Methodologic Assessment of Studies on Endoscopic Sinus Surgery

Judith E. Cho Lieu, MD; Jay F. Piccirillo, MD

Background: Functional endoscopic sinus surgery (ESS) has become the procedure of choice for surgical treatment of chronic rhinosinusitis. Does the published literature support the claims of greater efficacy than medical treatment alone or older sinus procedures?

Objective: To analyze the methodology of the published literature regarding the efficacy of ESS.

Data Sources: MEDLINE search for primary studies published in 1987 through 2001, written in the English language, reporting results on more than 100 patients, using the MeSH (medical subject headings) terms sinusitis [subheadings surgery or therapy] and endoscopy.

Study Selection: Of 512 studies initially identified, 29 studies met the inclusion and exclusion criteria for further assessment of methodologic criteria. An additional 6 studies were found when the reference lists of reviews or included studies were searched.

Data Extraction: Articles were evaluated for 4 core (inclusion/exclusion criteria, control group, intervention, and clear outcome measure) methodologic criteria essential to the determination of efficacy of ESS. Eight additional methodologic criteria were also used to rate the articles.

Data Synthesis: Three studies met all 4 core methodologic criteria. Only 4 of 35 studies used a control group to evaluate efficacy of ESS to treat chronic sinusitis. Mean number of criteria met was 7.2, with a range of 2 to 11.

Conclusions: Absence of a control group is the most important reason that studies are unable to scientifically assess the comparative efficacy of ESS to medical therapy or other sinus procedures. The methodologic criteria described here can be used to evaluate studies of interventions for other disorders.


The use of endoscopes for sinus surgery allowed Messerklinger to introduce mucosal-sparing techniques that focused on removing key areas of obstruction to allow the restoration of normal mucociliary function.1(p12) The term “functional endoscopic sinus surgery” was coined by Kennedy et al2 in 1985 to describe these techniques. Since then, many studies have been published reporting the results of endoscopic sinus surgery (ESS) for the treatment of chronic sinusitis and application of these techniques to treat other disease states. Most otolaryngologists now consider ESS to be the standard of care for treating chronic sinusitis that is not responsive to medical therapy, replacing older procedures such as Caldwell-Luc and external ethmoidectomy.

Cochrane3 defined 3 terms that are useful in describing treatments or interventions: efficacy, efficiency, and effectiveness. Efficacy asks the question, does it work? Does the treatment show better results than placebo or no treatment? Efficiency asks the question, does it work better than something else? Does the new treatment work better than standard therapy? Effectiveness asks the question, does it work in the real world? Does the new treatment work in the real world, outside the constraints of a clinical trial? In other words, are the results generalizable? By these definitions, efficacy, efficiency, and effectiveness cannot be determined by case series, but require study designs with control groups: case-control or cohort studies, or clinical trials.

Using Cochrane’s definition, is ESS an efficacious intervention for chronic sinusitis? Does the current medical literature demonstrate the efficiency of ESS over medi
cal therapy? Has the superiority of ESS over other sinus procedures for chronic sinusitis clearly been established? These questions still need to be answered because ESS has not cured all of chronic sinusitis; clearly, many patients continue to require long-term medical therapy or revision sinus surgery to manage their symptoms.

The aim of this study was to analyze the quality of the published literature regarding ESS as an intervention for chronic sinusitis using methodologic principles that enhance the applicability, accuracy, and reproducibility of clinical research. Our intent was to identify problems in the classification of medical data; to determine how data were observed and recorded and how recorded data were combined, arranged, analyzed, and reported; and to suggest standards that may improve future studies on ESS and chronic sinusitis.

**METHODS**

**DATA SOURCE**

A MEDLINE search of the medical literature between 1987 and 2001 was performed to identify studies reporting results of patients who underwent ESS for chronic sinusitis. The MeSH (medical subject headings) headings sinusitis [subheadings surgery or therapy] and endoscopy were used to find original articles published in 1987 through 2001. We then searched for additional original articles by looking at the references of the included studies or recent reviews. The search was meant to be representative, but not exhaustive.

**STUDY SELECTION**

Criteria for inclusion into this methodologic survey were that the articles were written in the English language, involved human subjects, and reported patient-based outcome results from ESS. A patient-based outcome was considered to be any outcome in which patient symptoms, signs, or test values were measured. Studies were excluded if they were review articles, focused on pediatric patients or one particular subcategory of patients with chronic sinusitis (such as asthma or cystic fibrosis), described surgical innovations or instrumentation, focused on infectious etiologies, or reported fewer than 100 patients in the study population.

The MEDLINE search identified 512 studies using the stated search terms. Twenty-nine articles met each of the stated inclusion and exclusion criteria. An additional 6 articles were identified when the references of recent reviews or these 29 studies were searched, bringing to 35 the total number of articles reviewed for this study.

**METHODOLOGIC CRITERIA**

**Core Criteria**

Methodologic criteria used to assess the studies were considered “core criteria” if they are essential to the determination of efficacy of ESS. Core criteria included (1) description of inclusion and exclusion criteria, (2) presence of a control or comparison group, (3) description of the intervention, and (4) well-defined outcome measure. Descriptions of these criteria follow.

**Inclusion and exclusion criteria:** Explicit criteria for both the inclusion and exclusion of patients allow for comparison of the study population with other groups of patients. Inclusion criteria may consist of specific surgical indications, diagnoses, or “all” consecutive patients who were treated for a condition during a specific time period. Exclusion criteria may consist of specific diagnoses (eg, cystic fibrosis or ciliary dysmotility syndromes), previous treatments (eg, sinus surgery), or age groups (eg, <18 years old). This criterion was considered met if either inclusion or exclusion criteria were listed.

The use of explicit inclusion or exclusion criteria helps to avoid susceptibility or selection bias, in which groups of patients are at baseline unequal in their susceptibility to an outcome. For instance, patients with cystic fibrosis who have chronic sinusitis are expected to fare much differently than patients with chronic sinusitis but no underlying disorder.

**Control group:** The gold standard control group for an intervention is a placebo or not-treated group. Because of ethical concerns about not offering therapy when it is available, a medically treated control group consisting of those who meet the indications for ESS would be an excellent alternative. However, a group of patients who underwent older surgical techniques such as the Caldwell-Luc procedure or transantral ethmoidectomy would also be a suitable control. Historical controls would be acceptable if the only factor that differentiated the historical group from the ESS groups was secular time (eg, the treatment protocols, surgical indications, surgeon[s], outcome evaluations, were the same). This criterion was considered met if the results of those receiving ESS were compared with any control group who did not undergo ESS.

Having a control group eliminates some interpretative bias that comes from not measuring the outcome in a group receiving an alternative therapy (placebo, medical, or other surgical). The best way to “show” an intervention works is to have no control group.

**Description of intervention:** Intervention involving ESS includes the preoperative evaluation and therapy (eg, prednisone prior to excision of nasal polyps), surgical technique, and the usual protocol for postoperative care (eg, antibiotics, cleansing, follow-up visits). These should all be described in adequate detail so that the study may be duplicated or compared. This criterion was considered met if at least the surgical technique was described.

**Description of the intervention can lessen performance or proficiency bias (in which how an intervention is performed can affect outcome), but not always eliminate it. Clearly, the technique used in performing sinus procedures is believed to be an important contributing factor to outcome, as is the surgeon performing the procedure. The protocols for preoperative and postoperative management may affect outcome as well.**

**Clear outcome measure:** Explicit subjective or objective outcome measures give sufficient detail about the criteria so that use of the outcome measure can be replicated. Subjective measures may include patient questionnaire, symptom scale follow-up sheet, or clinical rating by the surgeon or investigator. Objective measures may include endoscopic examination, acoustic rhinometry, computed tomographic scan, or ciliary beat frequency. This criterion was considered met if either a subjective or objective measure was used to report patient results. Credit was not given if criteria for the terms “improved, unchanged, and worse” were not defined.

The description and use of a clear outcome measure helps to lessen measurement bias. Ideally, the measurement of outcome uses a reliable, accurate, valid, and objective scale. For subjective outcomes, techniques can be used to “harden” data and make them more reliable, accurate, and objective. Describing the outcome measure and the techniques used to eliminate subjectivity helps to assess how much bias may still remain.

**Additional Criteria**

Other methodologic criteria used to assess these studies were considered “additional criteria,” which are important in de-
scribing the patients studied and their baseline clinical state, describing the design and methods, and describing the outcome events. These criteria help to identify and reduce bias that could affect the results of these studies, and help to determine the generalizability of a study's results. These criteria are described in further detail as follows:

**Patient demographic description:** Common demographic variables often used to define the patient population include age, sex, race/ethnicity, educational level, and income or occupation. Description of age should include the mean or median as well as the range. Since the anatomy of children is clearly different from that of adults, and other physiologic parameters are still developing, studies of children should be separated from that of adults. This criterion was considered met if at least age or sex were described.

**Description of comorbidities:** Comorbid conditions may affect the prognosis and outcome independent of the illness of interest. Sinusitis-related comorbidities often quoted include asthma, atopic disease, nasal polyposis, Samter’s triad, immunodeficiency, cystic fibrosis, and ciliary dysfunction. Other general medical comorbid conditions, such as diabetes, renal or hepatic failure, or major affective disorder could potentially affect the outcome of therapy for sinusitis. Family history, genetic predisposition, or habits such as smoking could also potentially be considered comorbidities. This criterion was considered met if any comorbidities of the patients were described.

**Staging system:** Since most medical conditions have a spectrum of disease from mild to severe, staging systems are valuable in comparing patients of different levels of severity. A patient with recurrent facial pressure from a unilateral maxillary sinusitis has a clearly different spectrum of disease than a patient with recurrent nasal airway obstruction and anosmia from bilateral nasal polyposis. Staging systems may be based on diagnoses, imaging studies, clinical symptoms, prognoses, or other factors. This criterion was considered met if any such stratification of disease severity was used.

**Inception cohort:** An inception cohort is a study population demarcated by date, place (hospital, clinic, or surgeon), visit or intervention (eg, initial visit or date of surgery), and a similar point in their disease course. These 4 components define the “zero-time” from which follow-up time is determined. This criterion was considered met if 2 of the 4 components were included in the study’s description, and at least 1 of those components was a “similar point in their disease course.”

**Adequate length of follow-up:** Since chronic sinusitis often has a seasonal component with acute exacerbations and a tendency to recur, a minimum of 12 months’ follow-up on all patients should be obtained. A consistent protocol for follow-up at regular intervals increases the credibility of an outcome if a trend can be demonstrated over these periods. This criterion was considered met if the mean or median follow-up period was at least 12 months.

**Statistical analysis:** Univariate (or descriptive) analysis (eg, mean, median, percentages, ranges) is the most elementary of statistical analyses. Bivariate analysis looks for the effect of patient, treatment, and other variables on the outcome of interest, and multivariate analysis evaluates the effect of multiple variables (such as treatment groups, preoperative and postoperative subjective scores, possible prognostic indicators, or risk factors) interacting with other on the outcome of interest. This criterion was considered met if at least bivariate analysis was performed.

**Missing data:** Missing data may bias the results of a study because the investigator usually does not know whether the missing patients are doing well, not requiring further medical treatment, or doing poorly and seeking treatment elsewhere. This criterion was considered met if the study mentioned how many patients were “lost to follow-up,” or if all patients were followed up.

### RESULTS

The 35 studies reported that ESS was performed on 10,149 patients. The following tabulation shows the number of core methodologic criteria met by the studies. Only 3 (8%) studies met all 4 “core methodologic criteria.”

<table>
<thead>
<tr>
<th>No. of Core Criteria Met</th>
<th>No. (%) of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3 (8)</td>
</tr>
<tr>
<td>3</td>
<td>10 (28)</td>
</tr>
<tr>
<td>2</td>
<td>21 (60)</td>
</tr>
<tr>
<td>1</td>
<td>1 (3)</td>
</tr>
</tbody>
</table>

Table 1 shows the frequency of studies that met each criterion. The criterion “clear outcome measure” was met by the greatest number of studies (33), while the criterion “control group” was met by the least (4).

<table>
<thead>
<tr>
<th>Methodologic Criterion</th>
<th>No. (%) of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion and exclusion criteria</td>
<td>29 (83)</td>
</tr>
<tr>
<td>Control group</td>
<td>4 (11)</td>
</tr>
<tr>
<td>Description of intervention</td>
<td>30 (86)</td>
</tr>
<tr>
<td>Well-defined outcome measure</td>
<td>33 (94)</td>
</tr>
<tr>
<td>Patient demographic description</td>
<td>27 (77)</td>
</tr>
<tr>
<td>Description of comorbidities</td>
<td>21 (60)</td>
</tr>
<tr>
<td>Staging system</td>
<td>13 (37)</td>
</tr>
<tr>
<td>Inception cohort</td>
<td>11 (31)</td>
</tr>
<tr>
<td>Length of follow-up</td>
<td>16 (46)</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>15 (43)</td>
</tr>
<tr>
<td>Missing data</td>
<td>24 (68)</td>
</tr>
<tr>
<td>Description of complications</td>
<td>28 (80)</td>
</tr>
<tr>
<td>Total</td>
<td>35 (100)</td>
</tr>
</tbody>
</table>

©2003 American Medical Association. All rights reserved.
Only 4 of the 35 studies surveyed were not case series. One study compared a group receiving only medical therapy with a group who underwent ethmoidectomy.14 Another study compared middle turbinate resection vs middle turbinate preservation in the setting of ESS.15 Two studies compared Caldwell-Luc procedures with ESS.28,29

Description of all 3 components of the intervention (preoperative evaluation/therapy, surgical technique, and postoperative care) occurred in 13 of the 30 studies that met this criterion.

All but 2 of the 33 studies that met the criterion “clear outcome measure” used patient-based questionnaires or scores. Eleven studies reported use of endoscopy to measure outcome, while 3 used other objective tests. Other outcome measures included a quality-of-life measure, percentage of patients who needed revision surgery, and numbers of antibiotic courses for recurrent sinusitis per year.

Patients’ ages ranged from 14 months to 87 years. Of the 26 studies that gave age as a demographic descriptor, 12 included patients younger than 13 years. Either the numbers of males and females or their ratios were documented in 18 studies.

Most of the studies (21) listed sinusitis-related comorbidities. Five studies considered other medical conditions, habits, previous therapies, or family history.

In the 13 studies that used any form of staging system, 7 used sinus computed tomographic scans at least in part. The others staged patients by disease (eg, nasal polyposis vs chronic sinusitis), by endoscopic findings, or by clinical severity.

The “inception cohort” criterion was usually met by the demarcation of dates and place or surgeon. Eleven gave sufficient detail to determine that patients were at a similar point in their disease course, most commonly a lack of satisfactory improvement after a stated minimum length of medical therapy.

Follow-up periods for all 35 studies ranged from 1 month to 10 years. Of the 16 studies that met the minimum standard for this criterion, 12 followed up all patients for at least 1 year.

Four of the 15 studies that met the minimum standard for statistical analysis attempted multivariable analysis, to control for the numerous patient and clinical factors that could affect outcome. One study clearly performed bivariate statistical tests, but failed to report which test was used.

The criterion of “missing data” was usually met by the authors stating that follow-up was obtained in all patients, or giving the numbers lost to follow-up. No study discussed the possible bias that could result from not having these follow-up results, or attempted to compare the baseline characteristics of those patients who reported postoperative outcomes from those who did not return for follow-up.

Complications were usually discussed in detail, with reports of the actual numbers of patients with each complication and their resolution. Authors differed in what they considered a complication, especially for “minor” problems such as synechiae and middle meatus stenosis.

We found that the overwhelming majority of articles reviewed in our survey (92%) failed to meet all 4 core methodologic criteria that are essential to the determination of efficacy of ESS. Furthermore, although we deliberately used liberal standards for meeting each of the methodologic criteria, 13 (37%) of 35 met half or fewer of the 12 criteria we used to rate the articles. The lack of a control group is the most important reason that studies currently in the literature are unable to scientifically compare the efficacy of ESS with that of medical therapy or other sinus procedures. However, it does appear that the number of criteria met seems to be increasing with time. The findings of this study are similar to other otolaryngologic reviews of methodologic quality.41,42

In the 33 articles surveyed for this study, patient improvement associated with ESS ranged from 68.9% (“good outcome”)51 to 94% (“at least 50% improvement”).25 Due to the heterogeneity of outcome responses, it is of course impossible to synthesize these results into a single summary number. Several authors suggest that comorbidity plays an important part in predicting outcome, and variability in type and severity of comorbidity in the different studies may account for the wide variation in results reported.
Absence of a control group is the most important reason that studies are unable to scientifically assess the comparative efficacy of ESS to medical therapy or other sinus procedures. More attention to the description of the baseline patient state, description of the intervention, and description of the outcome events can improve the general methodologic quality of future reports on ESS for sinusitis. These methodologic criteria can also be used to evaluate studies of interventions for other disorders, particularly in situations where randomized clinical trials are not feasible or not considered ethical.

Submitted for publication November 8, 2002; final revision received January 30, 2003; accepted February 18, 2003.

Corresponding author: Judith E. Cho Lieu, MD, Department of Otolaryngology—Head and Neck Surgery, Washington University School of Medicine, One Children's Place, Room 35 35, St Louis, MO 63110 (e-mail: lieuj@msnotes.wustl.edu).

REFERENCES


CONCLUSIONS

Of the 3 studies that met all 4 core criteria, one study compared middle turbinate resection with middle turbinate preservation in patients undergoing ESS, so that the determination of efficacy of ESS itself was not the main objective. The investigators found that for patients with more severe disease (stages 3 and 4), middle turbinate resection resulted in fewer revision operations than middle turbinate preservation with a mean follow-up time of 50 months. The other 2 studies meeting all 4 core criteria compared the Caldwell-Luc procedure with ESS, and found equivocal results. Narkio-Makela and Qvarnell found equivocal results. Narkio-Makela and Qvarnell compared the Caldwell-Luc procedure with ESS, and found the determinant of efficacy of ESS itself was not the main objective. The investigators found that for patients with more severe disease (stages 3 and 4), middle turbinate resection resulted in fewer revision operations than middle turbinate preservation with a mean follow-up time of 3.8 years. Penttilä et al reported the superiority of ESS over Caldwell-Luc at 1 year postoperatively, with patients evaluating themselves as globally “markedly improved” in 50.7% of the Caldwell-Luc group and in 76.7% of the ESS group. However, at 5 to 9 years postoperatively, 82% of Caldwell-Luc and 76% of ESS patients reported this outcome.

In the single study that compared medical therapy with surgical therapy, the groups of patients were not equivalent at baseline in their use of medications, quality-of-life scores, or comorbid illnesses; the nonsurgical group was closer to normal (as defined by a quality-of-life instrument) than the surgical group. The surgical group was also followed for a longer time (6-12 months vs 3 months for the nonsurgical group). These differences in the baseline state and follow-up between medical and surgical groups do not allow for definite conclusions about the relative efficiency of these interventions for chronic rhinosinusitis in this study population.

The main limitation of our study is that we missed some articles on rhinosinusitis and ESS because we limited our search to MEDLINE only and used a single search strategy. However, this study was not meant to be a meta-analysis or systematic review of the literature, so that our search for articles was not exhaustive. Because we examined the references of recent reviews and the studies that were found with our search, it is unlikely that we missed many additional studies that would have significantly altered our results and conclusions.

The main reason to go through the trouble of adhering to methodologic standards is to eliminate bias. In the “Methods” section we mentioned very briefly how each of the core criteria helps to eliminate bias. Cohort (observational) (level 2 evidence) and case-control (level 3 evidence), studies on this subject would provide much higher levels of evidence than another case series (level 4 evidence), primarily because they include a comparison or control group. Advocating the use of control groups does not imply that every study has to be a randomized controlled trial. Well-done cohort and case-control studies can offer evidence in deciding about efficacy, efficiency, or effectiveness, without the ethical, logistical, and methodologic hurdles of a randomized controlled trial. Thus, we caution against a rigid insistence that only randomized controlled trials can provide evidence to improve patient care.


