Closing in on the Truth About Recombinant Human Bone Morphogenetic Protein-2: Evidence Synthesis, Data Sharing, Peer Review, and Reproducible Research

Readers may think they are seeing double. They are not. This issue includes 2 systematic reviews (1, 2) that use the same data to address the same question: Compared with iliac crest bone grafting, does recombinant human bone morphogenetic protein-2 (rhBMP-2) safely improve outcomes of spinal fusion surgery? The Yale University Open Data Access Project commissioned the reviews, believing that confidence in the findings would be greatest if both reached the same conclusions. Both include patient-level meta-analyses of data from randomized, controlled trials (RCTs) obtained from Medtronic (Minneapolis, Minnesota), the manufacturer of rhBMP-2 (3–5). Annals’ review involved 2 internal teams of physician and statistical editors and distinct sets of external reviewers. Each team handled 1 manuscript, and neither had access to the other manuscript or associated reviews until both were accepted for publication. The clinical question is of great interest to orthopedic surgeons and relevant to internists who encounter patients who have had or are considering having spinal fusion. Yet, beyond the clinical issues, these dual reviews spotlight the power of evidence synthesis, data sharing, peer review, and reproducible research.

The Sum Really Is More Than Its Parts

Although the RCT may be the “king” of study designs, 1 RCT is rarely definitive. The first published rhBMP-2 trial (n = 14) reported that disability scores improved sooner with rhBMP-2 than with bone graft, quality of life improved, and no adverse events occurred (6). Although few would consider the results of this small, initial trial as a solid basis for treatment decisions, some may rely on the largest, later trial. That trial (n = 577) reported that rhBMP-2 improved outcomes, including function, pain, and return to work (7). Yet, after systematic evaluation and synthesis of all available evidence, both systematic reviews published here independently conclude that rhBMP-2, compared with iliac crest bone grafting, does not improve pain or function and increases adverse events, possibly including cancer (1, 2). For the outcome of fusion, the focus of initial enthusiasm for rhBMP-2, 1 review concluded that it improved by a marginally statistically significant amount with rhBMP-2 (overall fusion, 1.14 [95% CI, 1.03 to 1.25]), and the other concluded that it did not (anterior fusion, 1.05 [CI, 0.88 to 1.24]; posterior fusion, 1.16 [CI, 0.96 to 1.41]). Of importance, the CIs for both sets of estimates substantially overlap, and neither review found differences in pain or function, the outcomes that reflect patients’ well-being.

This shows us yet again that conclusions not based on a careful synthesis of all available evidence can be misleading. For that reason, Annals considers systematic reviews to be the best evidence for clinical practice.

Share and Share Alike (Or a Problem Shared Is a Problem Halved)

The reviews in this issue are not the first to synthesize evidence about rhBMP-2 in spine surgery. A 2011 review by Carragee and colleagues (8) suggested that adverse events associated with rhBMP-2 use ranged from 10% to 50% depending on the surgical approach. Because these authors did not have access to all patient-level RCT data, they had limited ability to precisely determine the benefits and harms of rhBMP-2.

Although access to individual-patient data cannot overcome inherent biases in individual studies derived from design or implementation flaws, our experience illustrates that routine sharing of individual-patient RCT data, perhaps planned for prospectively with other researchers as exemplified by the Early Breast Cancer Trialists’ Collaborative Group (9) or efforts to promote sharing of National Institutes of Health–funded research (http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm), could lead to more rapid, precise, generalizable, and trustworthy knowledge about medical interventions than the “closed data” world that currently defines medical research. To learn how to access the data used in the current reviews, visit http://medicine.yale.edu/core/projects/yodap/index.aspx.

What You See Is Not What We Got

These reviews provide an opportunity to reflect on peer review. As new methods of rapidly disseminating research proliferate, some question the benefits of traditional peer review and editing (10). The published reviews are accompanied by supplemental material that includes the initial submitted manuscript and correspondence between editors and authors. The changes from submission to acceptance included greater focus on individual-patient data, more sophisticated handling of missing and sparse data, more detailed methodological appendices (flow diagram, search strategy, details on analyses, and SAS codes), and more thorough assessment of study quality.

The perfect peer-review process does not exist, but this peek into the black box of our procedures should show readers that what they see in the journal can be very different (and, we contend, better) than what the authors initially submit. We see helping authors improve their papers to make their methods more transparent and better...
align the evidence and the conclusions as one of our core missions. Improving reports of medical science facilitates effective postpublication review by the broader medical community, arguably the most important part of the review process, which ultimately improves the quality of the information available for patient care. The substantive changes we noted between submission and publication of these 2 reviews are typical for most articles we and other major journals publish; consistent with the positive changes seen when we formally evaluated the Annals peer review and editing process nearly 2 decades ago (11); and relevant to concerns about alternatives currently being proposed to prepublication, journal-based peer review, and editing. Although peer review continues after publication, the publication of cursorily vetted research can materially impede the attainment of truth.

**The Value of Reproducible Research**

Was there added value in soliciting 2 separate systematic reviews on this topic? As advocates of “reproducible research” (12), we think so. Despite the frequent use of rhBMP-2, there remained substantial controversy about its benefits and harms. The fact that 2 independent groups armed with the same question and the same patient-level data, along with the guidance of 2 independent sets of reviewers and editors, arrived at essentially the same conclusions should greatly temper enthusiasm for the intervention. Beyond replicating results, this exercise demonstrates additional value in having different scientists tackle the same data. The Table summarizes key characteristics of each review. There were modest differences in approach, and in fact, they came to somewhat different conclusions about the effect of the 2 interventions on bone fusion. This occurred because of differences in evidence selection and different decisions about analyzing posterior and anterior lumbar interbody fusion separately or together. There was another difference worth noting. Fu and colleagues (2) included a second aim of assessing bias in the original trial reports, and the editors handling their manuscript did not discourage this effort. Simmonds and colleagues (1) fo-

### Table. Key Features of Systematic Reviews and Meta-analyses of rhBMP-2

<table>
<thead>
<tr>
<th>Feature</th>
<th>Fu and Colleagues (2)</th>
<th>Simmonds and Colleagues (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stated objectives</td>
<td>Evaluate effectiveness and safety of rhBMP-2</td>
<td>Evaluate effectiveness and safety of rhBMP-2</td>
</tr>
<tr>
<td></td>
<td>Assess reporting bias in original trial reports</td>
<td></td>
</tr>
<tr>
<td>Data sources</td>
<td>17 Medtronic trials</td>
<td>17 Medtronic trials</td>
</tr>
<tr>
<td></td>
<td>MEDLINE and other database searches from 1996 to 2012</td>
<td>MEDLINE and other database searches through 2012</td>
</tr>
<tr>
<td>Evidence identified</td>
<td>12 Medtronic RCTs (n = 1879)</td>
<td>11 Medtronic RCTs (n = 1302)</td>
</tr>
<tr>
<td></td>
<td>1 other RCT (n = 102*)</td>
<td>1 other RCT (n = 106*)</td>
</tr>
<tr>
<td></td>
<td>31 cohort studies</td>
<td>35 observational studies for safety outcomes</td>
</tr>
<tr>
<td></td>
<td>47 intervention series</td>
<td>MedWatch reports</td>
</tr>
<tr>
<td></td>
<td>34 case series or reports</td>
<td></td>
</tr>
<tr>
<td>Trials identified but excluded</td>
<td>1 Medtronic RCT (n = 3), unclear which other of the 17 Medtronic trials excluded and why</td>
<td>4 single-group trials</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 trial that did not use iliac crest bone graft as the comparator</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Primary: “Overall success” as defined by individual trials, which was mostly fusion</td>
<td>Primary: Defined as outcomes that authors considered likely to be important to patients, which was pain (rated by the Oswestry Disability Index score)</td>
</tr>
<tr>
<td></td>
<td>Secondary: Pain disability, neurologic status, function, return to work; adverse events</td>
<td>Secondary: Fusion, adverse events</td>
</tr>
<tr>
<td></td>
<td>Outcomes examined at 6 wk; 3, 6, 12, and 24 mo</td>
<td>Outcomes examined at 6 wk; 3, 6, 12, and 24 mo</td>
</tr>
<tr>
<td>Analytic approach</td>
<td>Mixed-effects models stratified by spinal area and surgical approach were used for primary analyses</td>
<td>2-stage meta-analysis using random-effects model (primary analysis)</td>
</tr>
<tr>
<td></td>
<td>Small numbers of events dictated pooling across all spinal areas and surgical approaches for cancer and death end points</td>
<td>Mixed-effects models (sensitivity analysis)</td>
</tr>
<tr>
<td></td>
<td>2-stage meta-analysis was used for sensitivity analyses or when problems were encountered in fitting mixed-effect models.</td>
<td></td>
</tr>
<tr>
<td>Effectiveness conclusion</td>
<td>rhBMP-2 and bone graft had similar effects on overall success, fusion, and other effectiveness outcomes.</td>
<td>At 2 y after surgery, Oswestry Disability Index scores with rhBMP-2 were 3.5% lower than with bone graft (95% CI, 0.5%–6.5%), an amount judged to be clinically insignificant. Radiographic fusion was 12% (CI, 2%–23%) more common with rhBMP-2.</td>
</tr>
<tr>
<td>Safety conclusion</td>
<td>The risks of any adverse event were high (77%–93% at 2 y) and similar for both groups.</td>
<td>At or shortly after surgery, several adverse events were more common among patients receiving rhBMP-2. Data on adverse events suggest that heterotopic bone formation, dysphagia, osteolysis, and retrograde ejaculation may be more common with rhBMP-2. Cancer was nearly twice as common among patients receiving rhBMP-2 (RR, 1.98 [95% CI, 0.86–4.54).</td>
</tr>
<tr>
<td></td>
<td>For anterior lumbar interbody fusion, rhBMP-2 was associated with a nonsignificant increase in urogenital problems. rhBMP-2 was associated with wound complications and dysphagia in anterior cervical spinal fusion.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At 24 mo, cancer risk was increased with rhBMP-2 (RR, 3.45 [95% CI, 1.98–6.00]), but event rates were low.</td>
<td></td>
</tr>
<tr>
<td>Reporting bias conclusion</td>
<td>Early journal publications misrepresented the effectiveness and harms through selective reporting, duplicate publication, and underreporting.</td>
<td>The authors report their evaluation of bias in the full evidence report at <a href="http://medicine.yale.edu/core/projects/yodap/index.aspx">http://medicine.yale.edu/core/projects/yodap/index.aspx</a>.</td>
</tr>
</tbody>
</table>

RCT = randomized, controlled trial; rhBMP-2 = recombinant human bone morphogenetic protein-2; RR = relative risk.

* A total of 106 patients initially entered the study, and 102 patients had results from the 2-year follow-up.
cused their efforts on analysis of the data rather than on identifying bias in the initial reports, a focus supported by the editorial team handling their manuscript. Different scientists (and different editors) can approach the same question and even the same data in different ways. This variety of perspectives can only enhance our understanding of the evidence and clarify evidence-based disagreement.

In summary, this Yale University Open Data Access–orchestrated project is a novel exercise that illustrates the value of evidence synthesis, data sharing, peer review, editing, and reproducible research in helping us get closer to the truth. Although it is infeasible to apply this elaborate process to every clinical question, we owe it to our patients to reduce the obstacles that prevent similar efforts from being more common.

Christine Laine, MD, MPH
Editor in Chief

Eliseo Guallar, MD, PhD
Associate Editor, Statistics

Cynthia Mulrow, MD, MSc
Senior Deputy Editor

Darren B. Taichman, MD, PhD
Executive Deputy Editor

John E. Cornell, PhD
Associate Editor, Statistics

Deborah Cotten, MD, MPH
Deputy Editor

Michael E. Griswold, PhD
Associate Editor, Statistics

A. Russell Localio, JD, MPH, MS
Associate Editor, Statistics

Anne R. Meibohm, PhD
Associate Editor, Statistics

Catharine B. Stack, PhD, MS
Deputy Editor, Statistics

Sankey V. Williams, MD
Deputy Editor

Steven N. Goodman, MD, PhD, MHS
Associate Editor, Statistics

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M13-1163.

Requests for Single Reprints: Annals of Internal Medicine, American College of Physicians, 190 N. Independence Mall West, Philadelphia, PA 19106.


References
Correction: Closing in on the Truth About Recombinant Human Bone Morphogenetic Protein-2

In the table of a recent editorial, the cell in the last row under the heading Simmonds and Colleagues should read as follows: The authors report their evaluation of bias in the full evidence report at http://medicine.yale.edu/core/projects/yodap/index.aspx.

This has been corrected in the online version.

Reference