ELSEVIER

The Spine Journal 13 (2013) 1001–1005

THE SPINE JOURNAL

NASS Review

Black, white, or gray: how different (or similar) are YODA and the The Spine Journal reviews of BMP-2?

Christopher M. Bono, MD[a,*], F. Todd Wetzel, MD[b], on behalf of the North American Spine Society Executive Committee, endorsed by the North American Spine Society Section on Biologics

[a]*Department of Orthopaedic Surgery, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St, Boston, MA 02478, USA*
[b]*School of Medicine, Temple University, 3401 N. Broad St, Philadelphia, PA 19140-5103, USA*

Disclaimer: This document was prepared by Executive Committee members of the North American Spine Society (NASS) and reviewed by representatives of the NASS Section on Biologics. It did not undergo peer review by The Spine Journal.

Introduction

Recently, two Yale University Open Data Access (YODA) [1] reviews have been published about the outcomes and adverse events of recombinant bone morphogenetic protein 2 (rhBMP-2) use in the spine. The data that stimulated these independent reviews have spawned some controversy. This controversy was spawned predominantly by the review of Carragee et al. [2], which compared the safety and efficacy data of rhBMP-2 published in industry-sponsored trials with ''subsequently available'' Food and Drug Administration (FDA) data summaries, follow-up publications, and ''administrative and organizational databases.'' In this review, the authors concluded that ''Level I and Level II evidence from original FDA summaries, original published data, and subsequent studies suggest possible study design bias in the original trials, as well as a clear increased risk of complications and adverse events to patients receiving rhBMP-2 in spinal fusion.'' Furthermore, they documented that the ''risk of adverse events associated with rhBMP-2 is 10 to 50 times the original estimates reported in the industry-sponsored peer reviewed publications.''

This ''review of reviews'' is an attempt to delineate the bare facts of the issue and, we hope, to review these important findings in a concise and objective manner. Although we encourage every North American Spine Society (NASS) member to read the source articles in detail, we understand that the amount of data can be overwhelming and difficult to digest. Single sentences from press releases, or even abstracts of the reviews themselves, may be easily tailored for a desired effect. It is the ''Methods'' and, more importantly, the ''Results'' sections that readers must interpret for themselves to draw their own conclusions to construct their own personal synopsis of the information. Of note, both authors of the current review are members of the Executive Committee of the NASS and have fully listed their disclosures on the NASS Web site and at the conclusion of this article. We freely acknowledge that all authors have bias and by the mechanism of disclosure enjoin the reader to make their own determination of author credibility.

The reviews

Four publications are reviewed subsequently. First is the 2011 article by Carragee et al. [2], ''A critical review of recombinant human bone morphogenetic protein-2 trials in spinal surgery: emerging safety concerns and lessons learned.'' Second is the YODA article, Simmonds et al. [3], ''Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion.'' This review was also published in the *British Medical Journal* [4]. This

FDA device/drug status: Certain rhBMP-2 uses are approved; others are not.

Author disclosures: ***CMB:*** Consulting: Harvard Clinical Research Institute (C, reimbursed for time as part of the Trial Design Team, developing, and implementing protocols for spine research); United Health Care (B, member of Advisory Board); Board of Directors: North American Spine Society (Nonfinancial, Treasurer); Other Office: Intrinsic Therapeutics (B, data safety monitoring board, no remuneration yet, for prospective study of new device); JAAOS (B, Deputy Editor), *The Spine Journal* (Nonfinancial, Deputy Editor). ***FTW:*** Board of Directors: McKenzie Institute International (B, annual stipend and nonfinancial travel expenses); North American Spine Society (Nonfinancial, travel expenses for SIG, BOD meetings).

The disclosure key can be found on the Table of Contents and at www.TheSpineJournalOnline.com.

* Corresponding author. Department of Orthopaedic Surgery, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St, Boston, MA 02478, USA. Tel.: (617) 732-7238.

E-mail address: bonocm@me.com (C.M. Bono).

1529-9430/$ - see front matter
http://dx.doi.org/10.1016/j.spinee.2013.07.030

version of the analysis was substantially shorter and focused more on the difference between the individual-participant data (IPD) and published data (eg, differences in rates of reported adverse events). Last is the article by Fu et al. [5], ''Effectiveness and harms of recombinant human bone morphogenetic protein-2 in spine fusion.''

Methodology

The data synthesis methodologies of the YODA reviews [3,5] were similar and comparable. The methods used in both reviews were ''prespecified'' and registered in PROSPERO (An International Prospective Register of Systematic Reviews, National Institute for Health Research, The University of York, UK) in February 2012. Most importantly, the YODA reviews had access to the IPD from Medtronic (Memphis, TN, USA). These data had been collected and stored from the past company-sponsored studies of rhBMP-2 and could be tracked to subsequent publications.

In addition to the IPD, Fu et al. [5] also used internal reports from Medtronic. From the published methodology, Simmonds et al. [3] did not use these data. Furthermore, Fu et al. [5] accessed data from the FDA Web site, which stored the data submitted by Medtronic for subsequent approval and labeling. Both YODA groups additionally performed a comprehensive systematic review of published literature using recognized sources (eg, Medline, Embase, Cochrane Library).

According to the published methods, Carragee et al. [2] used Medline to retrieve published studies about rhBMP-2 use. This group did not have access to Medtronic IPD. However, they did use ''primary evidence from government and administrative databases,'' which included those of the FDA and Centers for Disease Control and Prevention, similar to those used by Fu et al. [5].

There are important additional methodological differences among the YODA reviews [3,5] and that of Carragee et al. [2]. Both of the YODA review groups performed meta-analyses of the data. In other words, they combined data from multiple studies, when statistically appropriate, to analyze a larger group of data than reported in any original single study. This method increases the probability of finding significant differences that were not apparent in smaller studies. In contrast, it could also ''wash out'' erroneous differences, created by small heterogeneous patient cohorts detected in smaller studies. Meta-analysis, however, is not without its limitations. By combining the results of different studies, methodological differences among studies can influence the test-outcome relationship. This can, if not accounted, lead to difficulty in determining the true causal relationship of differences found between the experimental and control groups.

Whereas Carragee et al. [2] did perform additional statistical analyses for individual studies (to compare published and government-derived data) according to CONSORT recommendations [6], they did not perform a meta-analysis. Furthermore, the YODA reviews examined not only adverse events but also efficacy. Although the two are inextricably linked in many ways, Carragee et al. [2] did not review outcomes, such as fusion rates comprehensively. Thus, this ''review of reviews'' will be focused primarily on adverse events as it represents the region of greatest overlap of the pertinent works.

In addition to their literature review, Carragee et al. [2] also reported the financial conflicts of interest of the authors of published studies reviewed. This was not analyzed or reported specifically in either of the YODA projects. This work will offer no opinion on conflict of interest. Rather, the focus is to analyze patient data and not to theorize about possible intentions or explanations for discrepancies.

Concerns raised in 2011

From their review in 2011, Carragee et al. [2] raised a number of concerns, as follows.

Posterolateral fusion

Carragee et al. [2] noted in both small and large studies of posterolateral fusion that there was a paradoxical effect toward increased leg pain in the rhBMP-2 group in the early postoperative period. In the article, the authors published a figure demonstrating higher percentages of ODI failures and leg pain scores in the BMP group with instrumentation compared with the iliac crest group with instrumentation. These data were from a small pilot study of 25 patients [7]. What was missing from this table and the discussion were data about the other group in this study, the BMP group without instrumentation. Regardless, Carragee et al. [2] also highlighted that there was a 10% rate of wound complications ''associated with rhBMP-2 use'' in the small pilot study, which they concluded was higher than that reported in other published studies from the same group. In one of the larger RCTs comparing BMP to autograft with posterolateral fusion [8], Carragee et al. [2] noted ''three times as many back and leg pain adverse events … during the first 3 months.''

Although Carragee et al. [2] designated a separate section and analysis for high-dose rhBMP-2 for posterolateral fusion, in this review, we do not make this distinction. In their analyses of these so-called ''high-dose'' studies, Carragee et al. [2] once again noted higher rates of early leg pain (and back pain) associated with BMP compared with iliac crest. In addition, the group noted that the FDA found ''notably increased cancer rates in the AMPLIFY [ie, BMP-2] group'' [9]. Specifically, Carragee et al. [2] reported a 3.8% rate of new cancers in the BMP group compared with 0.89% in the iliac crest group. There was minimal overlap of the confidence intervals calculated for these two groups.

To summarize, Carragee et al. [2] noted higher rates of leg pain in the early (6 weeks to 3 months) postoperative period and wound complications with rhBMP-2 and an increased cancer risk.

Relevant data from Simmonds et al.

Via meta-analysis, Simmonds et al. [3] noted a mean improvement in leg pain at all time points. Notably, this was not specific to the posterolateral approach. In examining Figure 2 in their article that detailed clinical outcomes, the confidence intervals for leg pain scores did overlap. What is interesting, however, is that mean back pain was worse in the rhBMP-2 group than in the iliac crest group in the early postoperative period (presumably at 6 weeks) with very little overlap of the confidence intervals. By what appears to be the 3-month time point, back pain was on average better in the BMP-2 group.

Meta-analytical data of adverse events at or shortly after surgery were presented in Figure 4. Here, back and leg pain was significantly more common with rhBMP-2 than iliac crest, with no apparent overlap of the confidence intervals. This seems to be somewhat incongruous with the clinical outcome data reported in Figure 2. It was also noted that the rate of infection was at least 50% higher in the BMP group compared with the iliac crest group. Appendix Figure 7 displayed data with respect to these adverse events at 2-year follow-up. Back and leg pain remained higher in the BMP group, though infection was only marginally higher. In the analysis of four key adverse events (implant-related, infection, neurologic, and pain) over all time periods for all patients, the only clear finding was increased pain at or shortly after surgery. Again, these data were not specific to posterolateral fusion.

Regarding cancer risk, Simmonds et al. [3] found that cancer ''was nearly twice as common in the rhBMP-2 recipients'' with a relative risk of 1.98. They did note that the 95% confidence intervals were wide, ranging from 14% lower to 454% higher, and that the risk was not different between so-called high-dose and low-dose applications.

Relevant data from Fu et al.

The Oregon group [5] performed separate meta-analyses for the different surgical approaches. From their analysis of IPD of posterolateral fusion, they found no significant difference in the overall incidence of adverse events between the rhBMP-2 and iliac crest groups except in early postoperative back and leg pain scores (4 weeks), which were higher in the BMP group.

The Oregon group calculated the relative risk for cancer to be 3.45 in the BMP group compared with the iliac crest group at 24-month follow-up, which was statistically significant. The difference was no longer significant at 4-year follow-up. They stated there were insufficient numbers to determine if the cancer risk was dose dependent.

Resolution of the findings

With careful examination of these findings, it appears that the concern of Carragee et al. [2] about leg pain in the early postoperative after surgery is substantiated by both reviews. Regarding cancer risk, Fu et al. [5] found a stronger association than both Carragee et al. [2] and Simmonds et al. [3] (though this risk was not found at 4-year follow-up). One might ask why discrepant cancer rates were calculated among these reviews. This is best explained by the exclusion of a single study in one analysis that was included in another. Wound complication rates (that might include infections as reported by Simmonds et al.) may be higher in the rhBMP-2 groups, but this is unclear.

Anterior lumbar interbody fusion

Carragee et al. [2] reviewed a number of industry-sponsored trials of rhBMP-2 use with anterior lumbar interbody fusion (LIF). The concerns included underreporting of the incidence/prevalence of osteolysis, device subsidence, reoperation rates, incidence of retrograde ejaculation and urinary retention, and infections in the BMP groups.

To begin, Carragee et al. [2] referenced commentaries of some of the original industry-sponsored studies that noted large degrees of osteolysis (''some extending 50% of the vertebral height'') and subsidence (''collapse of the disc space by 50%'') in the figures published. Carragee et al. [2] found FDA source documentation that noted ''[t]he incidence of adverse events that were considered device-related, including implant displacement/loosening, implant malposition and subsidence were all greater in the investigational [rhBMP-2] groups compared to the control group'' [9]. Other numbers cited from the review of nonindustry-sponsored publications were ''70% of levels showed signs of early lucency'' and ''more than 10% graft subsidence with a mean collapse of 27%.'' Most illustrative is their Table 2 in which subsidence, implant malposition/displacement/loosening, or reoperation for device-related adverse event was not reported in industry-sponsored publications in 2003 and 2004 and reported in 7, 9, and 7 patients in 2009 and reported in 7, 10, and 22 patients to the FDA in 2002. All three reports were presumably on the same group of patients at 2 years from surgery. From this same table, the industry-sponsored publications from 2003, 2004, and 2009 reported no early infections, delayed infections, retrograde ejaculation, or urogenital adverse events at 2-year follow-up. In contrast, the 2-year FDA data cited 26, 12, 12, and 36 events, respectively.

Relevant data from Simmonds et al.

Simmonds et al. [3] reported meta-analytical results from 11 industry-sponsored trials in Figure 4. Once again, this was not specific to approach. They found that the risk for implant-related events, retrograde ejaculation, and wound complications was increased by at least 50% in the BMP group at or shortly after surgery. Infections urogenital events and implant-related adverse events were also higher in the BMP group. In Appendix Figure 7 (2-year adverse event meta-analysis), infection, urogenital events, implant-related adverse events, and wound complications remained somewhat higher in the BMP group. Of note, retrograde ejaculation was substantially higher in the BMP group (odds ratio 4.76).

Relevant data from Fu et al.

As stated previously, Fu et al. [5] performed a meta-analysis on IPD for outcomes and adverse events separately for the different surgical approaches. For anterior LIF, they found 38% and 45% rates of overall adverse events by 4 weeks for BMP and iliac crest, respectively. However, they did find that retrograde ejaculation, subsidence, and urogenital problems were more common with BMP (both at 4-week and 24-month follow-up). These differences were not statistically different, and the confidence intervals were wide and overlapping. Regardless, this group calculated the relative risk for retrograde ejaculation to be 4.36 at 2 years, similar to the value found by Simmonds et al. [3].

Resolution of the findings

Simmonds et al. [3] appear to be in agreement with Carragee et al. [2] regarding concerns of retrograde ejaculation, urogenital complications, subsidence, infections, and implant-related adverse events (if this can be used as a proxy for osteolysis). Fu et al. [5] seem to corroborate these concerns, finding that retrograde ejaculation, subsidence, and urogenital problems were more common with BMP (though not statistically significant). Both reviews reported similar relative risks for retrograde ejaculation. Based on all data presented, we can offer no resolution on the issue of reoperation rates, as these were reported only by the YODA reviews.

Posterior lumbar interbody fusion

Carragee et al. [2] reviewed a number of trials of the use of BMP-2 with posterior lumbar interbody fusion (PLIF). The group raised concerns regarding bone overgrowth into the spinal canal with BMP use, ''clinical failures'' compared with iliac crest patients 6 weeks and 2 years after surgery, and higher rates of reoperation, radiculitis, osteolysis, and loss alignment. Of note, Carragee et al. [2] relied primarily on the published data from one of the original industry-sponsored RCTs that compared iliac crest with BMP-2 [10] to conclude that clinical outcomes were perhaps worse with BMP-2.

Relevant data from Simmonds et al.

The York group [3] found limited data concerning other specific adverse events with other surgical approaches. These data were derived only from published studies and presumably were specifically recorded in the IPD or government databases. From their literature review, Simmonds et al. [3] found that heterotopic bone formation and osteolysis were more common in the BMP-2 groups in comparative studies. With the exception of one study, as noted in Figure 6, these findings were derived from studies of some type of posterior interbody fusion (transforaminal LIF, PLIF, LIF).

Relevant data from Fu et al.

Fu et al. [3] did not include a specific meta-analysis of BMP studies with PLIF. Of note, Fu et al. [3] discussed that no ''trial defined radiculitis, and adverse events consistent with possible radiculitis were variously classified as back and leg pain, neurologic events, or spinal events.'' This fact makes it difficult to perform a critical analysis of these complications. This group did, however, discuss the same article that Carragee et al. [2] analyzed (comparing rhBMP-2 with iliac crest with PLIF) [10], which reported statistically higher rates of ectopic bone formation with BMP. Fu et al. [3] noted that ''the authors emphasized the lack of association between ectopic bone formation and leg pain and gave an incomplete account of the reasons for study termination.''

Resolution of the findings

In the apparent absence of IPD, Carragee et al. [2], Simmonds et al. [3], and Fu et al. [5] all relied on limited published data regarding PLIF. The common conclusion of the three groups was that heterotopic (ie, ectopic) bone formation was much more common with rhBMP-2 using this approach.

Additional concerns

Carragee et al. [2] also reviewed data concerning complications with anterior cervical fusion. We feel that this complication has not been widely contested, as it has resulted in an FDA Public Health Notification in 2008. Though the YODA reviews did include these data, these data do not seem pertinent to the current discussion. In addition, Carragee et al. [2] devoted the last section of their ''Results'' section to an exploration of possible study design features that may have biased results against iliac crest and favored rhBMP-2. Although it was addressed in some detail by Fu et al. [5] and less so by Simmonds et al. [3], this will not be addressed in the current review, as we focus primarily on the factual content of the reported data. Suffice it to say that Fu et al. [5], similar to Carragee et al. [2], found instances in which ''no adverse events because of rhBMP-2'' were reported when in fact IPD analysis found that they in fact did occur (and were recorded). We understand that this could be explained by differences of interpretation of the meaning of ''due to''; readers are encouraged to examine this closely and draw their own conclusions.

Summary

With very few exceptions, scientific studies do not demonstrate ''all or none'' or ''black or white'' outcomes. It is clear from this review of reviews that different groups examining the same articles with similar or identical methodology will not report exactly the same findings. Carragee et al. [2] raised concerns in 2011. The YODA project was intended to determine if these concerns were indeed derived from objective examination of the data or from a selective or incomplete analysis. It is certainly no secret that the work of Carragee et al. [2] has elicited the entire spectrum of sentiments ranging from appreciation and affirmation to disdain. It is the current authors' fear that the ''BMP

issue'' is or will be divisive and polarizing. This is most emphatically neither in the spirit of critical scientific thought nor in the spirit of collegiality, which has always permitted differences of opinion among people of good conscience. This is an issue of undeniable importance and one in which it is the privilege and responsibility of every provider affected to examine the facts and reach their own conclusions. No authors, including the authors of this review of reviews, are free of bias. This underscores the need for every member of NASS to look at the facts as they are presented. The current authors are hopeful, at day's end, that we will all be grateful to innovators and reviewers alike for helping to advance the art and science of medicine and for holding those advances to the highest standards.

References

[1] Krumholz HM, Ross JS, Gross CP, et al. A historic moment for open science: the Yale University Open Data Access Project and Medtronic. Ann Intern Med 2013;158:910–1.

[2] Carragee EJ, Hurwitz EL, Weiner BK. A critical review of recombinant human bone morphogenetic protein-2 trials in spinal surgery: emerging safety concerns and lessons learned. Spine J 2011;11:471–91.

[3] Simmonds MC, Brown JV, Heirs MK, et al. Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion: a meta-analysis of individual patient data. Ann Intern Med 2013;158:877–89.

[4] Rogers MA, Brown JVE, Heirs MK, et al. Reporting of industry funded study outcome data: comparison of confidential published data on the safety and effectiveness of rhBMP-2 for spinal fusion. BMJ 2013;346:3981–95.

[5] Fu R, Selph S, McDonagh M, et al. Effectiveness and harms of recombinant human bone morphogenetic protein-2 in spine fusion: a systematic review and meta-analysis. Ann Intern Med 2013;158: 890–902.

[6] Ioannidis JPA, Evans SJW, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004;141:781–8.

[7] Boden SD, Kang J, Sandhog H, Heller JG. Use of recombinant human bone morphogenetic protein-2 to achieve poster lateral lumbar spine fusion in humans: a prospective, randomized clinical pilot trial: 2002 Volvo Award in clinical studies. Spine 2002;27: 2662–73.

[8] Dimar JR, Glassman SD, Burkus JK, et al. Clinical and radiographic analysis of an optimized rhBMP-2 formulation as auto graft replacement in poster lateral lumbar spine arthrodesis. J Bone Joint Surg 2009;91:1377–86.

[9] Department of Health and Human Services, Food and Drug Administration. Food and Drug Administration Executive Summary for P050036 Medtronic's AMPLIFY™ and rhBMP-2 Matrix Orthopedic and Rehabilitation Devices Advisory Panel 2010. Gaithersburg, MD, 2010.

[10] Haid RW, Branch CL, Alexander JT, Burkus JK. Posterior lumbar interbody fusion using recombinant human bone morphogenetic protein type 2 with cylindrical interbody cages. Spine J 2004;4:527–38.