Recent Randomized Controlled Trials in Otolaryngology

Sarfaraz M. Banglawala, MD1, Lauren A. Lawrence, MD1, Emily Franko-Tobin1, Zachary M. Soler, MSc, MD1, Rodney J. Schlosser, MD1, and John Ioannidis, MD, DSc2,3,4,5

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

Abstract

Objective. To assess recent trends in the prevalence and quality of reporting of randomized controlled trials (RCTs) in 4 otolaryngology journals.

Study Design. Methodology and reporting analysis.

Setting. Randomized controlled trials in 4 otolaryngology journals.

Subjects and Methods. All RCTs published from 2011 to 2013 in 4 major otolaryngology journals were examined for characteristics of study design, quality of design and reporting, and funding.

Results. Of 5279 articles published in 4 leading otolaryngology journals from 2011 to 2013, 189 (3.3%) were RCTs. The majority of RCTs were clinical studies (86%), with the largest proportion consisting of sinonasal topics (31%). Most interventions were medical (46%), followed by surgical (38%) and mixed (16%). In terms of quality, randomization method was reported in 54% of RCTs, blinding in 33%, and adverse events in 65%. Intention-to-treat analysis was used in 32%; P values were reported in 87% and confidence intervals in 10%. Research funding was most often absent or not reported (55%), followed by not-for-profit (25%).

Conclusions. Based on review of 4 otolaryngology journals, RCTs are still a small proportion of all published studies in the field of otolaryngology. There seem to be trends toward improvement in quality of design and reporting of RCTs, although many quality features remain suboptimal. Practitioners both designing and interpreting RCTs should critically evaluate RCTs for quality.

Keywords
evidence-based medicine, randomized control trial, clinical trial, quality of evidence, otolaryngology journals

The quantity of research is increasing in otolaryngology and in medicine overall.1,2 Research output has been reported to approximately double every 7 years.3 Such a voluminous quantity of studies can be limiting when physicians cannot assess the quality of the research, synthesize it, and apply it to patient care. Evidence-based medicine (EBM) has provided a tool to evaluate research, and randomized controlled trials (RCTs) are considered the gold standard for appraising the benefits and risks of interventions as they are designed to minimize error and bias. It is important for otolaryngologists to be aware of the makeup of the general body of otolaryngology literature, as well as the quality of the design and reporting of RCTs in the field.

Not all RCTs are created equally. Despite improvements in the understanding of the optimal design and reporting4 of RCTs over the past decades (eg, as supported by the Consolidated Standards of Reporting Trials [CONSORT] effort), many RCTs follow suboptimal design and are poorly reported; inferences based on them may be spurious.

A number of previous empirical studies have reported on the quantity and quality of the body of otolaryngology literature, focusing also explicitly on RCTs.1,2,5-8 A number of authors have reported an increase in the quality of otolaryngology research, noting that more studies with a higher level of evidence, such as RCTs or systematic reviews, are being published.5,6 Others have specifically analyzed RCTs, comparing them with CONSORT recommendations, and have concluded that their reporting is not improving and is

1Department of Otolaryngology–Head and Neck Surgery, Medical University of South Carolina, Charleston, South Carolina, USA
2Department of Medicine, Stanford Research Prevention Center, Stanford University, Stanford, California, USA
3Department of Health Research and Policy, Stanford University School of Medicine, Stanford University, Stanford, California, USA
4Department of Statistics, Stanford University School of Humanities and Sciences, Stanford University, Stanford, California, USA
5Meta-Research Innovation Center at Stanford, Stanford University, Stanford, California, USA

Corresponding Author:
Sarfaraz M. Banglawala, Department of Otolaryngology–Head and Neck Surgery, Medical University of South Carolina, 135 Rutledge, MSC-550, Charleston, SC 29425, USA.
Email: sirf03@yahoo.com
not up to optimal standards.\textsuperscript{7,8} This is in contrast to the general body of medical literature, which has seen improvements in RCT reporting since the release of the CONSORT statement.\textsuperscript{9,10} One analysis used the 5 highest impact internal medicine journals conforming to CONSORT standards,\textsuperscript{9} and the other compared the \textit{Journal of the American Medical Association}, which has adopted to CONSORT standards, with the \textit{New England Journal of Medicine}, which has not.\textsuperscript{10}

In 2007, Yao et al\textsuperscript{2} reported on trends in RCTs published in 2000-2005 in otolaryngology, noting that they made up a small percentage of the overall body of literature and that they were likely to improve in quality of reporting as otolaryngology journals embraced CONSORT standards. We aim to further delineate developments in quantity and reporting quality of RCTs in the past decade by examining articles published in 2011-2013 in the 4 leading peer-reviewed journals in otolaryngology.

\textbf{Methods}

Institutional approval was not required for this study. No patient information from our institution was used. Only published, publicly available journal articles were used for this study. Data collection was performed similarly as previously reported by Yao et al\textsuperscript{2} so as to maximize comparability of our results to that previous evaluation. The study sample consisted of all original RCTs published from 2011 to 2013 in the 4 most widely circulated peer-reviewed otolaryngology journals: \textit{The Annals of Otolaryngology, Rhinology, and Laryngology}; \textit{JAMA Otolaryngology—Head and Neck Surgery} (formerly \textit{Archives of Otolaryngology—Head and Neck Surgery}); \textit{The Laryngoscope}; and \textit{Otolaryngology—Head and Neck Surgery}. Any study not specifically stating that it was a randomized trial in the abstract or methods section was excluded. Studies were categorized by type of study (clinical, animal, or basic science), the type of intervention (medical, surgical, or both), and by subspecialty (neck or upper aerodigestive tract, otology, sinonasal, oncology, plastic or reconstructive surgery, or miscellaneous).

Conclusion scores regarding preference between the experimental and control groups were assigned using a 6-point scale introduced by Als-Nielsen et al.\textsuperscript{11} This was the same scale used by Yao et al.,\textsuperscript{2} which allows for direct comparison between literature from the past decade to this decade. The following grading scores were used: 1 (control intervention is preferred and should be considered the standard of care), 2 (control intervention preferred to experimental intervention, but experimental intervention might be promising under some circumstances), 3 (experimental and control interventions about equal, but the control cheaper, easier to administer), 4 (experimental and control interventions about equal, but the experimental cheaper, easier to administer), 5 (experimental intervention preferred to control, but further trials still indicated; experimental possibly more costly), and 6 (experimental intervention is preferred and should be considered the standard of care). Jadad scores were assigned as a measure of the reporting quality, with points allocated for including randomization, blinding, and accounting for withdrawals or dropouts.\textsuperscript{12} Two points were given for randomization, another 2 points for blinding, and 1 point for withdrawals. Additional points were added or subtracted from the score based on methods of randomization and blinding. Scores ranged from zero (lowest quality) to 5 (highest quality).

Method of statistical analysis was assessed for use of intention-to-treat (ITT) analysis, reporting of $P$ values, and reporting of any confidence intervals. Reporting of blinding, blinding method, randomization, and randomization method was also assessed separately. We also assessed whether there was mention of adverse events and withdrawals and, if so, whether any adverse events occurred and whether there were differences between the compared arms. Funding source was examined for industry and nonindustry support as well as type of support received (grant/money, supplies, statistician, industry author, or not specified). Undisclosed relationships between authors and industry were investigated when a study was funded by a for-profit industry by performing an Internet search (Google search engine) with the name of the lead author and the industry.

Data were collected into a Microsoft Excel spreadsheet (Microsoft Corp, Redmond, Washington). Articles were reviewed by 2 authors (L.A.L. and E.F.-T.) and discrepancies resolved through consensus. Categorical data were analyzed using the Fisher exact test and nonnormal data using a Mann-Whitney $U$-test to compare our results with data from RCTs published in the 2000-2005 time period provided by Yao et al.\textsuperscript{2} All statistical analysis was conducted using SPSS 20.0 (SPSS, Inc, an IBM Company, Chicago, Illinois).

\textbf{Results}

We identified 5729 research articles published between 2011 and 2013 in these 4 journals, of which the final sample included 189 (3.3\%) RCTs. Twenty-four (13\%) were published in \textit{The Annals of Otolaryngology, Rhinology, and Laryngology}; 26 (14\%) were published in \textit{JAMA Otolaryngology—Head and Neck Surgery} (formerly \textit{Archives of Otolaryngology—Head and Neck Surgery}); 74 (39.2\%) were published in \textit{The Laryngoscope}; and 65 (34\%) were published in \textit{Otolaryngology—Head and Neck Surgery}.

Clinical research was the predominant theme in these RCTs ($n = 162$ [86\%]), while a minority of trials were on animals ($n = 27$ [14\%]). The distribution of included RCTs by subject category was 43 (23\%) neck or upper aerodigestive tract, 49 (26\%) otology, 59 (31\%) sinonasal, 10 (5.3\%) oncology, 7 (3.7\%) plastic or reconstructive surgery, and 21 (11\%) miscellaneous, a distribution that was similar to what had been described for trials published in 2000-2005 (Figure 1). The experimental intervention was medical for 87 (46\%) trials, surgical for 71 (38\%) trials, and mixed for 31 (16\%).

The reporting quality of included RCTs was fair, with a median Jadad score of 3 (out of a maximum of 5 points); however, one third of the trials achieved a score of 4 or 5 (33\%). In addition, most investigators (57\%) concluded that
the experimental condition was preferred or highly preferred over the control intervention (conclusion scores of 5 or 6). The randomization score was not specified for 86 trials (47%), specified and valid for 82 trials (43%), and specified and invalid for 21 (11%). Only 62 of the studies included conducted a double-blind protocol, of which 61 (32%) were valid and 1 was invalid (0.5%). One hundred eleven trials (59%) were conducted without double-blinding, and 17 trials (9%) did not specify blinding protocol. Only 113 trials (60%) mentioned withdrawals or dropouts.

The median sample size was 50, with slightly fewer subjects included in the final clinical analyses (median, 48). Most studies (n = 179 [95%]) reported results as P values, and 10% reported with confidence intervals. Thirty-two percent of studies (n = 60) mentioned an ITT method of analysis, although several trials used multiple methods for data analysis. Adverse events were mentioned by 123 (65%) of authors, and of those who reported an adverse event occurring (n = 52), 31 (59%) reported a similar incidence for experimental and control groups. Of the remaining trials that mentioned adverse events occurring, 71 (58%) reported none occurring. Conversely, 66 (35%) authors made no mention of adverse events.

Research funding support was not reported for 103 trials (55%). Among the other trials, funding was provided by industry (for-profit) in 19 (10%), by not-for-profit in 47 (25%), and by both in 12 trials (6%). However, 8 trials (4%) made no mention of funding sources or disclosure statement. Of the 31 trials that included industry support, an Internet search yielded no undisclosed relationships. Not-for-profit support was monetary (84%), supplies (14%), or statistical (2%). In contrast, industry support was monetary (61%), supplies (45%), statistical (3%), or authorship (3%).

When trials with any industry support or funding were compared with all others, there was no difference in the incidence of conclusions favoring the experimental group (73% for unfunded vs 65% for funded; P = .166; Table 1). Similarly, studies that were unfunded did not have significantly different Jadad scores (median, 3 vs 2; P = .07) and sample sizes (median, 54 vs 40; P = .15) as funded studies.

**Table 2** summarizes the features and reporting quality of RCTs in our sample and compares them with the data from RCTs published in 2000-2005 as published by Yao et al.2 As shown, there is some evidence for statistically significant improvement in the proportion of trials with a Jadad score above 3 and those reporting the mode of randomization and double-blinding, but there is also a statistically significant decrease in the proportion of trials that mention withdrawals and use of ITT.

**Discussion**

For the 3-year period we examined, we identified 5279 otolaryngology articles published in 4 journals. This is in contrast to 5467 articles identified by Yao et al2 in the same 4 journals during a 6-year period and represents an approximate doubling in volume of articles published. In this time period, the proportion of RCTs has remained roughly equal, with 3.7% previously and 3.3% currently, although the absolute number per year has increased, with 202 reported for the 2000-2005 time period and 189 identified in our study, which was for the 2011-2013 time period.2 The increase in quantity of articles overall, and the proportional increase in RCTs, is consistent with the overall trend in medical literature and reports previously within the field of otolaryngology.1-3

While quantity is clearly increasing, there is room for improvement in the reported quality of RCTs. Most of the measures of reporting quality we examined were similar to the previous study period examined by Yao et al.2 Despite CONSORT recommendations, in the 7 years between the Yao et al study and ours, a number of elements of RCT design and reporting have apparently failed to improve.

The reporting of withdrawals and of ITT analyses has decreased in the recent sample. Fewer participants may be withdrawing from trials, investigators may be failing to report withdrawals, or there may be the possibility that the design of some RCTs is planned to limit follow-up and potential for withdrawals. For example, 1 study examined the effects of different concentrations of remifentanil on cough during laryngeal microsurgery, effectively limiting the length of data collection to the beginning and end of surgery.13 Lack of ITT is worrisome as it can introduce substantial bias in a randomized trial.

Conversely, the reporting of the method of randomization and use of double-blinding have increased. This also explains the 1-point improvement in the median Jadad score compared with the assessment of RCTs in 2000-2005. Most RCTs reported the method of randomization, and among those, the majority used valid means to randomize study participants. A previous report examined RCTs from 1966 to 1995 and compared those before and after the CONSORT statement, reaching the conclusion that RCTs had not improved in 30 years.7 Currently, with regard to 2 important aspects of study design, we see some improvement. However, this is countered by worsening in 2 other features.

Of the RCTs of treatment efficacy included, the vast majority, 86%, were clinical studies. Oncology and plastic
surgery were less well represented, consisting of 8% of the articles combined. This is consistent with previous reports.2 Randomized controlled trials in oncology may be seen less often because it is not ethical to provide a patient with cancer with a possibly ineffective treatment, and patients undergoing plastic surgery procedures may be less willing to be involved in RCTs for cosmetic reasons. Randomized controlled trials in oncology may also be published in higher impact oncology specialty journals.

There was not any significant relationship between conclusion scores in favor of the intervention and type of support. In contrast to previous reports,2 there was no correlation between industry support and sample size or median Jadad scores. This may be due to lack of power, since there are few trials sponsored by the industry. There was less overall support from industry compared with previous evaluations, 10% of trials currently and 21% previously. In recent years, over concern for impact on patient care, the medical community has worked to limit the influence of industry by limiting payment of gifts, dinners, travel, lodging, or other personal expenses to physicians.14 In contrast to our current findings regarding industry involvement in RCTs, Sun et al15 recently found that 22% of RCTs for head and neck cancer were industry supported. It is possible that industry-supported trials are being reported in otolaryngology journals other than the ones included in our study or that they are being published in medical journals with a more general interest. An additional possibility is that otolaryngology-related industries focus on medical devices rather than drugs. Some of these devices have less stringent Food and Drug Administration regulatory requirements, and thus probably fewer RCTs are performed compared with drugs. However, the most likely explanation for the low rate of industry involvement is that most interventions in otolaryngology are not such that the industry may show interest in them.

### Table 1. Investigator Conclusions for 184 Randomized Controlled Trials Delineated by Type of External Support or Funding.

<table>
<thead>
<tr>
<th>Investigator Conclusion Score</th>
<th>None</th>
<th>Industry (For-Profit) Only</th>
<th>Not-for-Profit Only</th>
<th>Industry and Not-for-Profit</th>
<th>Support Not Mentioned</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18 (60)</td>
<td>4 (13.3)</td>
<td>4 (13)</td>
<td>3 (10)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>2</td>
<td>45 (58)</td>
<td>5 (6.5)</td>
<td>20 (26)</td>
<td>5 (7)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>3</td>
<td>12 (50)</td>
<td>3 (12.5)</td>
<td>6 (25)</td>
<td>1 (4)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>4</td>
<td>12 (52)</td>
<td>0</td>
<td>7 (30)</td>
<td>2 (9)</td>
<td>2 (9)</td>
</tr>
<tr>
<td>5</td>
<td>3 (33)</td>
<td>2 (22)</td>
<td>3 (33)</td>
<td>0</td>
<td>1 (11)</td>
</tr>
<tr>
<td>6</td>
<td>11 (52)</td>
<td>4 (19)</td>
<td>5 (24)</td>
<td>1 (5)</td>
<td>0</td>
</tr>
</tbody>
</table>

aValues are presented as number (%).

### Table 2. Comparison of Reporting in Otolaryngology Randomized Controlled Trials between 2000-2005 and 2011-2013.

<table>
<thead>
<tr>
<th></th>
<th>2011-2013</th>
<th>2000-2005</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size, median</td>
<td>48</td>
<td>53</td>
<td>.10</td>
</tr>
<tr>
<td>Jadad score, median</td>
<td>3</td>
<td>2</td>
<td>.0005</td>
</tr>
<tr>
<td>Achieving 4 or 5</td>
<td>62 (33)</td>
<td>32 (16)</td>
<td>.0001</td>
</tr>
<tr>
<td>Randomization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not specified</td>
<td>86 (46)</td>
<td>149 (74)</td>
<td>.0001</td>
</tr>
<tr>
<td>Randomized and valid</td>
<td>82 (43)</td>
<td>45 (22)</td>
<td>.002</td>
</tr>
<tr>
<td>Randomized and invalid</td>
<td>21 (11)</td>
<td>8 (4)</td>
<td>.10</td>
</tr>
<tr>
<td>Blinding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not double-blind</td>
<td>111 (59)</td>
<td>142 (70)</td>
<td>.14</td>
</tr>
<tr>
<td>Double-blind and valid</td>
<td>61 (32)</td>
<td>35 (17)</td>
<td>.02</td>
</tr>
<tr>
<td>Double-blind and invalid</td>
<td>1 (1)</td>
<td>2 (1)</td>
<td>1.00</td>
</tr>
<tr>
<td>Not specified</td>
<td>17 (9)</td>
<td>24 (12)</td>
<td>.65</td>
</tr>
<tr>
<td>Withdrawals mentioned</td>
<td>112 (60)</td>
<td>194 (96)</td>
<td>.0001</td>
</tr>
<tr>
<td>P value</td>
<td>179 (95)</td>
<td>178 (88)</td>
<td>.13</td>
</tr>
<tr>
<td>Confidence interval</td>
<td>19 (10)</td>
<td>22 (11)</td>
<td>1.00</td>
</tr>
<tr>
<td>ITT</td>
<td>60 (32)</td>
<td>117 (58)</td>
<td>.0004</td>
</tr>
<tr>
<td>Adverse events mentioned</td>
<td>123 (65)</td>
<td>139 (69)</td>
<td>.65</td>
</tr>
</tbody>
</table>

Abbreviation: ITT, intention to treat.

Values are presented as number (%) unless otherwise indicated.
Many studies from other surgical specialties collected data that allow comparisons to our study (Figure 2). Important characteristics of study design reporting examined in each included randomization method, blinding, adverse events, and use of ITT analysis. Of these characteristics, adverse events appear to be most commonly reported among most specialties and the use of ITT analysis most variable between specialties. Previous research has noted that adverse events are often improperly reported and information is distorted. In these analyses, reporting of adverse events was assessed just in terms of overall inclusion and not evaluated for proper reporting. For example, a large number of trials in our sample (71/189) reported that no adverse events occurred. This is spurious as it is hard to believe that so many interventions had no adverse events at all. Overall, no specialty appears to have consistently superior RCT reporting standards compared with the others.

Research reporting quality in other fields has also been evaluated in additional ways. Reports on neurosurgery and orthopedic surgery research have used EBM levels of evidence (LOE) to assess quality, determining an increase in quality to be a higher number and proportion of higher LOE articles being published. The neurosurgery article, published in 2003, found LOE to be insufficient. The orthopedic surgery article examined data a decade later and found articles of higher LOE to be increasing, representing an overall improvement in research quality. Similarly conducted studies have been performed for otolaryngology research, with consistent results.

For this analysis, we used previously published measures for authors’ conclusions and reported quality scores, allowing for ease of reproducibility. We also used the same collection form previously used to evaluate the quality of RCTs from 2000-2005, allowing us to directly compare reported quality items between that time period and now. The body of otolaryngology literature is constantly evolving, and it is important to know what improvements we have made, if any.

There are also several limitations to this study. We evaluated RCTs only, and this should not be extrapolated to the larger body of otolaryngology research overall, which includes mostly other types of designs. We chose the 4 most widely circulated peer-reviewed journals in otolaryngology to examine RCTs, but they also may not be reflective of the entire breadth of otolaryngology research. Several relevant trials may be published in general surgical or even general medical journals. In addition, our ability to evaluate the literature is limited by what is reported. With a number of RCTs failing to report certain aspects of study design, our conclusions about the quality of study design are limited, although it does provide us with the information of what areas might need more attention.

**Conclusion**

The body of otolaryngology research published in the main specialty journals has doubled in the past decade, but the proportion of RCTs remains very low, and there is no strong evidence for consistent improvements in reported quality of RCTs, perhaps with the exception of reporting of the method of randomization and double-blinding. There are still considerable deficiencies. A number of quality measures have not seen substantial improvements in their reporting in the past decade. However, otolaryngology research is relatively on par with research in other surgical fields. Improved design and reporting are essential for improving the credibility of the evidence on interventions in otolaryngology.

**Author Contributions**

Sarfaraz M. Banglawala, conception, design, data analysis, revision of manuscript, final approval, agreement to be accountable for all aspects of the work; Lauren A. Lawrence, data collection, drafting and revision of manuscript, final approval, agreement to be accountable for all aspects of the work; Emily Franko-Tobin, data collection, data analysis, drafting of manuscript, final approval, agreement to be accountable for all aspects of the work; Zachary M. Soler, revision of manuscript, interpretation of data, final approval, agreement to be accountable for all aspects of the work; Rodney J. Schlosser, revision of manuscript, interpretation of data, final approval, agreement to be accountable for all aspects of the work; John Ioannidis, interpretation of data, drafting and revision of manuscript, final approval, agreement to be accountable for all aspects of the work.

**Disclosures**

**Competing interests:** Zachary M. Soler, consultant for Brainlab; Rodney J. Schlosser, consultant for Brainlab, Olympus, and Arrinex, grant support from Optinose and Intersect ENT; John Ioannidis, The Meta-Research Innovation Center at Stanford (METRICS), is funded by a grant by the Laura and John Arnold Foundation.

**Sponsorships:** None.

**Funding source:** None.

**References**


