Cervical Fusion

To the Editor: We appreciated the article by Fraser and Härtl (Fraser JP, Härtl R: Anterior approaches to fusion of the cervical spine: a metaanalysis of fusion rates. J Neurosurg: Spine 6:298–303, April, 2007). It is an important topic, and the authors took on a challenging task. However, the statistical methods in the article are incorrect, and we are concerned that the results and conclusions may not be accurate. The purpose of this letter is to emphasize the nature of a true meta-analysis and to point out the pitfalls of pooling data across studies.

Abstract

Object. Anterior cervical discectomy (ACD), ACD with interbody fusion (ACDF), ACDF with placement of an anterior plate system (ACDFP), corpectomy, and corpectomy with plate placement are used to fuse the cervical spine. The authors conducted a metaanalysis of studies published after 1990 in which fusion rates achieved with each procedure were reported for patients with degenerative disease at one, two, and three disc levels.

Methods. Twenty-one papers each included data on at least 25 patients. In each of the 21 studies the average clinical follow up was more than 12 months, and the results were evaluated according to radiographic evidence of fusion and delineated by the number of levels fused. Chi-square and Fisher exact tests were used for comparisons. The mean age of the patients was 46.7 years, 46.6% were female, and the mean follow-up period was 39.6 months. The studies included 2682 patients and the overall fusion rate was 89.5%. For single disc–level disease, fusion rates were 84.9% for ACD, 92.1% for ACDF, and 97.1% for ACDFP (p = 0.0002). For two disc–level disease, fusion rates were 79.9% for ACDF, 94.6% for ACDFP, 95.9% for corpectomy, and 92.9% for corpectomy with plate placement (p = 0.0001). For three disc–level disease, fusion rates were 65.0% for ACDF, 82.5% for ACDFP, 89.8% for corpectomy, and 96.2% for corpectomy with plate placement (p = 0.0001). The use of anterior plates significantly improved fusion for one-level (p < 0.0001), two-level (p < 0.0001), and three-level (p < 0.05) ACDF. There was no significant difference in fusion rates between two-level ACDF and corpectomy with plate placement.

Conclusions. The anticipated fusion rate is one of several factors that may guide surgical decision making. Anterior cervical decompression and fusion results in high fusion rates. The results of the authors’ study show that regardless of the number of levels fused, the use of an anterior cervical plate system significantly increases the fusion rate. For two-disc–level disease, there was no significant difference between ACD with a plate system or corpectomy with a plate system. For three-disc–level disease, however, the evidence suggests that corpectomy with plate placement is associated with higher fusion rates than discectomy with plate placement.

The authors provide interesting and useful information regarding fusion rates following anterior cervical surgery. Their study constitutes a systematic review of 21 studies but is not a proper meta-analysis as such. By merely pooling data from the various studies into contingency tables and analyzing them with chi-square and Fisher’s exact test procedures, they ignore problems frequently encountered in meta-analysis, and, thus, the paper may result in incorrect numerical estimates of association with potentially wrong conclusions. Some frequent problems that need to be addressed in a meta-analysis (and that were not addressed in this article by Fraser and Härtl) include the following: 1) differences among studies in the prevalence of a risk factor (for example, surgical method) and in the proportion of patients with a specified outcome (for example, fusion); 2) differences among studies in the observed effect of the risk factor on the outcome (for example, differences among studies in fusion rates for each treatment: ACDF vs ACDFP vs corpectomy vs corpectomy with plate placement); and 3) publication bias.

Table 1 shows no association between the treatment and the outcome in either Study 1 or Study 2 (odds ratio OR = 1.0 in both cases). However, pooling of these tables into a single contingency table produces a misleading result (Table 2). Table 2 shows an OR for success, given the treatment, calculated as OR = (4/2)/(6/4) = 1.33, supplying erroneous “evidence” that the treatment improves the outcome. This nonexistent association is only induced by the improper pooling of data across the 2 studies. If a proper meta-analysis of these 2 studies were carried out, it would yield a synthesized combined effect of OR = 1.0, correctly indicating no association between the treatment and the outcome.

The second issue that needs to be addressed in any meta-analysis (and which was not addressed in the study being considered here) is the choice between a fixed effects and a random effects meta-analysis. The meta-analysis proceeds along different lines according to whether the association between a risk factor and the outcome varies across studies more than can be readily explained by chance. Random versus fixed effects meta-analysis models weight studies differently.

The last issue noted earlier and to be addressed by a meta-analysis is publication bias. It may happen that small studies, studies with no effect, or studies with an effect in the opposite direction than expected are not published, possibly resulting in a biased estimate of association.

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment Success</th>
<th>Treatment Failure</th>
<th>Control Success</th>
<th>Control Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Study 2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Nowadays it is standard in the scientific literature to utilize appropriate meta-analytical methods, such as fixed or random effects meta-analysis, tests for heterogeneity, and tests and adjustments for publication bias. These methods are superior to the approach used in this paper because they optimally weight the effect sizes across studies and guard against various forms of bias, some of which we have mentioned here. In particular, the statistical approach used in this paper completely ignores the heterogeneity that almost always exists across a group of observational (nonrandomized) studies or even randomized studies.

Even a meta-analysis of high quality randomized controlled trials should be subjected to both fixed effect and random effect models to test for the degree of heterogeneity of the studies.\(^1\) It is unclear whether the authors included studies with 25 or more patients, as stated in their abstract, or studies with more than 25 patients, as stated in their Table 1.

We were puzzled by one more issue in the paper. The exclusion of smaller studies is not justified in the text and may introduce a bias in the current analysis.\(^2\)\(^,\)\(^4\) It is unclear whether the authors included studies with 25 or more patients, as stated in their abstract, or studies with more than 25 patients, as stated in their Table 1.

We will be interested in the response by authors Fraser and Härtl. The basis for using these data to guide clinical decisions would be strengthened immeasurably if the authors’ present conclusions were confirmed with a proper meta-analysis.

W. BRADFORD DELONG, M.D.
University of California
San Francisco, California
NAYAK POLISSAR, PH.D.
BLAZERI NERADILEK, M.S.
The Mountain-Whisper-Light Statistical Consulting
Seattle, Washington

References


RESPONSE: We thank Dr. DeLong and his colleagues for their rigorous examination of our analysis on cervical spine fusion rates from anterior procedures. The statistical methods that they describe are important tools for analyzing published data. We recognize their arguments about presentation of statistical data. In retrospect, we should have presented the data differently. The goal of our study was to provide a set of benchmarks for spine surgeons regarding fusion rates. For the practicing spine surgeon, what should she or he quote in terms of fusion rate for any given procedure? Our goal was not to determine the superiority of discectomy and fusion or corpectomy; as stated in Discussion, we aimed to provide practical benchmarks for expected fusion rates. Often in the neurosurgical literature, and especially in issues of spine surgery, there is no Class I evidence to answer questions such as the ones posed in this study. However, many groups have published their data on fusion rates for a variety of cervical procedures. We aimed to summarize and synthesize these data. Our response to Dr. DeLong’s letter is 2-fold. First, we recognize the limitations of our statistical methods and submit a statistical summary of the data that is more sound but reaches similar results. Second, we make note of the limitations that we discussed in our paper.

The methods of meta-analysis discussed by Dr. DeLong and his colleagues are more statistically rigorous than those that we used. We, hereafter, explain our decision to use certain criteria, and provide a new analysis of the data using their requested methods. First, our inclusion/exclusion criteria were strictly designed prior to screening the studies. We elected to exclude studies of 25 patients or fewer (the Abstract was not worded properly). One could argue that doing so introduces another form of bias, but we aimed to evaluate large series, and to exclude case reports and small groups of publications aimed at demonstrating feasibility of new techniques. Second, only four studies were randomized trials; the majority represented large series of patients. In addition, in most cases, the analyzed groups of ACD, ACDF, ACDFP, corpectomy, and corpectomy with plate placement were not compared within each study. For example, Martin et al.\(^1\) published a series of 317 consecutive patients treated only with ACDF. If we only included studies that directly compared ACDF versus ACDFP, this study would be excluded. From a statistical point of view, in retrospect, we could have provided an analysis of publication bias for each fusion rate obtained for each level and procedure. This would definitely have made our analysis stronger, and we recognize this as a limitation of our study. We therefore submit Table 1, which summarizes a meta-analysis of the data. We present both fixed effects and random effects models for analysis, and we conducted an analysis of bias using both the Begg and Mazumdar Rank Correlation and Egger tests. Statistically, the Begg and Mazumdar Rank Correlation test is powerful with larger numbers of studies. An alternative, the Egger test, may be more sensitive in detecting bias, but may elicit a false-positive for

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data from the two studies pooled for analysis. The OR shows an erroneous positive association between treatment and outcome</td>
</tr>
<tr>
<td>Pooled Data From Studies 1 &amp; 2</td>
</tr>
<tr>
<td>Success</td>
</tr>
<tr>
<td>treatment</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

J. Neurosurg.: Spine / Volume 8 / May 2008 495