Lessons from the infuse trials: do we need a classification of bias in scientific publications and editorials?

Sohaib Hashmi · Mohamed Nouredin · Safdar N. Khan

Published online: 31 May 2014
© Springer Science+Business Media New York 2014

Abstract The original 13 Food and Drug Administration industry-sponsored recombinant human bone morphogenetic protein-2 (rhBMP-2) trials investigating its use in spinal fusion all reported no associated adverse events. However, subsequent series of studies began reporting complication rates that were much higher than those that were initially published. Critical analysis of the original rhBMP-2 industry-associated data found systematic alignment favoring positive outcomes with no proven clinical advantage over bone graft. The sources of potential bias leading to inaccurate reporting of original rhBMP-2 efficacy and safety profile include flawed study design, methodological technique, data reporting and analysis, and significant financial conflict of interest. As such, to ensure the integrity of the scientific literature, further measures should be taken by researchers, surgeons, authors, journal editors and reviewers to assess for potential sources of bias.

Keywords Infuse trials · BMP · Bias · Adverse events · Bias classification system

Introduction

The original discovery of bone morphogenetic protein (BMP) in 1965 by Urist [1] was followed by extensive preclinical and translational research in the application of these growth factors in spine fusion surgery. There was a need in selected patients to investigate alternatives and adjuncts to autologous iliac crest bone graft (ICBG) that would prove to be effective and

safe. ICBG is considered the gold standard in spinal fusion with inherent osteoinductive, osteoconductive, and neovascularization properties, fusion rates varying from 40 % to 100 % [2].

The Food and Drug Administration (FDA) initially approved rhBMP-2 (Infuse; Medtronic Sofamor Danek, Memphis, TN) in the use of anterior lumbar interbody fusions (ALIF) with the LT-CAGE (Medtronic Sofamor Danek, Memphis, TN) in patients with degenerative disc disease at 1 level from L4–S1. The authors of the original industry-sponsored or industry-associated studies of rhBMP-2 all reported the efficacy in applications including both anterior and posterior cervical and lumbar fusion surgery [3–14]. Additionally, the authors of these studies stated similar safety profiles with no adverse effects or complications attributable to rhBMP-2 in any study. Favorable research findings regarding the use of rhBMP-2 in spinal fusion surgery has undoubtedly contributed to its increased utilization from 0.7 % in 2002 to 25 % in 2006, as the off-label use of rhBMP-2 expanded to nearly 73 % of the total use by 2007 [15].

With the complexity of mechanisms and interactions of growth factors, the nature and time course of adverse events would prove to be difficult to predict. However, clear concerns in the clinical application of rhBMP-2 were cited including bony overgrowth, interaction with nearby neural structures, local and systemic toxicity, osteoclastic activation, and potential for cancer risk [16]. A divergence from the previously reported safety profile surfaced as the literature reported a greater frequency of complications associated with the use of rhBMP-2 in spinal fusion.

This disparity in the findings resulted in scrutiny by the scientific community, press, and government regarding the safety of BMP and the overarching peer review and editorial process. In a systematic review, Carragee et al concluded that the original industry-sponsored publications were systematically aligned in favor of BMP use for spinal fusion surgery, in

S. Hashmi · M. Nouredin · S. N. Khan (✉)
Division of Spine Surgery, Department of Orthopaedic Surgery,
Wexner Medical Center at The Ohio State University, 456 W 10th
Ave, Columbus, OH 43210, USA
e-mail: Safdar.Khan@osumc.edu

turn contributing to the widespread off-label use and associated the complications [17••].

The reporting and publication of the inaccurate safety profile and results of original industry-sponsored rhBMP-2 trials occurred as a result of clear bias. Bias has been defined as any process at any stage of inference that tends to produce results or conclusions that differ systematically from the truth [18]. Substantial methodological, operational, reporting, editorial, and financial bias leads one to question the efficacy of rhBMP-2 in both on-label and off-label uses. The critique of the original rhBMP-2 trials varied from the scientific community to the government including Consumer Reports, The Wall Street Journal, the Milwaukee Journal Sentinel, US Senate, the New York Times, and the Department of Justice all chiming in on this controversial topic [19••]. Following this investigation, Medtronic commissioned the Yale University Open Data Access (YODA) project to report the safety and efficacy of rhBMP-2. Using data from the original 13 trials and 31 cohort studies, these authors concluded that rhBMP-2 has no proven clinical advantage over bone graft and may be associated with significant sequelae, thereby making it difficult to identify clear indications for rhBMP-2 [20••]. In light of the subsequent conflicting assessments of rhBMP-2 with regards to efficacy and safety, it is crucial to clarify the systematic flaws of the original trials in order to improve understanding and reporting of bias in the scientific literature.

Lessons from systematic failures of the infuse trials

Systematic review of the original rhBMP-2 by Carragee et al [17••] fostered awareness of the current shortcomings in the peer review process, specifically as it pertains to reporting of novel biological technologies for clinical application. The authors' analysis of the complete data presented by the 13 industry-sponsored studies shed light upon several areas of bias, whether intentional or unintentional, from the level of the author to publication editors. The importance of maintaining a high degree of scientific integrity relies on professionalism from the investigators and peer review for clinicians and the public to understand the published literature [21•]. Appraisal in limitations of the formative publications provides understanding of potential sources of bias in the peer review process and guidance of direction in how to face these challenges.

The FDA's premarket approval (PMA) investigations were open label investigations in which researchers, surgeons, and patients knew treatment allocation of the involved subjects. Nonblinded assessment is a significant source for potential conscious or unconscious bias. A recent meta-analysis of 16 trials with subjective outcomes demonstrated that nonblinded assessors exaggerated the effectiveness of treatment with a pooled effect size of 68 % (95 % CI, 14 %–230 %). The authors concluded that the significant evidence of observer

bias in randomized clinical trials with subjective scale outcomes and failure to blind the assessors of outcomes resulted in a high risk of substantial bias [22]. In a review of randomized clinical trials published from 2002–2004 in the *Journal of Bone and Joint Surgery (American Volume)* and *Spine*, approximately half did not report attempts to reduce bias in the Materials or Methods sections [23]. While surgical studies may be more difficult to blind compared with nonsurgical trials, efforts must be made in order to minimize potential bias.

All of the original BMP studies employed noninferiority trial designs rather than superiority trial designs. This model is intended to show that a new treatment is “not acceptably worse” than the current standard treatment. Debate regarding the use of such trials began in the mid-1900s as some authors have argued that the complexity of the study design and difficulty in verification as reasons against the utilization of noninferiority models [24]. Furthermore, noninferiority trials often carry “substantial methodological flaws” and subsequent risk for type II errors [25]. A review including 162 randomized controlled trials of noninferiority and equivalence hypotheses reported that only 20 % of articles fulfilled reporting requirements specific to this study type, in addition to 12 % of trials with misleading conclusions [26]. It has been suggested that BMP use would be better proven through superiority trials [21•].

Carragee et al [17••] included analysis of control group techniques in the assessment of the industry-sponsored BMP trials. The authors reported several discrepancies from the established literature of ICBG and local bone autograft use in spine fusion. The posterolateral fusion trials comparing rhBMP-2 with ICBG did not incorporate facet preparation, thus, fusion at facet joints was not radiographically assessed. The quantity of ICBG used was also disproportionately lower with as little as 7 cc. In addition, local bone autograft harvested from the decompression was discarded rather than being applied to augment the ICBG volumes. The control group of the posterior lumbar interbody fusion (PLIF) trial may have had unnecessary ICBG harvest, as local bone autograft alone has been shown to be sufficient for fusion, thereby contributing to preventable handicap. In addition, Carragee et al noted that the reporting of ICBG harvest-site morbidity was higher (50 %–95 %) than previous estimates. The methodological techniques used in the original BMP studies place the control ICBG groups at higher risk for poorer quality fusion, nonunion, and potential failure when compared with the standard surgical method for lumbar spinal fusion [17••].

Each of the 13 original BMP trials under-reported complications and adverse events related to BMP use when compared with additional data available through the FDA and unsponsored trials. As a result, a near-perfect safety profile of BMP was established. However, Carragee et al reported

that the complication rates associated with BMP were 5 %–15 % in anterior procedures and included implant displacement, subsidence, infection, urogenital events, and retrograde ejaculation. The authors also cited complication rates associated with posterior fusion procedures ranging from 25 %–50 % including radiculitis, ectopic bone formation, osteolysis, and poorer global outcomes [17••]. Of serious concern, Carragee et al demonstrated a greater apparent risk of a new malignancy with groups treated with the 40 mg dose of rhBMP-2 per level. However, even more alarming was the lack of reporting on this issue in the original publication even though it was significantly discussed in the FDA Executive Summary [14]. Carragee et al attributed “fundamental error in the statistical analysis of uncommon and serious adverse events within each of the original studies” behind the low rates of complications. In addition, the authors cited the need for sensitivity in the detection of adverse events in noninferiority trials [17••].

As all of the original BMP trials were industry-sponsored, the authors had significant financial conflict of interests. Since the authors involved in these trials received payments during and after the study periods, a strong conflict of interest was present. Carragee et al estimated the magnitude of the financial gain, with a median payment of \$12 million for 12 of the 13 studies with no information available for 1 study [17••]. Conflict of interest statements reported in these trials were marred with vague and inconsistent statements. In combination with an absence of standardized methodology and ascertainment protocol, including successful fusion outcome, the authors were prone to bias in favor of BMP use [21•].

Recommendations of bias classification in scientific publication

Critical assessment of the original industry-sponsored rhBMP-2 trials has provided insight into the potential sources of bias that must be carefully be addressed to maintain a high degree of scientific validity. The systematic bias leading toward reporting and publication of favorable outcomes for BMP use in respected orthopedic journals brings into question the current state of the editorial review process. In addition to the use of current guidelines to ensure quality reporting, including Consolidated Standards of Reporting Trials (CONSORT) [27, 28], checklist to evaluate a report of a nonpharmacologic trial (CLEAR NPT) [29]; Preferred reporting items for systematic reviews and meta-analyses (PRISMA) [30]; Strengthening the Reporting of Observational studies in Epidemiology (STROBE) [31]; Meta-analysis Of Observational Studies in Epidemiology (MOOSE) [32], a scientific classification system of bias is necessary to further appraise the quality of the published literature.

The Levels of Evidence rating system [33] was introduced in 2003 to categorize literature based upon the study quality

and design. This 5-level rating system, with Level I evidence being randomized control trials with least bias, allowed for critical appraisal for researchers, editorial reviewers, clinicians, and the public at large regarding the quality rating of scientific articles. Obremskey et al reviewed clinical scientific articles published from January to June 2013 in 9 different orthopedic journals and reported that 11.3 % were Level I, 20.7 % were Level II, 9.9 % were Level III, and 58.1 % were Level IV [34]. The scarcity of higher level of evidence literature is further compounded by a variability of quality between the level I evidence articles. In a review of level I evidence from January 2003 to December 2004, Poolman et al demonstrated low Cochrane reporting quality scores among the individual methodological safeguards with further that scores that did not significantly differ between Level I and Level II studies [35].

The CONSORT statement checklist was established to improve the reporting of randomized control trial design, conduct, analysis, interpretation, and assessment of the validity of results (Table 1) [27, 28]. CONSORT does not deem the use of arbitrarily determined and set statistical significance ($P < 0.05$) as a criterion to identify possible associations with infrequent but serious adverse events, as those under-reported in the original rhBMP-2 trials during the pharmacologically active period [17••]. Divergence from the recommended CONSORT guidelines is not exclusive to the rhBMP-2 FDA trials. In an evaluation of the adherence to CONSORT guidelines of 196 randomized control trials in the orthopedic trauma literature, Bhandari et al reported an adherence rate of 36 % of the critical CONSORT items and 32 % of the total CONSORT items [36]. Recognizing the differences between medical and surgical trials and the limited applicability of CONSORT in surgical trials, a more applicable 10-item CLEAR NPT checklist was developed (Table 2) [29]. Chan and Bhandhari evaluated the methodological quality of 87 randomized control trials in 8 orthopedic journals from 2004 to 2008 [37]. The authors demonstrated that 84 % of studies did not clearly report the concealment of the treatment allocation while only 20 % described the experience of the surgeon. Furthermore, blinding was unclear in 55 % of patients, 72 % of ward staff, 74 % of rehabilitation staff, 46 % of clinical outcome assessors, and 38 % of nonclinical outcome assessors. A recent review of the randomized control trials in spine surgery from 2005 to 2010 reported that only 10 % report surgical experience, which is usually a description of the learning curve [38]. Equally as alarming, 77 % of studies were unclear about the concealment of treatment allocation, 68 % were unclear regarding the blinding of participants, 77 % were unclear regarding the blinding of outcome assessors, and 67 % were unclear of adhering to the intention-to-treat principle [38]. The implementation of guidelines such as CONSORT and CLEAR NPT aim to minimize methodological bias and improve the quality of reporting among the randomize control

Table 1 CONSORT 2010 updated checklist

Title and abstract	Identification as a randomized trial in the title Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts 21–31)
Introduction	
Background	Scientific background and explanation of rationale
Objectives	Specific objectives or hypotheses
Methods	
Trial design	Description of trial design (such as parallel, factorial) including allocation ratio Important changes to methods after trial commencement (such as eligibility criteria), with reasons
Participants	Eligibility criteria for participants Settings and locations where the data were collected
Interventions	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered
Outcomes	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed Any changes to trial outcomes after the trial commenced, with reasons
Sample size	How sample size was determined When applicable, explanation of any interim analyses and stopping guidelines
Randomization	
Sequence	Method used to generate the random allocation sequence
Generation	Type of randomization; details of any restriction (such as blocking and block size)
Allocation concealment mechanism	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned
Implementation	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions
Blinding	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how If relevant, description of the similarity of interventions
Statistical methods	Statistical methods used to compare groups for primary and secondary outcomes Methods for additional analyses, such as subgroup analyses and adjusted analyses
Results	
Participant flow	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome For each group, losses and exclusions after randomization, together with reasons
Recruitment	Dates defining the periods of recruitment and follow-up Why the trial ended or was stopped
Baseline data	A table showing baseline demographic and clinical characteristics for each group
Numbers analyzed	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups
Outcomes and estimation	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95 % CI) For binary outcomes, presentation of both absolute and relative effect sizes is recommended
Ancillary analyses	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory
Harms	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms 28)
Discussion	
Limitations	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses
Generalizability	Generalizability (external validity, applicability) of the trial findings
Interpretation	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence
Other information	
Registration	Registration number and name of trial registry
Protocol	Where the full trial protocol can be accessed, if available
Funding	Sources of funding and other support (such as supply of drugs), role of funders

From [28], with permission

Table 2 Final checklist of items to assess quality of randomized controlled trials of nonpharmacologic treatment

	Possible answers
1. Was the generation of allocation sequences adequate?	Yes; No; Unclear
2. Was the treatment allocation concealed?	Yes; No; Unclear
3. Were details of the intervention administered to each group made available?	Yes; No; Unclear
4. Were care providers' experience or skill in each arm appropriate?	Yes; No; Unclear
5. Was participant (ie, patients) adherence assessed quantitatively?	Yes; No; Unclear
6. Were participants adequately blinded?	Yes; No, because blinding is not feasible; No, although blinding is feasible; Unclear
6.1. If participants were not adequately blinded	
6.1.1. Were all other treatments and care (ie, co-interventions) the same in each randomized group?	Yes; No; Unclear
6.1.2. Were withdrawals and lost to follow-up the same in each randomized group?	Yes; No; Unclear
7. Were care providers or persons caring for the participants adequately blinded?	Yes; No, because blinding is not feasible; No, although blinding is feasible; Unclear
7.1. If care providers were not adequately blinded	
7.1.1. Were all other treatments and care (ie, co-interventions) the same in each randomized group?	Yes; No; Unclear
7.1.2. Were withdrawals and lost to follow-up the same in each randomized group?	Yes; No; Unclear
8. Were outcome assessors adequately blinded to assess the primary outcomes?	Yes; No, because blinding is not feasible; No, although blinding is feasible; Unclear
8.1. If outcome assessors were not adequately blinded, were specific methods used to avoid ascertainment bias (systematic differences in outcome assessment)?	Yes; No; Unclear
9. Was the follow-up schedule the same in each group?	Yes; No; Unclear
10. Were the main outcomes analyzed according to the intention-to-treat principle?	Yes; No; Unclear

From [29], with permission

trials, especially in the current time of reliance upon comparative effectiveness research. Tangible progress, however, may only be realized with more standardized and strict adherence to the established guidelines.

With the strong relationship between the medical device industry and the orthopedic surgery specialty, conflicts of interest are simply unavoidable. Favorable outcomes like those seen in device-sponsored rhBMP-2 trials are not unique. Results are favorable to the implant manufacturer in 73 % of industry-sponsored studies compared with 44 % of independent studies (odds ratio, 3.3; 95 % CI, 2.4–4.5) for spine implants; 93 % and 37 %, respectively, for hip implants; and 75 % and 20 % for knee implants [39]. With the Food and Drug Administration and Public Health Service guidelines suggesting that \$5000 to \$25,000 of financial relationships with a study sponsor pose a risk of systematic bias, Carragee et al argue that it is impossible to estimate the potential bias from the multimillion-dollar financial industry dealings that involved the original rhBMP-2 trials [40•].

In an effort to increase transparency and potential bias, required disclosures of industry relationships been implemented by many journals. Yet, there is substantial variability of reported financial disclosure between journals. One industry-sponsored study investigating rhBMP-2 use with posterior spinal instrumentation disclosed that “The manuscript submitted does not contain information about medical device(s)/

drug(s). Institutional funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript” [12]. Furthermore, the accuracy of industry payment disclosure has significant incongruity. In an evaluation of inconsistencies between physician disclosures and payments reported by industry, authors found that Medtronic and Depuy Spine payments were disclosed by 12.1 % and 8.75 % of physicians attending the 2011 North American Spine Society annual meeting, respectively [41•]. This physician reporting discrepancy is especially concerning, as The Physician Payment Sunshine Act (PPSA) will require the industry to disclose anything of value that has been given to physicians by 2014. It is clear that more standardized practices must be adopted to decrease disclosure discrepancy and increase transparency.

The role for greater transparency may extend to editorial staff of journals in addition to authors and researchers. Journal publication of industry-sponsored studies with inherent systematic alignment for favorable results contributes to a significant source of potential bias in spine literature. In a retrospective review of articles published in the journal *Spine* from January 2002 to July 2003, the authors reported that industry support was noted for 15.9 %, foundation support for 12.7 %, government support for 10.2 %, and institution support for 3.2 % [42]. The odds ratio regarding the association between

industry funded reporting of positive results was 3.3 times that of studies with other funding sources ($P < 0.001$). The authors suggest that study design bias, experimental technique bias, result interpretation bias, or publication bias as potential sources of bias. One author with significant financial conflicts of interest in the original rhBMP-2 trials also served as the Editor-in-Chief of a journal that published these articles [5, 8]. Editorial conflicts of interest have played a role in the systematic failure surrounding the original rhBMP-2 papers reporting favorable outcomes [19••]. To further minimize financial conflicts of interest, editorial disclosures may address the increasingly industry-sponsored landscape of the scientific literature.

Conclusions

The original 13 rhBMP-2 industry-sponsored trials significantly under-reported adverse events and complications associated with rhBMP-2 use in spinal fusion surgery, thereby contributing to widespread promotion of on- and off-label use. The subsequent literature review, however, found serious complications ranging from 20 % to 70 % in all settings, which were 10–50 times the original estimates from the industry-associated trials [17••]. The YODA Project, commissioned to report on the safety and efficacy of rhBMP-2, concluded that “rhBMP-2 has no proven clinical advantage over bone graft and may be associated with important harms, making it difficult to identify clear indications for rhBMP-2. Earlier disclosure of all relevant data would have better informed clinicians and the public than the initial published trial reports did” [20••]. A significant degree of systematic bias undoubtedly contributed to the alignment of favorable outcomes reported in earlier industry-sponsored rhBMP-2 trials, while later independent investigations of the same data from FDA randomized control trials led to differing conclusions. In an effort to better assess the extent of the present and potential bias in the scientific literature, standardized classification and reporting of study design, methodology, data analysis and reporting, and financial conflicts of interests must be advocated. This responsibility must be shared between the scientific community including researchers, surgeons, authors, editors, and reviewers to ensure sufficient integrity of the published literature, which guides clinical decision-making and patient treatment.

Compliance with Ethics Guidelines

Conflict of Interest Sohaib Hashmi, Mohamed Noureldin, and Safdar N. Khan declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

1. Urist MR. Bone: formation by autoinduction. *Science*. 1965;150: 893–9.
2. Grabowski G, Cornett CA. Bone graft and bone graft substitutes in spine surgery: current concepts and controversies. *J Am Acad Orthop Surg*. 2013;21:51–60. doi:10.5435/jaas-21-01-51.
3. Boden SD, Zdeblick TA, Sandhu HS, Heim SE. The use of rhBMP-2 in interbody fusion cages. Definitive evidence of osteoinduction in humans: a preliminary report. *Spine*. 2000;25:376–81.
4. Boden SD, Kang J, Sandhu H, Heller JG. Use of recombinant human bone morphogenetic protein-2 to achieve posterolateral lumbar spine fusion in humans: a prospective, randomized clinical pilot trial: 2002 Volvo Award in clinical studies. *Spine*. 2002;27: 2662–73. doi:10.1097/01.brs.0000035320.82533.06.
5. Burkus JK, Gornet MF, Dickman CA, Zdeblick TA. Anterior lumbar interbody fusion using rhBMP-2 with tapered interbody cages. *J Spinal Disord Tech*. 2002;15:337–49.
6. Burkus JK, Transfeldt EE, Kitchel SH, Watkins RG, Balderston RA. Clinical and radiographic outcomes of anterior lumbar interbody fusion using recombinant human bone morphogenetic protein-2. *Spine*. 2002;27:2396–408. doi:10.1097/01.brs.0000030193.26290.dd.
7. Baskin DS, Ryan P, Sonntag V, Westmark R, Widmayer MA. A prospective, randomized, controlled cervical fusion study using recombinant human bone morphogenetic protein-2 with the CORNERSTONE-SR allograft ring and the ATLANTIS anterior cervical plate. *Spine*. 2003;28:1219–24. doi:10.1097/01.brs.0000065486.22141.ca. discussion 25.
8. Burkus JK, Heim SE, Gornet MF, Zdeblick TA. Is INFUSE bone graft superior to autograft bone? An integrated analysis of clinical trials using the LT-CAGE lumbar tapered fusion device. *J Spinal Disord Tech*. 2003;16:113–22.
9. Haid Jr RW, Branch Jr 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. doi:10.1016/j.spinee.2004.03.025. discussion 38–9.
10. Boakye M, Mummaneni PV, Garrett M, Rodts G, Haid R. Anterior cervical discectomy and fusion involving a polyetheretherketone spacer and bone morphogenetic protein. *J Neurosurg Spine*. 2005;2:521–5. doi:10.3171/spi.2005.2.5.0521.
11. Burkus JK, Sandhu HS, Gornet MF, Longley MC. Use of rhBMP-2 in combination with structural cortical allografts: clinical and radiographic outcomes in anterior lumbar spinal surgery. *J Bone Joint Surg Am*. 2005;87:1205–12. doi:10.2106/jbjs.d.02532.
12. Glassman SD, Dimar III JR, Burkus K, Hardacker JW, Pryor PW, Boden SD, et al. The efficacy of rhBMP-2 for posterolateral lumbar fusion in smokers. *Spine*. 2007;32:1693–8. doi:10.1097/BRS.0b013e318074c366.
13. Dawson E, Bae HW, Burkus JK, Stambough JL, Glassman SD. Recombinant human bone morphogenetic protein-2 on an absorbable collagen sponge with an osteoconductive bulking agent in posterolateral arthrodesis with instrumentation. A prospective randomized trial. *J Bone Joint Surg Am*. 2009;91:1604–13. doi:10.2106/jbjs.g.01157.
14. Dimar II JR, Glassman SD, Burkus JK, Pryor PW, Hardacker JW, Carreon LY. Clinical and radiographic analysis of an optimized rhBMP-2 formulation as an autograft replacement in posterolateral lumbar spine arthrodesis. *J Bone Joint Surg Am*. 2009;91:1377–86. doi:10.2106/jbjs.h.00200.

15. Spengler DM. Resetting standards for sponsored research: do conflicts influence results? *Spine J.* 2011;11:492–4. doi:10.1016/j.spinee.2011.05.001. *Commentary on Carragee et al "A critical review of recombinant human bone morphogenetic protein-2 trials in spinal surgery: emerging safety concerns and lessons learned" raises questions regarding state of peer-review process, conflicts of interest, and regulation and use of rhBMP-2.*
16. Poynton AR, Lane JM. Safety profile for the clinical use of bone morphogenetic proteins in the spine. *Spine.* 2002;27(16 Suppl 1): S40–8.
17. 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. doi:10.1016/j.spinee.2011.04.023. *This article compared data published in the original 13 industry-sponsored rhBMP-2 studies with data provided in the FDA website and spinal literature investigating BMP use concluding that the original industry-sponsored articles systemically aligned to report positive results of BMP use in spinal fusion.*
18. Owen R. Reader bias. *JAMA.* 1982;247:2533–4. doi:10.1001/jama.1982.03320430037027.
19. Carragee EJ, Ghanayem AJ, Weiner BK, Rothman DJ, Bono CM. A challenge to integrity in spine publications: years of living dangerously with the promotion of bone growth factors. *Spine J.* 2011;11: 463–8. doi:10.1016/j.spinee.2011.06.001. *This editorial states the precedence for critical review and cause of discrepancy between the original industry-sponsored BMP studies compared with later studies reporting higher rate of adverse events with use of BMP.*
20. Fu R, Selph S, McDonagh M, Peterson K, Tiwari A, Chou R, 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. doi:10.7326/0003-4819-158-12-201306180-00006. *This independent systematic review and meta-analysis of 17 industry-sponsored studies and subsequent publications regarding rhBMP-2 use in spine fusion surgery concluded rhBMP-2 has no proven clinical advantage over bone graft and may be associated with important harms, unclear indications for proper use of rhBMP-2, and earlier disclosure of original BMP studies would have better informed clinicians and the public.*
21. Mirza SK. Folly of FDA-approval studies for bone morphogenetic protein. *Spine J.* 2011;11:495–9. doi:10.1016/j.spinee.2011.05.009. *This commentary on Carragee et al "A critical review of recombinant human bone morphogenetic protein-2 trials in spinal surgery: emerging safety concerns and lessons learned" discusses the systematic bias leading to the reporting of highly favorable results in the original industry-sponsored spinal fusion studies with rhBMP-2.*
22. Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ.* 2013;185:E201–11. doi:10.1503/cmaj.120744.
23. Khan SN, Mermer MJ, Myers E, Sandhu HS. The roles of funding source, clinical trial outcome, and quality of reporting in orthopedic surgery literature. *Am J Orthop.* 2008;37:E205–12. discussion E12.
24. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet.* 2007;370:1875–7. doi:10.1016/S0140-6736(07)61604-3.
25. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med.* 2006;145:62–9.
26. Le Henaff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA.* 2006;295:1147–51. doi:10.1001/jama.295.10.1147.
27. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA.* 1996;276:637–9.
28. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c332. doi:10.1136/bmj.c332.
29. Boutron I, Moher D, Tugwell P, Giraudeau B, Poiraudeau S, Nizard R, et al. A checklist to evaluate a report of a nonpharmacological trial (CLEAR NPT) was developed using consensus. *J Clin Epidemiol.* 2005;58:1233–40. doi:10.1016/j.jclinepi.2005.05.004.
30. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol.* 2009;62:1006–12. doi:10.1016/j.jclinepi.2009.06.005.
31. Elm EV, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ.* 2007;335:806–8. doi:10.1136/bmj.39335.541782.AD.
32. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA.* 2000;283:2008–12.
33. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am.* 2003;85-A:1–3.
34. Obrebsky WT, Pappas N, Attallah-Wasif E, Tornetta III P, Bhandari M. Level of evidence in orthopaedic journals. *J Bone Joint Surg Am.* 2005;87:2632–8. doi:10.2106/jbjs.e.00370.
35. Poolman R, Struijs P, Krips R, Siersevelt I, Lutz K, Bhandari M. Does a "Level I Evidence" rating imply high quality of reporting in orthopaedic randomised controlled trials? *BMC Med Res Meth.* 2006;6:44.
36. Bhandari M, Guyatt GH, Lochner H, Sprague S, Tornetta III P. Application of the Consolidated Standards of Reporting Trials (CONSORT) in the fracture care literature. *J Bone Joint Surg Am.* 2002;84-A:485–9.
37. Chan S, Bhandari M. The quality of reporting of orthopaedic randomized trials with use of a checklist for nonpharmacological therapies. *J Bone Joint Surg Am.* 2007;89:1970–8. doi:10.2106/jbjs.f.01591.
38. van Oldenrijk J, van Berkel Y, Kerkhoffs GM, Bhandari M, Poolman RW. Do authors report surgical expertise in open spine surgery related randomized controlled trials? A systematic review on quality of reporting. *Spine.* 2013;38:857–64. doi:10.1097/BRS.0b013e31827ecb1c.
39. Gelberman RH, Samson D, Mirza SK, Callaghan JJ, Pellegrini Jr VD. Orthopaedic surgeons and the medical device industry: the threat to scientific integrity and the public trust. *J Bone Joint Surg Am.* 2010;92:765–77. doi:10.2106/jbjs.i.01164.
40. Carragee EJ, Baker RM, Benzel EC, Bigos SJ, Cheng I, Corbin TP, et al. A biologic without guidelines: the YODA project and the future of bone morphogenetic protein-2 research. *Spine J.* 2012;12:877–80. doi:10.1016/j.spinee.2012.11.002. *This editorial comments on fundamental flaws in existing studies investigating the efficacy and safety involving BMP-2 in spine surgery, including limitations in methodological, operational, and reporting issues.*
41. Buerba RA, Fu MC, Grauer JN. Discrepancies in spine surgeon conflict of interest disclosures between a national meeting and physician payment listings on device manufacturer web sites. *Spine J.* 2013;13:1780–8. doi:10.1016/j.spinee.2013.05.032. *In this comparison of publically available disclosure/payment data, authors concluded discrepancy rates between what spine surgeons disclosed at NASS 2011 and what companies reported for their consultants were high demanding need for more standardized practice in reporting of financial conflict of interests.*
42. Shah RV, Albert TJ, Bruegel-Sanchez V, Vaccaro AR, Hilibrand AS, Grauer JN. Industry support and correlation to study outcome for papers published in *Spine*. *Spine.* 2005;30:1099–104. discussion 105.