevel surgery of the upper airway in adults with obstructive sleep apnea/hypopnea syndrome. *Laryngoscope*. 2008; 118:902–8. [PubMed: 18300704]
5. Schechtman KB, Sher AE, Piccirillo JF. Methodological and statistical problems in sleep apnea research: the literature on uvulopalatopharyngoplasty. *Sleep*. 1995; 18:659–66. [PubMed: 8560132]
6. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996; 125:605–13. [PubMed: 8815760]
7. Woodson BT, Steward DL, Weaver EM, Javaheri S. A randomized trial of temperature-controlled radiofrequency, continuous positive airway pressure, and placebo for obstructive sleep apnea syndrome. *Otolaryngol Head Neck Surg*. 2003; 128:848–61. [PubMed: 12825037]
8. Steward DL, Huntley TC, Woodson BT, Surdulescu V. Palate implants for obstructive sleep apnea: multi-institution, randomized, placebo-controlled study. *Otolaryngol Head Neck Surg*. 2008; 139:506–10. [PubMed: 18922335]
9. Ravesloot MJ, De Vries N. Reliable calculation of the efficacy of non-surgical and surgical treatment of obstructive sleep apnea revisited. *Sleep*. 2010; 33
10. Elshaug AG, Moss JR, Southcott AM, Hiller JE. Redefining success in airway surgery for obstructive sleep apnea: a meta analysis and synthesis of the evidence. *Sleep*. 2007; 30:461–7. [PubMed: 17520790]

Table 1

Framework for reporting results of OSA surgery trials

	Analysis and Statistical Technique *	Rationale
Baseline Characteristics	Baseline (proportions and T-tests) Change (variable characteristics, paired t-test) and 95% CI	Baseline characteristics are associated with treatment outcomes
Sleep study outcomes, including: Changes in apnea-hypopnea index—absolute change and 95% CI Response according to criteria (50% reduction below levels of 20/15/10/5 with no oxygen desaturation)	Change (paired t-test) and 95% CI ** Response (proportion) and 95% CI	Outcome measures to capture spectrum of OSA
Health-related outcomes		
Behavioral outcomes		
Social outcomes		
Comparison of change in primary outcome (e.g., apnea-hypopnea index) and change in other outcome measures	Regression analysis or tests of correlation	Determine whether changes in the primary outcome mirror those of other outcome measures May differentiate placebo vs. treatment effects for subjective measures
Comparison of changes in other (non-sleep study) outcomes between responders and non-responders	Changes in outcome measures within responder and nonresponder subgroups (paired t-test) Compare changes in outcome measures between subgroups (t-test)	Examine potential definitions of effective surgical treatment
Examination of baseline characteristics and association with outcomes	Changes in outcome measures (regression techniques or subgroup analyses with t-tests) Response (chi-squared tests for examination of dichotomous baseline characteristics)	Baseline characteristics are associated with treatment outcomes

CI: confidence interval

* Non-parametric alternatives may be more appropriate for smaller samples or those with non-normal distributions.

** Multiple regression techniques may be appropriate to adjust for changes in variable characteristics such as body mass index or body weight.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript