Introduction

Currently, the role of upper airway surgery as a treatment for adult obstructive sleep apnea (OSA) remains controversial, with perspectives on treatment efficacy varying considerably. In contemporary articles, opinion is divided; some claim that surgical intervention “should be considered a viable treatment option,” whereas others suggest that “the majority of sleep apnea patients do not benefit from surgery and may even be harmed by it.” Debate surrounds the interpretation of evidence about the relative effectiveness of various surgical interventions, independent of, but also in comparison to, existing non-surgical treatment alternatives. The Cochrane review in this area supports the restricted use of these procedures

Yet, in Australia, not only is the use of these procedures unrestricted, but it appears that they are widespread and increasing.

Study Objectives: The role of upper airway surgery as a treatment for adult obstructive sleep apnea (OSA) remains controversial, with perspectives on treatment efficacy varying considerably. Though debate may occur in the clinical sphere, it is necessary to appreciate the ever-increasing funding and policy focus on cost effectiveness and “efficacy” in health care.

Design: In this review, we examine contemporary evidence that highlights the importance of “highly effective treatment” over “sub-therapeutic treatment” as a necessity to confer improved health outcomes in OSA. We highlight that assumptions of surgical success inherent in most articles fail to assimilate contemporary, clinically significant indicators of success. We performed a literature search and present interpolated meta-analyses data from 18 surgical articles. Statistical meta-analyses highlight how surgical success decreases when new evidence-based criteria of success are applied.

Measurements and Results: Specifically, when the traditional definition is applied (50% reduction in apnea-hypopnea index [AHI] and/or ≤ 20) the pooled success rate for Phase I procedures is 55% (45% fail). However, at AHI ≤ 10, success reduces to 31.5% (68.5% fail) and, at AHI ≤ 5, success is reduced to 13% (87% fail). According to these definitions, Phase II success (fail) rates decrease from 86% (14%) to 45% (55%) and 43% (57%), respectively.

Conclusions: The evidence for clinical efficacy must define treatment “success.” We propose all future surgical audits report “objective cure” rates with success based on AHI outcomes of ≤ 5 and/or ≤ 10. We hope this paper serves as a catalyst for debate and consensus.

Keywords: Surgery, sleep apnea, evidence, treatment effectiveness, health outcomes

Citation: Elshaug AG, Moss JR, Southcott A et al. Redefining Success in Airway Surgery for Obstructive Sleep Apnea: A Meta Analysis and Synthesis of the Evidence. SLEEP 2007;30(4):461-467

The present authors speculate that the same is true in many other countries around the world. Furthermore, it would seem that much of the debate is currently restricted to the clinical sphere. Though this is important, it is also necessary to appreciate the ever-increasing policy focus on cost-effectiveness and efficacy in health care. One clear and growing example of this, both in Australia and internationally, is in pharmaceutical policy, where approved product expenditure is increasingly gauged on the downstream health outcomes of the populations served. With this as a backdrop, it can only be a matter of time before the controversies surrounding upper airway surgery for OSA gain attention in the policy arena. If this were to occur, greater clarity will be required by the relevant medical specialties of their position within the debate.

Recently, Kezirian and Goldberg contributed to this debate with a review of surgical therapy for OSA, in which they report, “successful outcomes are achieved in 35% to 62% of patients.” This somewhat imprecise conclusion is not surprising given that the various authors of the original research applied varying definitions of “success.” Traditionally, claims of treatment “response,” “effectiveness,” “success,” and even “cure” of OSA have been defined in the literature as a reduction in the Apnea/Hypopnea Index (AHI) of equal to, or greater than, 50%. Some go on to specify this criterion plus a reduction in AHI to 20 or less. Yet, based on these conservative and variable criteria, there is a notable interpretative quandary. That is, such reporting of surgical “success” bears only a limited relationship to the evidence-based criteria of what is “effective” treatment. We believe the Kezirian and Goldberg review, and indeed the assumptions of treatment success inherent in most reviewed articles, fails to adequately acknowledge and synthesize the growing depth of sleep medicine-related health outcomes research that has developed over recent years. We contend that current notions of surgical “success” are overdue for reevaluation.
In 2004, Young et al. reported that, in the United States of America, approximately 1 in 5 adults has at least mild OSA and 1 in 15 adults has OSA of moderate or worse severity. OSA has been well documented to cause hypertension, cardiac morbidity and mortality, automobile accidents, and neurocognitive deficits, as well as impaired quality of life, and increasingly, glucose intolerance. Continuous positive airway pressure (CPAP) is unequivocally regarded as the gold-standard treatment modality available. However, White et al. reported that, on the basis of published literature to the year 2000, there was limited evidence to show that CPAP affects the incidence or outcomes of other medical conditions associated with OSA, such as hypertension, cerebrovascular disease, cardiopulmonary disease, or road traffic accidents. Since 2000 however, a wealth of research evidence has emerged to illuminate the treatment-derived health outcomes associated with OSA; with this are indications of optimal treatment-related clinical endpoints. Indeed, an increasing number of contemporary findings demonstrate the importance of reducing the AHI, in many cases to near or below 5 events per hour of sleep (thereby controlling OSA) in order to improve numerous physiologic, health outcome, and quality-of-life measures. Becker and coworkers provide one landmark example of this: Apneas and hypopneas (AHI) were reduced by approximately 95% and 50% in the therapeutic and sub-therapeutic groups, respectively. Mean arterial blood pressure decreased by 9.9 ± 11.4 mm Hg with effective CPAP treatment, whereas no relevant change occurred with sub-therapeutic CPAP (p = .01). Conclusions: Effective CPAP treatment leads to a substantial reduction in both day and night arterial blood pressure. The fact that a 50% reduction in the AHI did not result in a decrease in blood pressure emphasizes the importance of highly effective treatment. The drop in mean blood pressure by 10 mm Hg would be predicted to reduce coronary heart disease event risk by 37% and stroke risk by 56%.

Similar results have also surfaced in areas related to cardiovascular disease, as well as to heart failure, and endocrinology, and health-related quality of life. All indicate the importance of “highly effective treatment” with a substantial decrease in the AHI over “sub-therapeutic treatment” as a necessity to confer improved health outcomes. The implications of this on clinical and surgical policy and practice are indeed considerable. For the purposes of this review, and to highlight the potential evidence-to-policy gap that currently exists, we have conducted 2 meta-analyses to highlight how surgical “success” rates decrease when contemporaneous, evidence-based criteria of effectiveness are applied.

METHODS

Study Design and Search Strategy

In describing the quality of research evidence, a hierarchy has been developed that assigns each study a “level of evidence” based on the study design. The 5 characteristic levels of evidence are as follows: (1) systematic review of or individual randomized, controlled trial or trials (RCT), (2) cohort study, (3) case-control study, (4) case series, (5) expert opinion.

In evidence-based medicine, level 1 evidence (the RCT) is unequivocally regarded as the gold-standard treatment modality available. Therefore, the primary inclusion criterion for this study was set as the reporting of pre-surgical and post-surgical AHI for individual cases. As demonstrated further in the results section, this minimum inclusion criterion allows for us to meet the aims of this study and to draw conclusions regarding the overall trends in surgery success illustrated by varying definitions of success.

Publications (levels of evidence 1-4) were selected from 2001 to 2005 in order to represent the latest in surgical techniques and therefore to be policy relevant. Prior to 2001, numerous surgical methods changed or were newly dispersed, and this is reflected, in part, by the publication in Sleep of revised practice parameters for the laser-assisted uvulopalatoplasty (LAUP); this report also consisted of recommendations relevant to the uvulopalatopharyngoplasty (UPPP).

Inclusion Criteria for the Selection of Articles

All papers were assessed for the traditional quality-appraisal items of selection bias, allocation bias, confounders, blinding, and data-collection methods. Because the majority of studies in this area are level 4 (case series designs), it was anticipated that objective quality ratings would be inherently lower than that expected in RCT designs. Therefore, in the results section, this minimum inclusion criterion allows for us to meet the aims of this study and to draw conclusions regarding the overall trends in surgery success illustrated by varying definitions of success.

RESULTS

Fourteen articles were identified for Phase I procedures and 4 for Phase II. Table 1 provides a summary of the articles that met the inclusion criteria for data interpolation and meta analyses.

| Method | Study Design and Search Strategy | Inclusion Criteria for the Selection of Articles | Statistical Model | RESULTS | Fourteen articles were identified for Phase I procedures and 4 for Phase II. Table 1 provides a summary of the articles that met the inclusion criteria for data interpolation and meta analyses. | Redefining Success in Airway Surgery for OSA—Elshaug et al. | SLEEP Vol. 30, No. 4, 2007 | 462

Downloaded from https://academic.oup.com/sleep/article-abstract/30/4/461/2708205/Redefining-Success-in-Airway-Surgery-for-OSA by guest on 16 September 2017
Quality Appraisal (Heterogeneity)

Seventeen out of the 18 studies comprise level 4 evidence, and, as such, there is greater scope for selection and allocation bias, though this potential is similar across all included studies (with the exception of the lone RCT). Although blinding is possible in level 4 study designs, it was not apparent in the included studies. Potential confounding in this area relates to 2 factors. (1) It is expected that the AHI is likely to be higher when sleep occurs in the supine compared to the lateral position; therefore, ideally, this variable (ie, time spent supine) needs to be compared and reported from pre-surgical and post-surgical polysomnography (PSG) assessments to identify potential confounding. (2) The second potential confounder is the body mass index (BMI) at pre-surgery versus post-surgery PSG. Again, weight gain or loss may influence OSA severity, thus potentially confounding the surgical results. For all included studies, proportion of time in supine sleep position was not reported, whereas BMI differences were in the majority of but not all studies.

For the quality item “data-collection methods,” these relate primarily to the uniformity of PSG methods and follow-up time from surgery to PSG. There is general PSG uniformity across the included studies; however, follow-up time from surgery to PSG does vary substantially not only within, but between procedures (studies). Though most claimed to aim for a 6-month follow-up period, actual PSG follow-up periods for all studies varied from 6 weeks to 12.3 months, thus indicating further clinical heterogeneity in this respect also. This point is likely to be a co-contributor to findings of heterogeneity.

Importantly, the presence of heterogeneity is not at all surprising in this context and does not so much point to weaknesses in our study design but more to the variability with which the surgical procedures have been applied and reported. For example, from Table 1, it is apparent that in only 1 study was UPPP performed without adjunctive procedures; 2 exist with UPPP plus repose system; 1 with UPPP + radiofrequency volume reduction (RFVR); 2 with UPPP plus genioglossus advancement (GA), and 1 with UPPP plus hyoid suspension (HS). Thus, clinical heterogeneity is apparent, and this appears to be the nature of clinical practice. This phenomenon is supported by a new report highlighting substantial procedural variability in the application of these surgical procedures. Moreover, in the LAUP subgroup, there appears to be homogeneity in terms of case characteristics (age, sex, BMI, etc.) but clinical heterogeneity in relation to the mean number of procedures performed per person across the studies, from 1.5 (Finkelstein), 1.8 (Berger), to 2.4 (Ferguson).

**Surgical Success Results**

The forest plots in Figure 1 illustrate the study-specific and combined proportions of surgical success reported in each of the Phase I surgical trials according to various definitions. The plots indicate a substantial degree of heterogeneity between the studies for each of the 3 definitions of success, and, as a result, overall pooled proportions were calculated using random-effects models (significant Cochran Q Statistics were achieved for each of the 3 success definitions, suggesting heterogeneity between studies). Under the “traditional” definition (ie, a 50% reduction in AHI, an AHI of ≤ 20, or both), the pooled success rate for Phase I surgical procedures was 55% (95% confidence interval [CI]; 0.44, 0.66). However, with success defined as...
a post-surgery AHI of ≤ 10, pooled success reduces to 31.5% (95% CI: 0.21, 0.43), and at AHI ≤ 5, success is reduced to 13% (95% CI: 0.07, 0.21).

Table 2 presents Phase I subgroup analyses together with I² statistics highlighting the degree of heterogeneity. Figure 2 presents plots for the study-specific and combined proportions of surgery success reported in the Phase II surgical trials. Across the 4 studies, 38 cases underwent phase II procedures. The majority of these (34/38) were uniform simultaneous mandible and maxilla advancements (MMAs: standard LeFort I, sagittal split osteotomies) and 4 were mandible advancements (MAs) only. This procedural difference is study specific, thus allowing comparisons to be made within the Phase II forest plots. Table 1 also includes the degree of mandible or maxilla advancement for the Phase II procedures. Although the plots in Figure 2 demonstrate a small amount of between-study heterogeneity, a fixed-effects model was used to pool the proportions due to the small number of studies available. According to the “traditional” definition (ie, a 50% reduction in AHI, an AHI of ≤ 20, or both), the pooled success rate for Phase II surgical procedures was 86% (95% CI: 0.74, 0.95). With success defined as a post-surgery AHI of ≤ 10, pooled success is reduced to 45% (95% CI: 0.30, 0.60), and, at AHI ≤ 5,

Table 2—Phase I Subgroup Analyses Highlighting “Success” According to Various Definitions And I² Statistics (for Heterogeneity)

<table>
<thead>
<tr>
<th>Sub-Group</th>
<th>(# of studies)</th>
<th>Success = 50% reduction in AHI and/or ≤ 20</th>
<th>Success = AHI ≤ 10</th>
<th>Success = AHI ≤ 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Successful</td>
<td>F²</td>
<td>% Successful</td>
<td>F²</td>
</tr>
<tr>
<td></td>
<td>95% C.I.</td>
<td></td>
<td>95% C.I.</td>
<td></td>
</tr>
<tr>
<td>UPPP</td>
<td>178 (7)</td>
<td>51.5</td>
<td>81.3*</td>
<td>34.1</td>
</tr>
<tr>
<td></td>
<td>(44.3 – 58.7)</td>
<td>(56.6 - 89.2)</td>
<td>(27.5 – 41.1)</td>
<td>(73.6 - 91.8)</td>
</tr>
<tr>
<td>LAUP</td>
<td>72 (3)</td>
<td>48.8</td>
<td>82*</td>
<td>19.2</td>
</tr>
<tr>
<td></td>
<td>(37.6 – 60)</td>
<td>(0 - 92.3)</td>
<td>(11.1 – 28.8)</td>
<td>(0 - 72.9)</td>
</tr>
<tr>
<td>TCRF</td>
<td>36 (2)</td>
<td>60.8</td>
<td>undefined</td>
<td>34.1</td>
</tr>
<tr>
<td></td>
<td>(45 – 75.5)</td>
<td></td>
<td>(20.1 – 49.7)</td>
<td></td>
</tr>
<tr>
<td>Varied</td>
<td>61 (2)</td>
<td>77.9</td>
<td>undefined</td>
<td>55.6</td>
</tr>
<tr>
<td></td>
<td>(66.9 – 87.2)</td>
<td></td>
<td>(43.2 – 67.5)</td>
<td></td>
</tr>
<tr>
<td>Phase I Combined</td>
<td>347 (14)</td>
<td>55</td>
<td>79.3*</td>
<td>31.5</td>
</tr>
<tr>
<td></td>
<td>(43.6 – 66.2)</td>
<td>(64.1 - 86.3)</td>
<td>(21.1 – 42.9)</td>
<td>(67 - 87)</td>
</tr>
</tbody>
</table>

F² describes the percentage of total variation across studies that can be attributed to heterogeneity. Phase I refers to combination of soft palate surgery (uvuolapatopharyngoplasty [UPPP], laser-assisted uvulopalatoplasty [LAUP]), hyoid suspension (HS), genioglossus advancement (GA), ± tonsillectomy and adenoidectomy; RFVR, radiofrequency volume reduction of soft tissue (eg, tongue base); TCRF, temperature-controlled radiofrequency volumetric reduction (tongue base); repose system, a titanium screw and a permanent suture anchors/stabilizes the tongue base to the inner mandibular cortex.

*Signifies significant Cochran’s Q Statistic.
success is further reduced to 43% (95% CI: 0.28, 0.58). The results of Phase II studies should be treated tentatively, however, due to the small number of studies available.

**DISCUSSION**

The statistical meta analyses highlight how surgical “success” decreases when new evidence-based criteria of success are applied. In light of the health outcomes data presented in this review, we argue that current policy and practice may indeed be lagging the increasing depth of evidence in this area. The obvious and problematic point is that a reduction in AHI to 20 or less still confers the status of at least mild to moderate OSA, and the number of recipients/patients who achieve even this modest reduction varies substantially. This has potentially significant implications both for the health outcomes of individuals who live with OSA and for what may be deemed clinically and cost-effective (efficacious) treatment. Although many surgery publications report statistically significant improvements in AHI, we believe a distinction must be drawn between statistical significance and clinical significance. There is relative ease with which differences in continuous outcomes can reach statistical significance. Yet, a 50% reduction in AHI, albeit statistically significant, does not necessarily support claims of an effective procedure. To call a procedure successful in this context, where significant syndrome, comorbid disease, and associated poor health may still be present, is potentially misleading and contrary to what may reasonably be considered efficacious clinical practice. This is particularly important when an alternative treatment modality such as CPAP or weight loss may be rejected or underutilized due to the perception or misperception of a surgical “cure.” We propose that, in all future surgical audits, in addition to reporting the statistical significance of findings, “objective cure” rates are also reported. That is, outcomes achieved based on AHI of ≤ 5, ≤ 10, or both.

We have also demonstrated that clinical heterogeneity is present in existing surgical studies; yet, we believe that the conclusions drawn from our results remain relevant and robust. Heterogeneity reduces the value of pooled summary measures, yet the overall trends in surgery success illustrated in the forest plots are still apparent and are in agreement with the conclusions made in this report. For example, if success is defined as AHI less than or equal to 5, not 1 phase I study reported a success rate of over 40%. Limitations in study quality is an issue we recommend should be addressed in future research by more thorough measuring and reporting of potential confounding factors, such as time spent supine at all PSG assessments, as well as BMI changes. Similarly, time of follow-up from surgery to PSG should be made uniform, though we acknowledge that this is often difficult in practice.

To date, authors of existing surgical audits have reported, “Both phase I and phase II procedures are effective in treating OSA” and, “…the modified MMA technique for the treatment of OSA is highly effective….” Such claims can influence clinical practice and policy, particularly as, for both of these examples, surgical success was reported at over 90% (ie, a 50% reduction in AHI, an AHI of ≤ 20, or both). However, at an AHI ≤ 5, success in these cases is revised down to 45% and 30%, respectively. These revised outcomes may have entirely different policy and practice implications. Our meta-analyses demonstrate that when the traditional definition is applied (ie, a 50% reduction in AHI, and AHI of ≤ 20, or both), the pooled success rate for Phase I procedures is 55% (45% fail). However, at an AHI ≤ 5, success is reduced to 13% (87% fail). Similarly, for Phase II procedures, success (fail) rates decrease from 86% (14%) to 43% (57%), respectively. Again, the policy and practice implications of this are substantial. The bar has been raised on what may be defined as “treatment success,” and this should ultimately impact on policy and practice. This discussion highlights an apparent gap in the translation of evidence into clinical practice. In Australia, there is limited policy guidance as to who should be offered various treatment options, under what circumstances, and by whom. These decisions occur according to individual assessments by general medical practitioners, physicians, and surgeons with their patients. Moreover, we do not currently know how much collective evidence is considered in these decisions nor, indeed, what the collective outcomes are for many who seek out, or are induced into, alternative therapies. We envisage that this issue will increasingly become more than merely an academic or clinical exercise in promulgating differences of opinion. Ultimately, this has potential to become a policy issue; hence, there is a need for the relative medical specialties to clarify notions of “treatment success” and “treatment effectiveness” so as they are no longer loosely defined and interchangeable. Before inter-specialty consensus on the efficacy and application of these treatments is achieved, the current state of evidence suggests that the findings described in this review should be used to inform practice and policy.
procedures can be achieved, intra-specialty consensus is required. Ultimately, this will deliver what is in the best interests of OSA cases around the world.

AUTHORS’ CONTRIBUTIONS

AE, JM, AMS, and JH jointly conceived of and conceptualized the article’s orientation and structure. AE carried out the initial literature review and statistical meta-analyses (see also acknowledgements). AE and JM completed the first draft of the manuscript. In subsequent drafts AE and AMS contributed expertise on sleep medicine content, and AE, JM and JH epidemiologic and health outcomes content. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

The authors would like to acknowledge and thank Mr. Thomas Sullivan and Dr. Amy Salter, Discipline of Public Health, The University of Adelaide, for their advice and assistance with the statistical analyses on this project. The authors would also like to acknowledge and thank the anonymous reviewers for their thoughtful and constructive comments on earlier versions of this report.

Declaration

This work was undertaken at The University of Adelaide and at The Queen Elizabeth Hospital, Australia. Adam Elshaug received funding for this project from The University of Adelaide Faculty of Health Sciences. The authors declare that they have no competing interests and that no conflicts of interests exist.

REFERENCES


Redefining Success in Airway Surgery for OSA—Elshaug et al


