Meta-analysis in Otolaryngology

Ramsey Alsarraf, MD, MPH; Nicole W. Alsarraf, MES; B. Maya Kato, MD; Neal D. Goldman, MD

Objective: To examine the results of meta-analyses in otolaryngology and compare these results with the individual component studies that constitute each meta-analysis.

Design: A retrospective review of the literature.

Main Outcome Measures: Studies that conducted pooled statistical systematic analyses indexed on MEDLINE for the 10-year period from January 1989 to January 1999 were selected for keyword or subject headings of meta-analysis and otolaryngology (N = 22). Analysis consisted of a modified funnel graph depiction of the individual studies that made up each meta-analysis. Each meta-analysis was evaluated for consistency among these individual studies and comparison of the median result with the weighted mean meta-analysis result. In addition, the methodologic quality of each meta-analysis was assessed in terms of the rigor with which component studies were evaluated.

Results: Ten (46%) of the 22 meta-analyses did not provide the individual study results that made up their meta-analyses. The results of 10 studies (46%) were similar to the median result of their individual component studies. The results of 2 studies (9%) differed from this median result, with widely heterogeneous component study results.

Conclusions: A large proportion of meta-analyses in otolaryngology (46%) fail to provide the individual study results necessary to analyze the meta-analysis result critically. Most remaining studies do provide results that accurately compare with the median of their component study results. Only a small proportion of meta-analyses were found to have disparate results, and each appropriately discusses the heterogeneity of the individual studies that comprise their meta-analysis.

Arch Otolaryngol Head Neck Surg. 2000;126:711-716

META-ANALYSIS can be defined as the statistical pooling of (published and unpublished) results available from different studies on the same topic in a systematic manner. Meta-analysis differs from traditional literature review, because it can provide a more objective or quantitative result compared with a simply descriptive or narrative compilation. The goal in meta-analysis is to comprehensively review the available evidence on a topic and provide the reader with an assessment of the combined trend that has been described. This assessment may take the form of a single, pooled result that stems from the combining of each of a series of study populations and evaluating the total group in terms of some common end point. The benefit of such pooling is the fact that many smaller studies that may not have obtained statistical significance may be combined in such a manner that the trend that results may itself now have the power to show such significance. In addition, such systematic review may serve to highlight existing clinical controversies that require further investigation.

There have been many criticisms of this method of analysis. Feinstein refers to meta-analysis as “statistical alchemy.” Bailar questions the validity of such pooling and contends that “any attempt to reduce the results [of meta-analysis] to a single value . . . is likely to lead to conclusions that are wrong, perhaps seriously so.” Sharpe provides a summary of many of these criticisms, each of which focuses on the basis of meta-analysis as a form of study or the theory that underlies meta-analysis. These theoretical issues in meta-analysis include the following: (1) there is often a bias toward including only published studies (the “file drawer” issue); (2) it is difficult to validly combine results from studies with different populations, with dif-
METHODS

A retrospective search was performed for the 10-year period from January 1989 to January 1999 of articles indexed on MEDLINE. Keyword and subject listings were searched for meta-analysis and otolaryngology. Although this approach misses articles not indexed in this way, our search was limited in this fashion to avoid an artificial bias in choosing among various subject headings. A total of 32 articles were found; 22 of these were studies in which actual pooled statistical evaluations (meta-analyses) had been undertaken.7–28 The remaining 10 articles were either discussions about meta-analysis or simple literature reviews that did not pool individual study results to obtain a single meta-result. These literature reviews were not included in our analysis.

Each meta-analysis was evaluated by means of a modified funnel graph diagram analysis for the accuracy with which the meta-analysis result characterized the overall trend of the individual studies that made up each meta-analysis study. Funnel graph diagram analysis, or a scatterplot of results, is the method by which a group of individual studies can be compared in terms of their distribution by study population size and is a method of critically analyzing meta-analyses that several authors have recommended.3,6 This form of analysis is a simple, graphic way to allow the reader to inspect the consistency and homogeneity of a grouping of studies without having to perform cumbersome statistical analyses. Our modification of this method is simply the use of bar graphs rather than a scatterplot to aid in the visual simplicity of understanding the distribution of individual component study results. The figures provided in the text thus have a y-axis representing study size (n) and an x-axis representing the effect or result of each individual study (eg, odds ratio or percentage). The following features of a meta-analysis can be evaluated in this manner: (1) the range of individual study results, (2) the comparison of the meta-result with the result of the largest previous study (mode), (3) the comparison of the meta-result with the median result of this population of studies, (4) the presence of a skewed distribution of studies or multiple rather than simply 1 mode, and (5) the presence of atypical outliers that might influence meta-analysis results. Our focus was the comparison of the meta-result with the median result of the group of component studies, as this median result appears to reflect the most stable indicator of what is accepted in the given literature.

Our study evaluated these 22 meta-analyses in otolaryngology to determine the accuracy with which each meta-result characterized the results of the individual studies in each meta-analysis, as represented by this median result. Meta-analyses that provided the results of these individual studies were categorized as providing a result that either did or did not differ markedly from the result of the median of studies already in the otolaryngology literature. Meta-analyses whose results differed from this median result were evaluated to determine the distribution of individual studies and the heterogeneity that may be present between their individual component studies. In addition to this numerical analysis, each meta-analysis was evaluated to determine the degree to which the authors addressed such heterogeneity, evaluated the quality of individual studies, or weighed other methodologic issues.

This analysis was conducted using the SPSS computer package (SPSS Inc, Chicago, Ill); simple graphic representations were obtained using Microsoft Excel and Powerpoint programs (Microsoft Corp, Redmond, Wash).

RESULTS

Investigating the methods of the meta-analyses evaluated in this study revealed that more than half investigated the quality of the individual studies that comprised their meta-analysis.

<table>
<thead>
<tr>
<th>Method</th>
<th>No. (%) of Metas-analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigated individual study quality</td>
<td>13 (59)</td>
</tr>
<tr>
<td>Discussed heterogeneity of component studies</td>
<td>9 (41)</td>
</tr>
<tr>
<td>Weighed studies by quality</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Used both published and unpublished studies*</td>
<td>9 (41)</td>
</tr>
<tr>
<td>Used unpublished data other than own study</td>
<td>0</td>
</tr>
</tbody>
</table>

*Unpublished data from own study only. Two meta-analyses used statistical techniques to assess the degree of possible publication bias, although they did not find any unpublished data to include in their meta-analyses.

A smaller percentage (41%) discussed the heterogeneity of these component studies, and only 18% weighed studies by some quality scale. Although approximately 40% of the meta-analyses evaluated used some form of...
ponent studies that approximated a normal curve with little skewness and no atypical outliers (Figure 1). In addition, the meta-analysis result of this study was similar to the median result of the funnel graph diagram analysis, with only an 11% change from the median to the meta-result (Table). This meta-analysis provided a result that differed from the largest previous study (mode) yet resulted in a weighted mean that appeared to accurately characterize the group of individual studies when examined collectively by the median result.

In contrast to the studies shown in the Table, there were 2 meta-analyses identified in which the meta-result differed markedly from the median result of the individual component studies. Unlike studies with near-normal distributions, the remaining 2 meta-analyses were found to be composed of a group of widely heterogeneous individual studies, with very wide ranges, atypical outliers, and meta-results that did not appear to describe a collective trend. These 2 studies were the meta-analyses of Shatkin et al and Rosenfeld et al. Each of these studies, when examined graphically, revealed skewed distributions with groups of atypical outliers (Figure 2) or multiple modes (Figure 3). Furthermore, the results of these meta-analyses differed greatly from the median results of our funnel graph analysis. One study’s meta-result was 54% less than the median result for the range of studies, whereas the other study’s meta-result was 75% greater than its similar median result. These 2 meta-analyses appear to combine heterogeneous individual study results, and, in fact, the authors appropriately discuss this fact in their articles. For example, in the study by Shatkin et al, there is a group of studies with a low level of mucosal allergy (0%-12%) and a separate group of studies with a high level of mucosal allergy (56%-60%). It is possible that these individual studies cannot be pooled for an accurate meta-analysis result (of 21%) if they represent different answers that need to be further explained. Similarly, the study by Rosenfeld et al presents a wide range of odds ratios (0.8-17.9) with 3 unpublished data in addition to published sources, none of these studies used any unpublished material other than their present study.

Of the 22 articles we found indexed in MEDLINE that contained meta-analyses during this 10-year period, 10 (46%) did not provide the results of the individual studies that were included in their meta-analysis (data from the study by Stell25 were contained in a separate article that was not identified by the MEDLINE search criteria used in this analysis).7,8,11,17,19,22,24-26 No further analysis could be conducted on these studies in terms of the consistency of these individual studies or the accuracy of the meta-analysis result. Unfortunately, the reader is thus left with no means of critical analysis in approximately half of the meta-analyses in the otolaryngology literature identified in this manner.

The remaining 12 articles provided the results of the individual studies that constituted their meta-analyses. Among this group of studies, 10 meta-analyses (46%) had results that were similar to the median result of those component studies in the literature (Table). The distributions of the individual studies that made up 8 of these meta-analyses approximated a normal distribution, with several smaller studies that did not appreciably influence the weighted mean meta-analysis result and a mode (largest study size) that approximated the median result. The distribution of 2 of these 10 studies also approximated a normal curve, but with a mode (largest study size) that differed from the median result significantly.15,28 Despite this difference, the meta-result accurately characterized the median of the overall group of component studies in the literature. These meta-analyses were found to be composed of studies with a relatively narrow range and provided meta-results that appeared to accurately characterize this range of individual results. For example, graphic evaluation of the study by Haughey et al14 revealed a distribution of component studies that approximated a normal curve with little skewness and no atypical outliers (Figure 1). In addition, the meta-analysis result of this study was similar to the median result of the funnel graph diagram analysis, with only an 11% change from the median to the meta-result (Table). This meta-analysis provided a result that differed from the largest previous study (mode) yet resulted in a weighted mean that appeared to accurately characterize the group of individual studies when examined collectively by the median result.

In contrast to the studies shown in the Table, there were 2 meta-analyses identified in which the meta-result differed markedly from the median result of the individual component studies. Unlike studies with near-normal distributions, the remaining 2 meta-analyses were found to be composed of a group of widely heterogeneous individual studies, with very wide ranges, atypical outliers, and meta-results that did not appear to describe a collective trend. These 2 studies were the meta-analyses of Shatkin et al and Rosenfeld et al. Each of these studies, when examined graphically, revealed skewed distributions with groups of atypical outliers (Figure 2) or multiple modes (Figure 3). Furthermore, the results of these meta-analyses differed greatly from the median results of our funnel graph analysis. One study’s meta-result was 54% less than the median result for the range of studies, whereas the other study’s meta-result was 75% greater than its similar median result. These 2 meta-analyses appear to combine heterogeneous individual study results, and, in fact, the authors appropriately discuss this fact in their articles. For example, in the study by Shatkin et al, there is a group of studies with a low level of mucosal allergy (0%-12%) and a separate group of studies with a high level of mucosal allergy (56%-60%). It is possible that these individual studies cannot be pooled for an accurate meta-analysis result (of 21%) if they represent different answers that need to be further explained. Similarly, the study by Rosenfeld et al presents a wide range of odds ratios (0.8-17.9) with 3
very different groupings. The meta-result (of 3.6) may not accurately characterize these 3 distinct groups of studies, because they may be far too heterogeneous to be combined in this manner. It is possible that there simply is no obvious pooled answer, given the individual studies that exist on this topic. In this way, these meta-analyses may highlight the clinical controversy in these fields of otolaryngology.

The following is a summary of the classifications of these 22 meta-analyses.

<table>
<thead>
<tr>
<th>Classification</th>
<th>No. (%) of Meta-analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual study results not provided</td>
<td>10 (46)</td>
</tr>
<tr>
<td>Results unchanged from median</td>
<td>10 (46)</td>
</tr>
<tr>
<td>Result of (consistent) component studies</td>
<td></td>
</tr>
<tr>
<td>Meta-result different from median result</td>
<td>2 (9)</td>
</tr>
<tr>
<td>(heterogeneous studies)</td>
<td></td>
</tr>
</tbody>
</table>

A large proportion of studies (46%) unfortunately provide too little information with regard to the component studies for any critical analysis to be performed. A similar number of meta-analyses provide a meta-result that appears to accurately characterize the overall trend that exists in the given literature, as represented by the median result of the individual studies grouped together. Only 2 studies (9%) appear to provide a meta-result of a widely heterogeneous distribution of individual studies that differs significantly from the median result of those studies.

The goal of this study is to provide the clinician with a simple method to evaluate a given meta-analysis in the otolaryngology literature. Thus, theoretical issues related to the quality or validity of any specific meta-analysis were not specifically addressed in our evaluation, although we did evaluate the degree to which each meta-analysis investigated the heterogeneity or quality of its individual component studies (Table).
The results of our study show that a large proportion of meta-analyses (46%) in otolaryngology indexed on MEDLINE during the last decade fail to provide the reader with enough information to evaluate the meta-result in any comprehensive fashion. These studies should be criticized for not providing the individual study results that are essential if the reader is to assess the consistency of these results or evaluate the accuracy of the meta-result when compared with these individual studies. It is difficult to accept a meta-analysis result as valid when there is no evidence provided as to the distribution of study results from which this result has been obtained.3

Of the remaining studies, the majority (46%) provide a meta-result that agrees favorably with the median result of the individual studies that have been done, whether or not this median result is represented by the largest study in the existing literature. These studies do indeed provide the reader with the results of each of the individual studies that constitute the meta-analysis. In each of these studies, however, there appears to be an obvious trend established in the existing literature on each topic, with one or several larger studies influencing this collective trend. For instance, the recent study by Hebert and Bent15 on the outcome of pediatric functional endoscopic sinus surgery presents a meta-analysis of 8 individual studies in addition to unpublished data from their own institution. The largest study (n = 500) found a “positive outcome” in 88% of patients, whereas the meta-result for all patients combined (n = 882) similarly had a positive result in 87% of patients. The remaining studies had 77, 44, 24, 24, 21, and 18 patients, with 50 patients from their unpublished data. These smaller studies had a range of positive outcome from 77% to 100%, and the total group approximates a normal distribution. Given this relatively narrow range and the strong influence of the large study included in their meta-analysis, the value of performing a pooled analysis in this setting is that it confirms the collective trend and may provide the statistical significance of the median effect size that the reader would not obtain from a narrative literature review.

The value of this sort of meta-analysis is demonstrated in the article by Hebert and Bent.15 These authors provide a wealth of information regarding the specific aspects that were reviewed in selecting the final 8 articles of the meta-analysis. These inclusion criteria included the number of patients, average follow-up, study design, and comorbidities. After a comprehensive weighting, 6 studies are excluded from their analysis, although 3 of these contain patients included in the large study that greatly influences their meta-result. Unfortunately, the results of these excluded studies are not available for the reader’s review; however, it is clear that the authors have made a significant attempt to focus their meta-analysis on the most consistent, highest-quality individual studies in the otolaryngology literature to “create a consensus of outcomes.”15

There were only 2 meta-analyses20,23 that provided results that differed from the median result of their component studies, and in both cases the authors provide the reader with a discussion of the heterogeneity of these study populations. As shown in Figures 2 and 3, if a group of individual studies are severely skewed or grouped into multiple modes across an inordinately wide range of results, with a marked discrepancy between the median and weighted mean results, it may be that there is no given meta-result that truly reflects whatever trend may or may not be described. Rather, this form of heterogeneity either seems to reflect specific differences in study design, population, or outcome measurements among the het-
erogeneous studies or simply the fact that no given trend has been described to date. The answer in these instances may not be a pooled meta-analysis but, in contrast, further individual studies that do indeed maintain some degree of homogeneity in their design, follow-up, and outcomes.

The authors of these 2 studies comment on this problem. Shatkin et al\textsuperscript{23} discuss that “inhomogeneity” leads to wide confidence intervals in their results, which is likely due to “differences in patient populations, testing procedures, or both of these.” Rosenfeld et al\textsuperscript{20} comment that “when significant heterogeneity is present, it is unlikely that the observed variations . . . are due solely to chance,” and go to great lengths to describe the differences between these heterogeneous studies, and thus conclude that they “cannot presently recommend [the use of oral steroids] . . . until more information is known about the treatment.” Each of these studies does, however, provide an excellent review of the literature and illustrate that further work is needed on these topics before an accurate conclusion may be reached.

One limitation of our study is the focus only on articles indexed on MEDLINE with the keyword or subject listing of meta-analysis or otolaryngology. There are some meta-analyses on topics in otolaryngology that are certainly missed in this way; however, our objective was to survey studies most readily available to the general reader of the otolaryngology literature during this period and provide one method of analysis that may aid the reader in evaluating meta-analyses in the future.

Meta-analysis is a valuable tool that provides a more objective and quantitative evaluation of a grouping of studies on a given topic compared with the more descriptive nature of traditional narrative literature review. Some authors contend that a pooled meta-analysis result may misrepresent the overall trend of such a group of individual studies.\textsuperscript{2} Our study illustrates the practical accuracy of 22 meta-analyses indexed on MEDLINE in otolaryngology from 1989 to 1999. A large proportion (46%) of these studies do not provide the reader with the individual study results that are required for any such critical evaluation. Of the remaining meta-analyses, only a small percentage (9%) provide information that disagrees with a given trend in the otolaryngology literature as represented by the median result of the individual component studies. The use of a modified funnel graph diagram analysis is a simple means of evaluating the consistency of a group of studies that make up a given meta-analysis and the accuracy of a meta-analysis result. The clinician should exercise caution when interpreting the results of any meta-analysis. Further studies are needed that investigate more closely those theoretical issues that surround the evaluation of meta-analysis as a statistical method; however, this study attempts to provide an initial assessment of the work of meta-analysis in otolaryngology during the last decade.

Accepted for publication January 11, 2000.

Reprints: Ramsey Alsarraf, MD, MPH, Box 356515, Department of Otolaryngology—Head and Neck Surgery, University of Washington School of Medicine, Seattle, WA 98195 (e-mail: amro@u.washington.edu).

REFERENCES


©2000 American Medical Association. All rights reserved.