1. Authors: Original Cover Letter (18 October 2012)

To the editor,

On behalf of my co-authors I have submitted the manuscript “The safety and efficacy of recombinant human bone morphogenetic protein-2 (rhBMP-2) for spinal fusion: An individual patient data meta-analysis” for consideration by Annals of Internal Medicine. This paper is drawn form a wider report on the same topic performed by the Centre for Reviews and Dissemination as part of the Yale Open Access Data (YODA) project. As such, we understand that the journal has provisionally agreed to publish this article as previously discussed with the YODA team.

To ensure that this submission remains entirely separate for the parallel submission on the same topic by the Oregon group, we are submitting only title and contact details via the online submission process, as instructed by Christine Laine.

We will email the full manuscript to Mary Beth Schaeffer directly for subsequent manual distribution to peer reviewers.

Yours faithfully,
Dr Mark Simmonds,
on behalf of the authors.
2. **Authors: Original Submitted Manuscript (18 October 2012)**

The safety and efficacy of recombinant human bone morphogenetic protein-2 (rhBMP-2) for spinal fusion

An individual patient data meta-analysis

**Running title:**
*Bone morphogenetic protein-2 in spinal fusion*

Mark C Simmonds¹
Jennifer VE Brown¹
Morag K Heirs¹
Julian PT Higgins¹,²
Richard J Mannion³
Mark A Rodgers¹
Lesley A Stewart ¹*

1 Centre for Reviews and Dissemination, University of York, UK
2 MRC Biostatistics Unit, Cambridge, UK
3 Addenbrookes Hospital, Cambridge, UK

* Corresponding author
Centre for Reviews and Dissemination
University of York, UK
YO10 5DD
Tel: 01904 321040
Email: lesley.stewart@york.ac.uk

Project funded by a competitive grant from the Yale Open Data Access (YODA) project.

Word count: 3400 words
Abstract

Background

Recombinant human bone morphogenetic protein-2 (rhBMP-2) is widely used to promote fusion in spinal surgery. Recent controversy around its safety has led to the need for a thorough evaluation of the evidence supporting its use.

Purpose

To evaluate the efficacy and safety of rhBMP-2 in patients undergoing spinal surgery through systematic review and meta-analysis of individual participant data (IPD).

Data sources

IPD were available from the manufacturers, via the Yale Open Data Access project. Data for one further trial were identified for the systematic review.

Study selection

Randomised controlled trials of rhBMP-2 versus Iliac crest bone graft in patients undergoing spinal fusion surgery for degenerative disc disease and related spinal conditions were eligible. For evaluation of safety, published evidence from comparative studies in the same types of participants was also eligible.

Outcomes

Pain and physical function (Oswestry disability index (ODI), SF-36), successful spinal fusion and a range of adverse events.

Data synthesis

IPD from 1411 patients were analysed. Two years after surgery rhBMP-2 patients had ODI scores 3.5 percentage points lower than ICBG patients (95% CI: 0.49 to 6.47); radiographic fusion was 12% more common (95% CI: 2% to 23%). At or shortly after surgery, several adverse events were more common among rhBMP-2 patients. Cancer was almost twice as common among rhBMP-2 patients (RR 1.84, 95% CI: 0.81 to 4.16). Adverse event data
suggested that heterotopic bone formation, dysphagia, osteolysis and retrograde ejaculation might be more common in rhBMP-2 surgery.

Conclusion

The use of rhBMP-2 over ICBG in spinal fusion surgery increases fusion rates but does not result in clinically significant improvements in pain. There is evidence that rhBMP-2 increases early post-surgical pain and may be associated with other adverse events including cancer.

**Introduction**

Recombinant human bone morphogenetic protein-2 (rhBMP-2) is widely used as an alternative to iliac crest bone graft (ICBG) to promote fusion in spinal surgery (1) (2). Following FDA approval of rhBMP-2 in anterior lumbar interbody fusion (ALIF) surgery (3), numbers of spinal procedures using rhBMP-2 have grown rapidly, including off-label indications (2) (4). Publications of manufacturer-sponsored trials reported positive findings with remarkably few adverse events (5). Recently, concerns have grown about adverse events, with a review of publicly available data suggesting that the risk of adverse events is 10 to 50 times higher than reported in trial publications (6).

The Yale Open Data Access (YODA) project invited the manufacturer, Medtronic, to provide full data from all their trials of rhBMP-2, to allow independent re-analysis. This provided an opportunity to undertake an individual participant data (IPD) meta-analysis, which is regarded as a ‘gold standard’ approach to synthesis, and allows more in-depth analysis than is possible in a review of the published literature. YODA subsequently funded the Centre for Reviews and Dissemination and one other group to conduct such an analysis in the context of a full systematic review, on the basis of a competitive bidding process.

Our aim was to provide a robust and fair assessment of the benefits and harms of rhBMP-2 in spinal fusion surgery, examining all available research evidence. We conducted a systematic
review and IPD meta-analysis of randomised controlled trials (RCTs) that compared spinal fusion surgery using rhBMP-2 with ICBG, and, for assessments of safety, we also analysed comparative randomised and non-randomised studies. Our research was conducted entirely independently of Medtronic. Our full report to YODA is publicly available and examines a wide range of issues around the efficacy and safety of rhBMP-2 in spinal surgery. It also includes an examination of the reliability of the literature published in this field (which will be submitted for publication separately) (7) (8). Here, we focus on the clinical outcomes following spinal surgery using rhBMP-2, including both efficacy and safety.

Methods

Eligibility, search, data collection and critical assessment

Our methods were, in advance of knowledge of the data to be provided, pre-specified in a protocol, registered in PROSPERO in February 2012 (registration number CRD42012001907) (9).

We sought IPD from all randomised controlled trials comparing rhBMP-2 with ICBG in spinal fusion surgery, irrespective of spinal level or surgical approach. We included trials of the licensed INFUSE® rhBMP-2 formulation and unlicensed AMPLIFY ™ and BCP formulations. For analyses of safety we additionally sought all randomised and non-randomised studies comparing rhBMP-2 with any other spinal fusion technique and reporting adverse events.

We performed a systematic search of multiple bibliographic databases including CENTRAL, MEDLINE, EMBASE and Science Citation Index in January 2012, and automated ‘current awareness’ searches up to June 2012, supplemented by a published call for evidence (see web appendix for details). We were provided with IPD from 17 trials undertaken by Medtronic. Twelve of these were RCTs and five were single-arm trials of rhBMP-2. We requested IPD
from authors of the two further RCTs, and obtained these from one (10). Data extraction from published reports was undertaken by two researchers independently with any differences resolved by discussion. We assessed risk of bias in the RCTs using the Cochrane Collaboration’s ‘risk of bias’ tool, and using the Newcastle-Ottawa scale for non-randomised studies (11)(12).

Efficacy analysis

We addressed outcomes likely to be important to patients: Oswestry Disability Index (ODI; or Neck Disability Index (NDI) for cervical spinal surgery); SF-36 Physical Component Score (SF-36 PCS); back and leg pain; and successful spinal fusion (as defined radiographically by Medtronic). Data on these outcomes were analysed at 6 weeks, 3 6, 12 and 24 months after surgery. We did not analyse the limited data for longer follow-up times because these were available for only a minority of the patients. We also considered four secondary outcomes: duration of hospital stay, operating time, successful return to work / usual activity and use of pain-relief medication.

For analyses of efficacy we used standard two-stage meta-analytic techniques (13) (14). IPD from each trial were analysed separately, using the same methods across trials; resulting summary statistics were then combined across trials. Separate meta-analyses were performed for each of the specified time points. For continuously distributed outcomes (such as ODI) we used mean differences between treatment arms in the change in score from pre-operative values. For dichotomous outcomes (such as successful fusion) we used relative risks. Heterogeneity was assessed in all meta-analyses using the Higgins’ $I^2$ statistic (15) and Cochran’s Q test. We examined whether effects varied according to the type of spinal surgery, or by rhBMP-2 formulation (INFUSE® or AMPLIFY ™) using subgroup analysis. We investigated whether patient-level factors (age, sex, smoking, alcohol consumption, body mass index, diabetic status, and history of spinal surgery for back pain) were associated with
the efficacy of rhBMP-2 surgery using a one-stage random-effects regression model (16) including interaction terms between patient-level factors and treatment. In an analysis not specified in our protocol, we also investigated whether successful fusion affected pain and function outcomes using similar one-stage random-effects linear regression models.

Safety analysis

For evaluation of safety, we examined numbers of adverse events, including cancer, but were limited by classifications of adverse events produced by Medtronic. Because the numbers of specific adverse events in most trials were generally small, we used one-stage random-effects logistic regression meta-analysis models (17). Results of these analyses are presented as odds ratios. From other published studies that compared spinal fusion surgery using rhBMP-2 with ICBG and other comparators, we extracted data on numbers of adverse events as specified in our protocol. This was restricted to comparative studies including more than ten adult participants.

Results

Our IPD meta-analysis was based on data from 1,411 patients included in 11 Medtronic RCTs that compared rhBMP-2 with ICBG surgery. Data from one Medtronic trial (18) were not included because the comparator arm was not ICBG surgery. Analyses also included IPD from one further RCT of 106 patients not sponsored by Medtronic (10) when the outcome was included in the data available. Table 1 summarises the 12 trials.

In assessments of risk of bias, we found that the randomisation and allocation concealment procedures were adequate for all trials. Neither patients nor physicians were blinded to the treatment received, and all pain and function outcomes were patient assessed, so there was a potential for bias in these outcomes. Assessment of successful fusion was performed blind to treatment received, minimising bias.

Efficacy
Pain and function

Figure 1 illustrates meta-analyses across the 12 trials for four pain and function outcomes. Points on the plot represent mean differences between rhBMP-2 and ICBG at each time point, with vertical lines showing the 95% confidence intervals for these. For the first three outcomes, points below the horizontal zero line indicate a benefit of rhBMP-2; for SF-36 PCS, points above the line indicate a benefit of rhBMP-2.

The use of rhBMP-2 generally reduced pain when compared to ICBG from six months after surgery onwards: at two years after surgery the ODI was approximately 3.5 percentage points better (MD -3.48, 95%CI: -6.47 to -0.49, $I^2 = 38\%$) and back pain was similarly better among rhBMP-2 patients, by more than one point on the 20-point scale used (MD -1.58 95%CI: -2.65 to -0.51, $I^2 = 44\%$). SF-36 PCS was approximately two percentage points higher for patients receiving rhBMP-2 (MD 1.93 95%CI: 0.63 to 3.22, $I^2 = 0$) at two years.

Improvements in pain and function were statistically significant but small. There was no evidence of a difference in leg pain between treatment groups. Patients in both groups improved considerably over time such that the extra benefit of rhBMP-2 over ICBG surgery was small in comparison (Appendix figure 1). ODI improved by 25 percentage points at two years for ICBG patients and by approximately 30 percentage points for rhBMP-2 patients.

Fusion

Meta-analyses of success of spinal fusion at 6, 12 and 24 months are illustrated in Figure 2. This shows that rhBMP-2 increased the rate of spinal fusion by 10 to 20% (where 69% of ICBG patients achieved fusion within two years). These differences in rates were consistent over time and were statistically significant. There was however substantial heterogeneity in the relative risk of successful fusion across trials at earlier time points, with $I^2$ at six months, one year and two years being 97%, 80% and 76% respectively. This heterogeneity is apparent...
in Figure 3 which presents a forest plot for successful fusion two years after surgery. Data on fusion were not available for the LT-CAGE®pilot and Glassman trials.

Investigation by surgical approach

We investigated whether the efficacy of rhBMP-2 varied across anterior lumbar interbody fusion (ALIF), posterior lumbar fusion (TLIF/PLIF/PLF), and anterior cervical spinal surgery (see Table 1) using subgroup analysis. Figure 4 shows the results of these subgroup analyses for ODI and successful fusion two years after surgery. There was moderate evidence of a difference between surgery types for ODI (p-value 0.065), but this was primarily due to the very large benefit of rhBMP-2 on the NDI in the single, small cervical surgery trial (23 patients), excluding this trial resulted in no statistically significant difference between anterior or posterior approaches (p = 0.17). There was no evidence of a difference in the relative risks of successful fusion at two years across surgery types (p-value 0.88). There was no evidence of any difference between surgery types for any other outcome.

We also considered whether a range of patient-level factors were associated with the effectiveness of rhBMP-2. There was generally no evidence of interactions between rhBMP-2 and the patient-level factors (age, sex, smoking, alcohol consumption, body mass index, diabetic status, and history of spinal surgery). One possible exception was that, for people with a history of previous spinal surgery, there was no difference in the effectiveness of rhBMP-2 and ICBG, either at reducing ODI or improving fusion rates (whereas there was an overall benefit of rhBMP-2 across all patients). Given the number of analyses performed, this result may be a chance finding.

Further outcomes

Among secondary outcomes there was no overall evidence of any meaningful difference in duration of hospital stay (mean difference -0.15 days, 95%CI: -1.35 to 1.06 days) or that rhBMP-2 surgery increased the probability of returning to work or usual activity when
compared with receiving ICBG (for example, at two years: RR 1.01, 95%CI: 0.88 to 1.17).
There was some evidence that using rhBMP2 shortened operating times, by 21 minutes (95%
CI: 15 to 27, see Appendix figure 2), from an average of 135 minutes. There was no evidence
that use of pain-relief differed between rhBMP-2 and ICBG arms at any time (see Appendix
figure 3).
We also explored the relationship between successful fusion and improvement in pain and
function scores. Figure 5 shows the clinical outcome scores (ODI, SF-36 PCS, back pain) for
the four categories of patient who: (i) received ICBG and achieved fusion; (ii) received ICBG
and did not achieve fusion; (iii) received rhBMP-2 and achieved fusion; (iv) received
rhBMP-2 and did not achieve fusion. For patients receiving ICBG surgery, successful fusion
had almost no impact on back pain or SF-36 PCS when compared to ICBG patients who had
not achieved fusion. One year after surgery there was an 11 percentage point greater
improvement on ODI among patients with successful fusion compared with those who did
not achieve fusion, which was statistically significant, but two years after surgery this
reduced to 4.5 points; this difference was not statistically significant. Thus it is not apparent
from these data that achieving fusion necessarily translates to improved patient outcomes.
For rhBMP-2 patients there were slightly larger differences in ODI and back pain (but not
SF-36 PCS) between patients who did and did not achieve fusion, greater (and statistically
significant) benefits being reported by those who fused compared with those who did not.
Also rhBMP-2 patients with fusion had slightly better outcomes than ICBG patients with
fusion who, and rhBMP-2 patients without fusion had worse outcomes than ICBG patients
without fusion. A possible explanation for this is that patient’s knowledge of both treatment
and fusion status led to bias in the reporting of pain and function outcomes.

Safety
The trials sponsored by Medtronic did not generally record as adverse events several of the outcomes that we had intended to explore, including heterotopic bone formation, osteolysis and radiculitis. We investigated the available adverse events according to the categorisations provided by Medtronic. The numbers of adverse events in the 11 Medtronic RCTs (again excluding the MAVERICK™ trial) are given in Table 2.

Figure 6 illustrates one-stage meta-analyses for adverse events across these 11 Medtronic RCTs at, or shortly after (up to four weeks), the time of surgery. Some types of adverse event were more common among rhBMP-2 patients. Arthritis/bursitis, back and leg pain, implant-related events, neurological events, other pain, retrograde ejaculation wound complications and vascular events all had at least a 50% increase in risk of occurring among rhBMP-2 patients. Because of the rarity of most events, these differences were statistically significant only for back and leg pain (OR 1.92, 95% CI: 1.14 to 3.25). Arm and neck pain, dysphagia and spinal events (including disc prolapse and stenosis) were less common among rhBMP-2 patients, but results were not statistically significant.

Figure 7 shows the results of meta-analyses for four key categories of adverse event (implant related, infections, neurological, any pain) across all time periods. These do not show evidence of a difference in the risk of adverse events from three months after surgery onwards. At, or shortly after, surgery implant related events were more common in rhBMP-2 patients, as were neurological events, but neither was statistically significant. Pain was significantly more common in the rhBMP-2 patients at or shortly after surgery (OR 1.80, 95% CI: 1.09 to 3.00), and was higher six weeks after surgery, but not different from ICBG patients thereafter. This contrasts with the results seen in the analyses of ODI, SF-36 PCS and back pain in the efficacy analyses where pain was generally significantly lower in the rhBMP-2 patients from three months after surgery onwards.

Cancer
Table 3 summarises the cancers observed in the 11 Medtronic RCTs. It excludes pre-existing cancers but does include three cases of cancer in the INFUSE®/LT-CAGE® open pivotal trial (one thyroid, one testicular, and one melanoma) identified through extended follow-up of only the rhBMP-2 patients. Five RCTs observed at least one cancer case. A one-stage random-effects meta-analysis model for the incidence of cancer (excluding the three cancers identified through extended follow-up) found that cancer was nearly twice as common among rhBMP-2 patients (RR 1.98, 95% CI 0.86 to 4.54) but the result was not statistically significant due to the rarity of cancers. A forest plot for the equivalent two-stage analysis is given in Figure 8 (RR 1.84, 95% CI: 0.81 to 4.16). The relative risk of cancer was similar across trials. In particular, the relative risk of cancer in the AMPLIFY™ trial, which used a different preparation of rhBMP-2 at a higher dose, was no greater than in those trials which used INFUSE. We note that there were also three cancers among rhBMP-2 patients in the MAVERICK™ trial and two in a single-arm Medtronic trial (not included in the analyses here).

**Adverse events reported in the wider literature**

We supplemented the above analyses with adverse events data from randomised and non-randomised comparative studies. These studies used a variety of spinal fusion techniques as a control, including ICBG, local bone graft, allograft and bone marrow aspirates. Most studies were observational, so confounding due to differences in the types of patients receiving the different types of surgery cannot be ruled out. For these reasons we did not combine studies in meta-analyses. These studies are summarised in Appendix table 1.

Despite these methodological issues there is some suggestive evidence of higher rates of particular adverse events among rhBMP-2 patients. In particular, heterotopic bone formation was more common among rhBMP-2 patients, though whether this led to any clinical consequences for those patients was unclear (see Appendix figure 4). Leg pain and...
radiculitis appeared to be more common among rhBMP-2 patients, as had been observed in the Medtronic trials (see Appendix figure 5). Osteolysis was more common among rhBMP-2 patients, but only two studies reported on this event. Dysphagia appeared to be more common for rhBMP-2 patients in cervical spinal surgery, although there were some inconsistencies in the results of the studies reporting on dysphagia (see Appendix figure 6). Comer et al (19) compared four consecutive-patient cohorts undergoing ALIF with or without rhBMP-2, reporting a statistically significantly higher rate of retrograde ejaculation among rhBMP-2 treated patients (6.3% vs 0.9%, p=0.0012).

Discussion

This IPD meta-analysis has shown that the use of rhBMP-2 in spinal surgery improves back pain and quality of life when compared with ICBG surgery at between six months and two years after surgery, but these differences are small, and fall below the thresholds of what are regarded to be clinically meaningful differences (estimated as being between 4 and 17 points for ODI and 5.4 for SF-36 PCS (20), (21). There was no evidence of a difference in reductions in leg pain between the treatment groups, nor did rhBMP-2 surgery reduce the use of pain-relief medication. Some ICBG patients experienced pain at the graft site but, as pain at this same anatomical site was not assessed in rhBMP-2 patients, no formal comparison could be made.

There is clear evidence from our analysis that rhBMP-2 results in improved rates of fusion, according to Medtronic definitions, compared with ICBG, although the majority of patients receiving either treatment fuse within two years. We found little evidence, however, that successful fusion \textit{per se} led to improvements in pain and function: improvements in pain after surgery were similar among ICBG patients whether or not they had achieved fusion. The fact that rhBMP-2 patients were not blinded to treatment received and fusion status may have biased outcome assessment in favour of rhBMP-2.
We emphasise that this analysis is based on fusion criteria determined by Medtronic. We recognise that a state of fusion is not truly dichotomous, and there will be a degree of subjectivity in its evaluation. The more stringent the criteria used to judge fusion, the more patients will be classified as “non-fused” who might actually be on their way to solid fusion. Given that rhBMP-2 is strongly osteogenic and that over growth of bone can in fact be a problem, the more stringent the criteria used to assess fusion, the more likely it would be to see a difference between rhBMP-2 and bone graft, particularly in the first year.

Meta-analyses of the IPD suggest an increased risk of adverse pain events associated with rhBMP-2 in the immediate post-surgical period. Analyses of the Medtronic data also indicate a possible increase in the risk of cancer associated with rhBMP-2. The overall absolute risk of cancer is however low in both groups.

This review included both the licensed INFUSE® preparation of rhBMP-2 and unlicensed AMPLIFY™ and BCP preparations as it was our intention to review all the relevant trials of rhBMP-2 in spinal surgery. We found no evidence of any difference in efficacy, safety or cancer risk between licensed and unlicensed preparations or rhBMP-2.

Data on adverse events in the wider literature raise concerns that rhBMP-2 increases the risk of heterotopic bone formation, osteolysis, radiculitis and retrograde ejaculation. However, as these observations are from non-randomised studies, they should be interpreted cautiously. Nonetheless, the evidence from these studies for heterotopic bone formation and retrograde ejaculation are compatible with the limited IPD available in the Medtronic sponsored trials.

**Conclusions**

The use of rhBMP-2 in spinal fusion surgery increases the likelihood of successful fusion at up to two years, but this does not seem to translate into clinically significant benefits in pain reduction, function or quality of life. The small improvements in fusion, pain and function, which manifest after six months, also seem to come at the expense of more pain in the
immediate post-operative period, and an increased number of cancer cases. We suggest that it is important that these findings are explained clearly to patients so that they are able to make informed choices about the type of surgery they would prefer.

Acknowledgements

We thank Leah Carron for providing the IPD for the Glassman trial, Kath Wright for performing the literature search and Charlotte Seneschall for assistance in extracting data for the safety evaluation.

References


36. Gray RJ, Rampersaud YR. Comparison of the incidence of radiculitis and radiographic adverse event following minimally invasive lumbar transforaminal interbody fusions (TLIF) with and without the use of bone morphogenetic protein (BMP). Spine J. 2010;10(9 Suppl. 1):8S-9S.


44. Rowan FE, O'Malley N, Poynton A. Recombinant human bone morphogenic protein is safe to use in postolateral spinal fusion surgery and is associated with better leg pain scores at late follow up. J Bone Joint Surg Br. 2011;93 B(Suppl IV):490-1.


47. Pimenta L, Marchi L, Oliveira L, Coutinho E. A prospective, randomized, controlled clinical and radiological study to evaluate and compare the use of silicated calcium phosphate and rh-BMP2 in interbody lumbar spine fusion: 36 month follow-up. Spine J. 2011;11(10 Suppl. 1):130S.
### Table 1: Summary of 12 RCTs providing IPD used in these analyses

<table>
<thead>
<tr>
<th>Trial</th>
<th>Surgical Approach</th>
<th>Number of rhBMP-2 patients</th>
<th>Number of ICBG patients</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT-CAGE Pilot</td>
<td>Single level open/laparoscopic anterior lumbar interbody fusion</td>
<td>11</td>
<td>3</td>
<td>2 years</td>
</tr>
<tr>
<td>LT-CAGE Open</td>
<td>Single level open anterior lumbar interbody fusion</td>
<td>143</td>
<td>136</td>
<td>2 years *</td>
</tr>
<tr>
<td>Bone Dowel Pilot</td>
<td>Single level open anterior lumbar interbody fusion</td>
<td>24</td>
<td>22</td>
<td>4 years</td>
</tr>
<tr>
<td>Bone Dowel Pivotal</td>
<td>Single level open anterior lumbar interbody fusion</td>
<td>55</td>
<td>30</td>
<td>2 years</td>
</tr>
<tr>
<td>Interfix PLIF</td>
<td>Single level posterolateral lumbar interbody fusion</td>
<td>34</td>
<td>33</td>
<td>2 years</td>
</tr>
<tr>
<td>Cornerstone Pilot</td>
<td>One or two level anterior cervical interbody fusion</td>
<td>18</td>
<td>15</td>
<td>2 years</td>
</tr>
<tr>
<td>Mastergraft Pilot</td>
<td>Single level posterolateral lumbar fusion</td>
<td>25</td>
<td>21</td>
<td>4 years</td>
</tr>
<tr>
<td>Interfix ALIF Pilot</td>
<td>Single-level open anterior lumbar interbody fusion</td>
<td>25</td>
<td>20</td>
<td>2 years</td>
</tr>
<tr>
<td>BCP US</td>
<td>Single level posterolateral lumbar fusion</td>
<td>22</td>
<td>5</td>
<td>2 years</td>
</tr>
<tr>
<td>BCP Canada</td>
<td>One or two level posterolateral lumbar fusion</td>
<td>98</td>
<td>99</td>
<td>2 years</td>
</tr>
<tr>
<td>Amplify</td>
<td>Open bilateral posterolateral lumbar fusion</td>
<td>239</td>
<td>224</td>
<td>5 years</td>
</tr>
<tr>
<td>Glassman (non-Medtronic)</td>
<td>Single or multilevel posterolateral lumbar fusion</td>
<td>52</td>
<td>54</td>
<td>2 years</td>
</tr>
</tbody>
</table>

* rhBMP-2 patients were followed up for six years
Table 2: Incidence of adverse events in 11 Medtronic RCTs

<table>
<thead>
<tr>
<th>Nature of adverse event</th>
<th>At all times</th>
<th>At or shortly after surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rhBMP-2 ICBG</td>
<td>rhBMP-2 ICBG</td>
</tr>
<tr>
<td>Total number of patients</td>
<td>672 609</td>
<td>672 609</td>
</tr>
<tr>
<td>Arm and neck (upper extremity) pain</td>
<td>13 12</td>
<td>1 2</td>
</tr>
<tr>
<td>Arthritis or bursitis</td>
<td>24 16</td>
<td>3 1</td>
</tr>
<tr>
<td>Back and leg (lower extremity) pain</td>
<td>268 207</td>
<td>39 24</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>93 99</td>
<td>60 65</td>
</tr>
<tr>
<td>Death</td>
<td>6 10</td>
<td>0 2</td>
</tr>
<tr>
<td>Dural injury</td>
<td>13 12</td>
<td>11 10</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>3 2</td>
<td>1 2</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>117 115</td>
<td>67 73</td>
</tr>
<tr>
<td>Graft related</td>
<td>--- 45</td>
<td>--- 17</td>
</tr>
<tr>
<td>Heterotopic bone formation *</td>
<td>24 4</td>
<td>--- ---</td>
</tr>
<tr>
<td>Implant</td>
<td>22 10</td>
<td>10 6</td>
</tr>
<tr>
<td>Infection</td>
<td>104 90</td>
<td>53 47</td>
</tr>
<tr>
<td>Neurological</td>
<td>152 119</td>
<td>29 20</td>
</tr>
<tr>
<td>Osteolysis **</td>
<td>12 ---</td>
<td>--- ---</td>
</tr>
<tr>
<td>Other pain</td>
<td>97 63</td>
<td>12 9</td>
</tr>
<tr>
<td>Respiratory</td>
<td>34 27</td>
<td>19 16</td>
</tr>
<tr>
<td>Retrograde ejaculation</td>
<td>5 1</td>
<td>3 1</td>
</tr>
<tr>
<td>Spinal</td>
<td>70 63</td>
<td>4 7</td>
</tr>
<tr>
<td>Trauma</td>
<td>153 137</td>
<td>8 8</td>
</tr>
<tr>
<td>Urogenital</td>
<td>81 63</td>
<td>42 35</td>
</tr>
<tr>
<td>Vascular</td>
<td>6 4</td>
<td>6 4</td>
</tr>
<tr>
<td>Vertebral fracture</td>
<td>3 1</td>
<td>3 0</td>
</tr>
<tr>
<td>Wound complication</td>
<td>6 4</td>
<td>4 3</td>
</tr>
</tbody>
</table>

* Only reported for Interfix PLIF trial as additional data beyond the protocol

** Only reported in rhBMP-2 arms of Bone Dowel trials as additional data beyond the protocol
Table 3: Incidence of cancer in the Medtronic RCTs.

<table>
<thead>
<tr>
<th>Cancer</th>
<th>rh-BMP-2</th>
<th>ICBG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skin cancers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanoma</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Basal cell carcinoma</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Breast</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Colon</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Larynx</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Leukaemia</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lung</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Non-Hodgkin’s lymphoma</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ovarian</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Pancreatic</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Prostate</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Stomach</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Testicular</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Thyroid</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total cancers</strong></td>
<td><strong>20</strong></td>
<td><strong>8</strong></td>
</tr>
<tr>
<td><strong>Total patients</strong></td>
<td><strong>672</strong></td>
<td><strong>609</strong></td>
</tr>
</tbody>
</table>
Figure legends

Figure 1: Meta-analyses of pain and function outcomes at 6 weeks, 3, 6, 12 and 24 months after surgery.

Figure 2: Meta-analyses of successful spinal fusion at 6, 12 and 24 months after surgery.

Figure 3: Forest plot of relative risk of successful fusion 2 years after surgery.

Figure 4: Subgroup analyses by surgical approach for ODI and successful fusion.

Figure 5: Mean improvement in ODI, SF-36 and back pain according to treatment (rhBMP-2 v ICBG) and fusion status at 6, 12 and 24 months after surgery.

Figure 6: Meta-analysis of adverse events by category at or shortly after time of surgery in 11 Medtronic RCTs

Figure 7: Meta-analyses of adverse events according to adverse event category.

Figure 8: Forest plot of cancer incidence in the Medtronic RCTs.

* Back pain and leg pain are 20-point scales, Oswestry (ODI) and SF-36 PCS are percentage scales (100 points)

Figure 9
Figure 10
Figure 11

### Estimates with 95% confidence intervals

<table>
<thead>
<tr>
<th>Trial</th>
<th>mBMP Fused/Total</th>
<th>ICBG Fused/Total</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT Cage Open</td>
<td>123/130</td>
<td>105/119</td>
<td>1.96 (0.99, 1.16)</td>
</tr>
<tr>
<td>Bone Dowel Pilot</td>
<td>24/34</td>
<td>13/19</td>
<td>1.44 (1.00, 1.99)</td>
</tr>
<tr>
<td>Bone Dowel Packed</td>
<td>43/44</td>
<td>23/27</td>
<td>1.20 (1.00, 1.44)</td>
</tr>
<tr>
<td>Interfix ALIF Pilot</td>
<td>16/20</td>
<td>13/14</td>
<td>0.86 (0.66, 1.12)</td>
</tr>
<tr>
<td>Comerstone Pilot</td>
<td>11/11</td>
<td>11/11</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Interfix PLIF</td>
<td>24/26</td>
<td>21/27</td>
<td>1.19 (0.94, 1.49)</td>
</tr>
<tr>
<td>M itself Pilot</td>
<td>16/19</td>
<td>14/20</td>
<td>1.36 (1.00, 1.84)</td>
</tr>
<tr>
<td>BCP US</td>
<td>17/21</td>
<td>3/3</td>
<td>0.81 (0.66, 1.00)</td>
</tr>
<tr>
<td>BCP Canada</td>
<td>86/89</td>
<td>64/91</td>
<td>1.37 (1.20, 1.56)</td>
</tr>
<tr>
<td>Amplify</td>
<td>188/194</td>
<td>151/169</td>
<td>1.97 (1.01, 1.14)</td>
</tr>
</tbody>
</table>

**Relative risk of fusion (mBMP-2 vs ICBG)**

ICBG better → mBMP-2 better
Oswestry Disability Index

Successfu fusion

Figure 12
Figure 13
Estimates with 95% confidence intervals

<table>
<thead>
<tr>
<th>Type of adverse event</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arm and neck pain</td>
<td>0.65 (0.06, 7.24)</td>
</tr>
<tr>
<td>Arthritis/bursitis</td>
<td>2.84 (0.29, 27.51)</td>
</tr>
<tr>
<td>Back and leg pain</td>
<td>1.92 (1.13, 3.25)</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>0.99 (0.68, 1.44)</td>
</tr>
<tr>
<td>Dural injury</td>
<td>1.49 (0.62, 3.62)</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>0.30 (0.02, 3.59)</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>1.04 (0.73, 1.48)</td>
</tr>
<tr>
<td>Implant</td>
<td>2.11 (0.73, 6.07)</td>
</tr>
<tr>
<td>Infection</td>
<td>1.23 (0.82, 1.85)</td>
</tr>
<tr>
<td>Neurological</td>
<td>1.73 (0.96, 3.12)</td>
</tr>
<tr>
<td>Other pain</td>
<td>1.68 (0.69, 4.07)</td>
</tr>
<tr>
<td>Respiratory</td>
<td>1.27 (0.64, 2.51)</td>
</tr>
<tr>
<td>Retrograde ejaculation</td>
<td>3.00 (0.31, 29.15)</td>
</tr>
<tr>
<td>Spinal</td>
<td>0.53 (0.15, 1.85)</td>
</tr>
<tr>
<td>Trauma</td>
<td>0.93 (0.34, 2.55)</td>
</tr>
<tr>
<td>Urogenital</td>
<td>1.38 (0.86, 2.12)</td>
</tr>
<tr>
<td>Vascular</td>
<td>1.94 (0.53, 7.18)</td>
</tr>
<tr>
<td>Wound complication</td>
<td>1.76 (0.39, 8.06)</td>
</tr>
</tbody>
</table>

Odds ratio of adverse events (rhBMP-2 vs ICBG)

More common with ICBG ← → More common with rhBMP-2

Figure 14
Figure 15
Figure 16

Estimates with 95% confidence intervals

<table>
<thead>
<tr>
<th>Total</th>
<th>rhBMP Cancers/ Total</th>
<th>ICBG Cancers/ Total</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT Cage-Open</td>
<td>2/143</td>
<td>1/137</td>
<td>1.92 (0.18, 20.89)</td>
</tr>
<tr>
<td>Bone Dowel Pilot</td>
<td>1/24</td>
<td>0/22</td>
<td>2.76 (0.12, 64.23)</td>
</tr>
<tr>
<td>Bone Dowel Pictoral</td>
<td>1/66</td>
<td>0/00</td>
<td>1.65 (0.07, 39.25)</td>
</tr>
<tr>
<td>BCP Canada</td>
<td>1/58</td>
<td>2/99</td>
<td>0.51 (0.05, 5.45)</td>
</tr>
<tr>
<td>Amplify</td>
<td>1/2048</td>
<td>5/2035</td>
<td>2.25 (0.81, 6.29)</td>
</tr>
<tr>
<td>Pooled</td>
<td></td>
<td></td>
<td>1.84 (0.81, 4.16)</td>
</tr>
</tbody>
</table>

Relative risk of cancer (rhBMP-2 vs ICBG)
Appendix tables and figures

Appendix figure 1: Reduction in pain scores from pre-operative results by treatment received.
Appendix figure 2: Mean difference in operating time between rhBMP-2 and ICBG in the Medtronic trials
Appendix figure 3: Meta-analysis of use of four types of pain-relief medication in the Medtronic trials
### Appendix table 1: Details of published studies reporting on the safety of rhBMP-2 surgery

<table>
<thead>
<tr>
<th>Comparator</th>
<th>Surgical approach</th>
<th>Author</th>
<th>rhBMP-2 No. Patients</th>
<th>Comparator</th>
<th>Comparator No. Patients</th>
<th>Adverse events reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICBG</td>
<td>ACDF/PCF</td>
<td>Buttermann (2008) (22)</td>
<td>30</td>
<td>ICBG</td>
<td>36</td>
<td>Wound complications</td>
</tr>
<tr>
<td>ICBG</td>
<td>ACDF/PCF</td>
<td>Crawford (2009) (23)</td>
<td>41</td>
<td>ICBG</td>
<td>36</td>
<td>Wound complications</td>
</tr>
<tr>
<td>ICBG</td>
<td>ALIF/ALF</td>
<td>Carragee (2012) (19)</td>
<td>239</td>
<td>ICBG</td>
<td>233</td>
<td>Retrograde ejaculation, infection, urinary retention</td>
</tr>
<tr>
<td>ICBG</td>
<td>ALIF/ALF</td>
<td>Pradhan (2006) (24)</td>
<td>9</td>
<td>ICBG</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>ICBG</td>
<td>PLIF/PLF/TLIF</td>
<td>Glassman (2007) (26)</td>
<td>91</td>
<td>ICBG</td>
<td>35</td>
<td>Hardware failure</td>
</tr>
<tr>
<td>ICBG</td>
<td>PLIF/PLF/TLIF</td>
<td>Lee (2010) (27)</td>
<td>86</td>
<td>ICBG</td>
<td>41</td>
<td>Infection, wound complications</td>
</tr>
<tr>
<td>ICBG</td>
<td>PLIF/PLF/TLIF</td>
<td>Mannion (2010) (28)</td>
<td>17</td>
<td>ICBG</td>
<td>19</td>
<td>Heterotopic bone formation</td>
</tr>
<tr>
<td>ICBG</td>
<td>PLIF/PLF/TLIF</td>
<td>Rihn (2009) (30)</td>
<td>86</td>
<td>ICBG</td>
<td>33</td>
<td>Heterotopic bone formation, osteolysis, infection, hardware failure, wound complication, leg pain/radiculitis, inflammatory cyst formation</td>
</tr>
<tr>
<td>ICBG</td>
<td>PLIF/PLF/TLIF</td>
<td>Singh (2006) (31)</td>
<td>41</td>
<td>ICBG</td>
<td>11</td>
<td>Heterotopic bone formation</td>
</tr>
<tr>
<td>Allograft</td>
<td>ACDF/PCF</td>
<td>Lu (2010) (32)</td>
<td>100</td>
<td>ICBG</td>
<td>50</td>
<td>Dysphagia</td>
</tr>
<tr>
<td>Allograft</td>
<td>ALIF/ALF</td>
<td>Slosar (2007) (33)</td>
<td>45</td>
<td>ICBG</td>
<td>30</td>
<td>Heterotopic bone formation, wound complications</td>
</tr>
<tr>
<td>Allograft</td>
<td>Various</td>
<td>Vaidya (2007b) (34)</td>
<td>36</td>
<td>ICBG</td>
<td>41</td>
<td>Hardware failure, dysphagia</td>
</tr>
<tr>
<td>Allograft</td>
<td>/ autograft</td>
<td>Mindea (2009) (35)</td>
<td>35</td>
<td>ICBG</td>
<td>8</td>
<td>Leg pain/radiculitis</td>
</tr>
<tr>
<td>Autologous bone</td>
<td>PLIF/PLF/TLIF</td>
<td>Gray (2010) (36)</td>
<td>82</td>
<td>ICBG</td>
<td>39</td>
<td>Heterotopic bone formation, osteolysis, hardware failure, leg pain/radiculitis</td>
</tr>
<tr>
<td>Autologous bone</td>
<td>PLIF/PLF/TLIF</td>
<td>Joseph (2007) (37)</td>
<td></td>
<td>ICBG</td>
<td></td>
<td>Heterotopic bone formation</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>ACDF/PCF</td>
<td>Longley (2009) (38)</td>
<td>84</td>
<td>ICBG</td>
<td>84</td>
<td>Dysphagia</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>ACDF/PCF</td>
<td>Smucker (2006) (39)</td>
<td>69</td>
<td>ICBG</td>
<td>165</td>
<td>Dysphagia</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>ACDF/PCF</td>
<td>Vaidya (2007a) (40)</td>
<td>22</td>
<td>ICBG</td>
<td>24</td>
<td>Dysphagia</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>ACDF/PCF</td>
<td>Williams (2011) (41)</td>
<td>4532</td>
<td>ICBG</td>
<td>652</td>
<td>Dysphagia, wound complications</td>
</tr>
<tr>
<td>Treatment Type</td>
<td>Procedure</td>
<td>Study Reference</td>
<td>Cases</td>
<td>Patients</td>
<td>Complications</td>
<td></td>
</tr>
<tr>
<td>------------------------------------</td>
<td>------------</td>
<td>----------------------------------</td>
<td>-------</td>
<td>----------</td>
<td>-------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>ACDF/PCF</td>
<td>Xu (2011) (42)</td>
<td>48</td>
<td>156</td>
<td>Heterotopic bone formation, infection, hardware failure, dysphagia, recurrent laryngeal neck palsy, wound complications</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>ACDF/PCF</td>
<td>Yaremchuk (2010) (43)</td>
<td>260</td>
<td>515</td>
<td>Dysphagia</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>PLIF/PLF/TLIF</td>
<td>Rowan (2011) (44)</td>
<td>64</td>
<td>40</td>
<td>Leg pain/radiculitis</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>PLIF/PLF/TLIF</td>
<td>Williams (2011) (41)</td>
<td>6049</td>
<td>18267</td>
<td>Wound complications, inflammatory cyst formation</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Various</td>
<td>Cahill (2009) (2)</td>
<td>17626</td>
<td>53026</td>
<td>Dysphagia, wound complications</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Various</td>
<td>Williams (2011) (41)</td>
<td>11933</td>
<td>43929</td>
<td>Wound complications</td>
<td></td>
</tr>
<tr>
<td>Bone marrow aspirate</td>
<td>ALIF/ALF</td>
<td>Gerszten (2011) (45)</td>
<td>45</td>
<td>54</td>
<td>Heterotopic bone formation, neurologic events, hardware failure, wound complications, vascular events</td>
<td></td>
</tr>
<tr>
<td>Bone marrow aspirate</td>
<td>PLIF/PLF/TLIF</td>
<td>Taghavi (2010) (46)</td>
<td>24</td>
<td>38</td>
<td>Hardware failure</td>
<td></td>
</tr>
<tr>
<td>Silicated Calcium Phosphate</td>
<td>PLIF/PLF/TLIF</td>
<td>Pimenta (2011) (47)</td>
<td>15</td>
<td>15</td>
<td>Heterotopic bone formation</td>
<td></td>
</tr>
</tbody>
</table>
Appendix figure 4: Heterotopic bone formation in five non-randomised studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Comparator</th>
<th>Surgery</th>
<th>rhBMP-2 events</th>
<th>Comparator events</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rihn</td>
<td>ICBG</td>
<td>TLIF</td>
<td>2/98</td>
<td>0/33</td>
<td>1.94 (0.10, 39.29)</td>
</tr>
<tr>
<td>Mannion</td>
<td>ICBG</td>
<td>LIF</td>
<td>2/17</td>
<td>0/19</td>
<td>5.57 (0.29, 100.27)</td>
</tr>
<tr>
<td>Gray</td>
<td>Autologous bone</td>
<td>PLIF</td>
<td>43/82</td>
<td>10/39</td>
<td>2.05 (1.15, 3.63)</td>
</tr>
<tr>
<td>Joseph</td>
<td>Autologous bone</td>
<td>PLIF/TLIF</td>
<td>50/38</td>
<td>1/10</td>
<td>2.17 (0.29, 16.30)</td>
</tr>
<tr>
<td>Pimensa</td>
<td>SCP</td>
<td>LIF</td>
<td>1/15</td>
<td>0/15</td>
<td>3.00 (0.13, 66.09)</td>
</tr>
</tbody>
</table>

Relative risk of heterotopic bone formation

More common with comparator ← → More common with rhBMP-2
Appendix figure 5: Leg pain or radiculitis in four non-randomised studies
Appendix figure 6: Dysphagia in six non-randomised studies of cervical surgery
Appendix material: call for evidence

Systematic review of bone morphogenic protein-2 (rhBMP-2) for spinal fusion - call for evidence

The Centre for Reviews and Dissemination (CRD) is undertaking a systematic review and individual participant data (IPD) meta-analysis of the comparative effectiveness of rhBMP-2 (marketed as INFUSE) for spinal fusion. The review has been commissioned by the Yale University Open Data Access (YODA) initiative as part of an overarching project to systematically review the safety and effectiveness of rhBMP-2, including re-analysis of IPD that have been made available to Yale on an unrestricted basis by the manufacturer (Medtronic Inc). YODA aims to improve access to patient-level data from clinical trials, and provide independent, scientifically rigorous, objective and fair analyses of such data.

CRD will undertake a comprehensive and rigorous systematic review and meta-analysis of individual participant data (IPD) of all relevant randomised controlled trials that have compared rhBMP-2 with standard bone graft therapy.

We will include all relevant randomised controlled trials, irrespective of whether conducted by the manufacturer or not, and irrespective of whether published or not.

We are therefore interested in hearing from anyone who has conducted, or is aware of, unpublished or partially published research in this area. For example, trials which have been presented at conferences but not fully reported elsewhere.

We are currently aware of 17 trials funded by the manufacturer and have searched the published literature but welcome any information regarding further unpublished research. If you know of any such trials please contact [CRD details deleted].

Link to CRD project page
http://www.york.ac.uk/inst/crd/projects_in_progress.cfm
Link to YODA page
http://medicine.yale.edu/core/projects/yodap/index.aspx
3. **Annals of Internal Medicine: External Peer Reviewer Comments (October 2012)**

Annals of Internal Medicine Mary Beth Schaeffer, Managing Editor, 215-351-2629
190 N. Independence Mall West, Philadelphia, PA 19106

M12-2603
Reviewer 1
Due: 11/12/12

Please review this manuscript as a Systematic Review. Please complete the following:

1) Summary Grades (below), which we use as a rough guide to your opinions and which are confidential.

2) Comments to the Editors, which are confidential and are an opportunity for you to tell us candidly what you think are the major strengths and weaknesses and your thoughts on suitability for publication.

3) Comments for the Authors, which are sent to the authors with our letter. Do not indicate on this section whether or not you believe the manuscript should be published.

4) Will reviewing this manuscript represent a conflict of interest for you (financial or otherwise)? Yes____ No____

Thank you for your help.

**Summary Grades (Circle one response for each question)**

<table>
<thead>
<tr>
<th>Relevance: 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accuracy: 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inaccurate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Completeness: 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information missing</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Appeal: 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Educational Value 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Overall Judgment (circle one):**

Reject Reconsider after major revision
Accept after satisfactory revision  Accept

Confidential Comments for the Editors:

Indicate briefly your opinion of the most important strengths and weaknesses of the manuscript and which of these were most important to you in making your recommendation.

Fairly carefully conducted person level meta-analysis of very important topic. Interpretation of data is relatively sound, but could be better
M12-2603 Reviewer 1
Comments for the Author:

Please give specific comments, as appropriate on importance of the research question, scientific strength, clarity of presentation and appropriateness for Annals readers. Include comments on style and presentation, length of text, and number, quality and importance of tables and figures. Group your main points under the heading Major Comments. Give further suggestions under the heading Additional Comments.

Major Comments:
1) Page 5- bias assessment method is noted here but not provided in results
2) Page 13- Need citation for rhBMP-2 and cancer risk with more discussion on biologic mechanisms. Unclear what last sentence in second paragraph means?.
3) Were the results sensitive to including the Medtronics trial which did not include ICBG as comparator? Where other sensitivity analyses performed?
4) Cancer conclusion is too strong given lack of statistical significance. Agree trend is concerning. But need to quality
5) Mention increased fusion rate in first paragraph of discussion.
6) Figure 6 does not contribute very much. Are there a priori hypotheses for these associations? Some are likely fully unrelated to BMP or ICBG

Additional Comments::
1) Figures are mis-numbered and hard to follow.
2) Many odd acronyms used, need to reduce.
3) Need to describe direction of change of pain scores and functional measures and what these scales mean
4) Introduction aim to provide “robust and fair assessment” -- tone down, let the reader decide.
5) Table 1- needs citation to trials references
6) Table 2- needs to better define “at or shortly after surgery” in table
M12-2603
Reviewer 2
Due: 11/12/12

Please review this manuscript as a Systematic Review. Please complete the following:

5) Summary Grades (below), which we use as a rough guide to your opinions and which are confidential.

6) Comments to the Editors, which are confidential and are an opportunity for you to tell us candidly what you think are the major strengths and weaknesses and your thoughts on suitability for publication.

7) Comments for the Authors, which are sent to the authors with our letter. Do not indicate on this section whether or not you believe the manuscript should be published.

8) Will reviewing this manuscript represent a conflict of interest for you (financial or otherwise)? No

Thank you for your help.

Summary Grades (Circle one response for each question)

Relevance:     1  2  3  4  5
               Low  Great

Accuracy:      1  2  3  4  5
               Inaccurate  Accurate

Completeness:  1  2  3  4  5
               Information missing  Complete

Appeal:        1  2  3  4  5
               Limited  Wide

Educational Value:  1  2  3  4  5
                   Low  High

Overall Judgment (circle one):

Reject          Reconsider after major revision
Accept after satisfactory revision Accept

Confidential Comments for the Editors:

Indicate briefly your opinion of the most important strengths and weaknesses of the manuscript and which of these were most important to you in making your recommendation.
1. IPD makes the message strong; conclusions will resonate with many readers
2. For the journal to decide: will this topic truly have interest among your readership? It will get searched online and get cited but that is somewhat different than interest for your core readership.
3. The methodology, though I suspect is excellent, is not transparent—even after looking up the CRD entry—and that makes it not only difficult to assess but, honestly, provides a feeling of disappointment that a sophisticated group such as this thought they could get away with it...
4. At the end, this is much more a meta-analysis than it is a systematic review; because of this idiosyncrasy, the authors need to especially clear if they choose to revise this and you choose to continue to pursue it
Reviewer 2  M12-2603

Comments for the Author:

Please give specific comments, as appropriate on importance of the research question, scientific strength, clarity of presentation and appropriateness for Annals readers. Include comments on style and presentation, length of text, and number, quality and importance of tables and figures. Group your main points under the heading Major Comments. Give further suggestions under the heading Additional Comments.

This manuscript reports the findings of an individual patient data (IPD) meta-analysis (MA)—primarily from the Yale Open Data Access (YODA) depository of Medtronic trial data—of known trials supplemented with a systematic review (SR) of adverse events regarding the surgical arthrodesis of the spine utilizing either iliac crest bone graft (ICBG) or recombinant human Bone Morphogenetic Protein 2 (rhBMP2). Authors report the findings of their MA and SR with the conclusions that rhBMP2 increases fusion rates but of questionable clinical significance, increases short term post surgical pain and the number of cancer cases.

The authors represent investigators from several academic centers. Their academic credentials are not specifically known to the reviewer nor is their prior knowledge of orthopaedic spine or neurosurgical procedures. Funding was received from the YODA project.

Overall, the issues and questions that the authors raise are timely and of topical interest to a broad spectrum of healthcare providers and consumers dealing with spine related pain and dysfunction. Authors develop the background on the topic and a compelling reason for their study including the current knowledge gap regarding the balance between improved outcomes and adverse events and importantly, why this is important for patients. They rationally propose a method of using MA of IPD from RCTs supplemented with additional data collected from SR to inform and help close that knowledge gap. Their methods appear appropriate and their conclusions are consistent with the results they present.

However, several major issues should be addressed to improve the manuscript:

1. Because this is not a simple standard linear progression of SR leading to data extracted leading to MA, the overall description of the study must be clear. It is not clear whether a SR resulted in 11 Medtronic studies + 1 non Medtronic study or if these were the a priori RCTs already known to exist. Without this knowledge and without knowing what studies were excluded (if a SR was indeed performed), the reader cannot tell if these are the best data (because they are complete IPD) or the most comprehensive.

2. Related to item 1, as a MA/SR, the overall methods must be clear and transparent such that the study could be independently replicated like an experiment. An examination of the CDR entry in PROSPERO did not clarify the search strategy sufficiently for this to occur. Inclusion and exclusion criteria should be explicitly stated. In fact, all search details should be included along with a flow diagram. In addition, was a manual bibliography search performed. SRs reported in the orthopaedic literature—to which this study pertains—indicate several instances where relevant literature was found in bibliographic searches.

3. Although the recently reported adverse events are of public concern—in fact, I too am concerned and biased toward the same conclusions—the authors are encouraged to take a more neutral scientific stance for this MA/SR. Or to provide more compelling reasons from the literature that make it reasonable to start with a skewed stance. This bias is also subtly present in the study design where efficacy is meta-analyzed using IPD but the adverse events are not.
Minor issues:

1. In the methods, please report the kappa agreement between the 2 reviewers in reviewing and agreeing on inclusion of studies.
2. In the methods, please describe the definition of fusion according the Medtronic and comment on its appropriateness.
3. In the methods, please clarify the use of linear versus logistic regression: eg why not logistic regression for fusion versus pain and function outcomes?
4. In the results, please utilize an accepted scoring system for study quality.
5. In the results, please report data in the text consistently throughout for clarity and readability; ie p values (instead of “statistically significant” and RR with CI when relevant)
6. In the results and discussion, it should be made clear that adverse events including rates of cancer are associations and are not derived from the same quality of data (eg IPD) as the MA of outcomes.
4. **Annals of Internal Medicine: Statistical Editor Comments (November 2012)**

MS 12-2602: “The safety and efficacy of recombinant human bone morphogenetic protein-2 (rhBMP-2) for spinal fusion – An individual patient data meta-analysis.”

**CORRESPONDING AUTHOR:** Dr. Lesley A. Stewart, University of York, UK

**STATISTICAL EDITOR’S COMMENTS**

**General comments**

1. The present manuscript describes an individual patient meta-analysis of randomized clinical trials comparing recombinant human bone morphogenic protein (rhBMP-2) vs iliac crest bone graft in spinal fusion surgery. The authors concluded that rhBMP-2 increased fusion rates but did not result in clinically relevant improvements in pain and it could result in serious adverse events. This paper was discussed by a team of senior editors and statistical editors at *Annals of Internal Medicine*, where the following methodological issues came up:

2. **Literature search.** Please provide a flow diagram of the study search process, indicating the number of studies that were included and excluded along the process and the final number of studies included in the individual patient meta-analysis and in the additional meta-analysis with published studies, separately for efficacy and safety endpoints. Please also complete a search of RCT registries to identify potentially completed but unpublished studies.

3. **Presentation of which studies contribute to different endpoints.** Throughout the Results section, it was difficult to follow all the different endpoints that were being evaluated as well as which studies contributed to each endpoint. Please provide a table or a figure (either in the main text or in an appendix) with a grid showing which studies contribute to each endpoint (including efficacy and safety endpoints). For efficacy endpoints, please detail availability by time period (6 weeks, 3, 6, 12 and 24 months). In addition to this Table, at least for the primary efficacy endpoints that are based on scale measurements, please provide some guidance (either in the Methods section or in the caption of for the reader as to what is the range of the scale and what are considered clinically meaningful differences.

4. **Description of statistical methods.** In general, we found the description of statistical methods insufficient for an informed reader to reproduce the analysis. For both efficacy and safety outcomes, please provide detail methods with respect to each of the following four aspects:
   - Analysis of individual studies with individual patient data.
   - Combination of studies with individual patient data.
• Combination of studies with individual patient data with studies with data available only from study reports (including more detailed description of the approach to incorporating non-randomized studies for safety endpoints).
• Sensitivity analyses.

Each of these items does not necessarily need a separate section, but should be clearly described in a structured way (please consider preparing a detailed Statistical Appendix with a more detailed description than that provided in the main text). With respect to the statistical methods, please:
• Specify exactly what is meant by one-stage and two-stage methods.
• When describing mixed models, please provide more detailed descriptions of the methods: what were the fixed and the random effects in the models, where there random intercepts only or also random coefficients, what were the estimation methods (for both linear and logistic mixed models, although this is particularly important for mixed logistic models)?
• Please also provide a detailed description of the analytical methods for each individual study and use methods consistent with appropriate modern standards. For instance, for continuous outcomes it seems that the authors simply calculated the difference in outcomes at each point in time with respect to baseline, while we would expect a more efficient use of the data from each individual study by fitting a mixed model for longitudinal data from each study (for both continuous and dichotomous endpoints, although alternative survival-type models may be considered for longitudinal dichotomous endpoints – see below). Notice that we also expect that the authors consider the elements of each individual study with the same degree of granularity as we would expect in the report of an individual study. This includes, for instance, indicating the extent and analytical considerations of missing data and losses to follow-up in longitudinal models (what was the approach used in the analysis?), indicating the approach taken about study centers in each individual study, specifying if the estimates calculated are marginal or population-averaged, specifying the approach to subgroup analyses (notice that with the current description, it is unclear how the subgroup analyses were actually done and tested for), etc.
• For the endpoint of spinal fusion, the investigators used relative risks at different points in time, but it is unclear if this is the most appropriate model given that the endpoint is theoretically permanent and the likely occurrence of losses to follow-up. It seems that a survival-type model that takes into account the fact that fusion occurs between two visits (and not exactly at each visit) seems much more appropriate than the model considered in the manuscript. Please base the analysis for this endpoint in a survival-type model or provide a solid argument justifying the current approach.
• For most of the analyses presented in the manuscript, the investigators had the option of providing one step or two step models. For the analysis of efficacy endpoints, the investigators selected two stage models. As sensitivity analyses, please use one-step models for efficacy endpoints.
5. **Presentation of Results:** Please consider the following issues:
   - We considered the presentation of the characteristics of individual studies insufficient to provide enough context for the readers of the paper. Please provide more detailed evidence tables with key characteristics of individual studies (for instance, it was unclear if we were dealing with similar studies in terms of patient characteristics and pre-intervention pain, or if we were dealing with very different types of studies). In addition, since this is an individual patient meta-analysis, the investigators have the opportunity of presenting much more detailed descriptive and analytical data for each individual study. Please take advantage of this when preparing evidence and descriptive tables of participant characteristics.
   - With respect to the Figures, it seems that the numeration of the Figures did not correspond to the description in the text. In the revised version, please check the figures and tables and consider limiting the number of figures in the main text and making more effective use of figures. Given the lack of synchrony between text and figures, we found it difficult at this time to provide detailed guidance on each figure, but we ask the investigators that they carefully consider the role of each figure in the revised manuscript and provide informative and carefully designed figures to support the Results section.

6. **Relationship between successful fusion and improvement in pain and function scores.** In the paper, the investigators present the impact of fusion (a post randomization event) on improvement in pain and fusion. While this type of mediation analysis is interesting, it is subject to potentially complex biases and does not add to the main aims of the paper. The authors considered that this analysis is best presented in a different paper.

7. **Discussion.** Overall, the editors found that the discussion was a bit too mechanical and did not integrate study characteristics and biological mechanisms with the findings in the analysis. In the revision of the paper, please provide a more integrated Discussion of the findings.
5. **Annals of Internal Medicine: Comments on Format Requirements**

**Annals of Internal Medicine**

**Revision Requirements**

<table>
<thead>
<tr>
<th>Manuscript No.:</th>
<th>M12-2603</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Author:</td>
<td>Simmonds</td>
</tr>
</tbody>
</table>

**Abstract**

- Please add the following section(s) to your abstract: Primary Funding Source. Be sure to provide information for all sections and do not combine any sections.
- Please do not exceed the 275-word limit when adding this section.
- If the work that you present in this manuscript is registered in the PROSPERO systematic review registry (see [http://www.crd.york.ac.uk/prospero](http://www.crd.york.ac.uk/prospero)), please add the registration number in parentheses at the end of the Primary Funding Source section of your abstract.

**Text**

- Please do not use (write out) the following abbreviations: ODI (in abstract), ICBG (in abstract), NDI, MD, BCP.
- Please provide academic degrees for all authors.
- On a separate page after the text, please include the following information for publication: acknowledgments, grant support (include specific grant numbers when available), address for reprint requests, and current mailing addresses for all authors. Please note that letters of permission must be provided for all persons acknowledged by name.
- Please use generic names in place of brand names where applicable.
- Please clarify the word "significant" throughout. If the term does not refer to statistical significance, please substitute it with a more appropriate term/phrase (e.g., "clinically significant," "substantial," or "important").
- If your study was funded, please add a paragraph at the end of the Methods section specifying the role of the funding source(s), if any, in the design, conduct, and analysis of your study and in the decision to submit the manuscript for publication. If your study was not funded, please state this at the end of the Methods section.
- Please report $P$ values as such: For $P$ values between 0.001 and 0.20, please report the value to the nearest thousandth. For $P$ values greater than 0.20, please report the value to the nearest hundredth. For $P$ values less than 0.001, report as "$P < 0.001."'
- If the work that you present in this manuscript is registered in the PROSPERO systematic review registry (see [http://www.crd.york.ac.uk/prospero](http://www.crd.york.ac.uk/prospero)), please note this (including the registration number) in the overview of your Methods section.
• As you revise your paper, please keep the total number of tables and figures intended for print publication to roughly 1 per 800 words of text; the average original research article can accommodate 4 to 5 graphics in total. Further tables and figures can be slated for online-only publication. Furthermore, please follow any advice recommended or requested from the Deputy Editor.

**Tables**
• Define all abbreviations in footnotes to the tables.
• Remove all vertical lines and all internal rule lines from your tables. Retain only horizontal lines beneath the title, beneath the column headings, and at the end of the tables.
• If you have not already done so, please prepare tables by using the Tables feature in Word. Data should appear in individual cells. Please do not use tabs.

**Figures**
• In the figure legends, define all abbreviations used in the figures.
• Please do not use 3-dimensional images in figures, and limit the use of shading.

**Permissions**
• Provide letters of permission from everyone listed in the acknowledgment section.

**Electronic Manuscript Submission**
• Please submit an electronic file for the manuscript, a separate file containing all tables, and individual JPEG files for each figure. For further instructions on how to convert files created in Microsoft Office into JPEGs, please e-mail the contact person below.
6.  *Annals of Internal Medicine*: Editor’s Letter Requesting Revision

November 27, 2012

Mark C. Simmonds Dr, MA PhD
Centre for Reviews and Dissemination
University of York
York, YO10 5DD
GBR

REF: M12-2603

Dear Dr. Simmonds:

Thank you for submitting to *Annals of Internal Medicine* your manuscript, “The safety and efficacy of recombinant bone morphogenetic protein-2 (rhBMP-2) for spinal fusion.” It was reviewed by our physician editors, statistical editors, and two external reviewers. While we did not find the current version suitable for publication in *Annals*, we would be interested in publishing a revision that satisfactorily addresses the issues raised in this letter and the accompanying reviews.

Please provide a letter with the revised manuscript that describes how you responded to each of the concerns raised in this letter, in the attached statistical review and two external reviews. We would like to receive your revisions within 4 weeks. Please let us know as soon as possible if you do not think you can meet this target date for resubmission. Also, it would streamline the review process if you could provide us with a copy of the full technical report that served as the basis for this review.

1. In reporting the findings of your systematic review and meta-analysis of the trials and of the observational studies, take care to follow the PRISMA and MOOSE recommendations, respectively ([www.equator-network.org](http://www.equator-network.org)).

2. While you have the great advantage of having individual patient-level data, many of your analyses reflect conventional meta-analysis. Similarly, you do not take full advantage of the longitudinal data that you have for many outcomes. In addition, the majority of the results section is devoted to the presentation of subset analyses while falling short of a clear, complete evaluation of the findings of the primary analyses. Refer to the statistical editor’s comments for further comment on these issues.

3. One of our major concerns relates to a lack of clarity about the extent of the patient-level data that you had available to you for these analyses. How does the data you have overlap with the data Carragee and colleagues report in reference 6? Given that determining the safety of rhBMP-2 was a major objective of this review, it is of utmost
importance that you include all available data on adverse events including data that was reported to the FDA even if it was not reported in any of the included trial or observational study reports. If you do not currently include such data, it is imperative that you do so.

4. Following on the above, you state that, “For evaluation of safety, we examined numbers of adverse events, including cancer, but were limited by classifications of adverse events produced by Medtronic.” Please provide more detail about these classifications so that the reader can better understand how they might influence your findings. Also, our understanding is that Medtronic was required to provide patient level data. As such, it seems that you should have the ability to look at the primary adverse event reports and reclassify adverse events in a matter that was less limiting.

5. The manuscript lacks clarity about how the data that Medtronic provided overlapped with the studies you identified in your own literature search. Please, describe the overlap more clearly.

4. Please include a figure that summarizes your search and selection process. Did you search clinical trials registries as well as publication databases when trying to identify all relevant studies?

5. The description of included outcomes lacks sufficient detail and clarity.

• Please clearly identify in both the abstract and the text the outcomes you sought to examine including the time points for these outcomes.

• What methods did included studies use to measure pain?

• In some places in the manuscript you refer to pain and disability/function as separate outcomes, but in other places you seem to be referring to them as a composite outcome, making the findings about effectiveness of rhBMP-2 difficult to follow.

• It also appears that you treat pain as both an efficacy outcome and an adverse event, which we find problematic.

• On page 5 you state, “Data on these outcomes were analyzed at 6 weeks, 3, 6, 12, and 24 months after surgery.” Did all included studies measure all outcomes and measure them at all of these time points? If not, the results must include details about which studies measured what and when. Missing outcome data is another issue on which you need to report greater detail.

• Although you report changes in pain and disability and the statistical significance of these changes you do not say anything about the clinical significance of these changes. For example, ODUI improved by 25 percentage points with ICBG vs. 30 percentage points with rhBMP-2. Are these findings clinically significant?
6. We found that the report lacked sufficient detail about your assessment of the quality of included studies and risk.

7. Please include evidence tables that summarize key characteristics of each included study. Complete tables will likely need to be presented as appendices, so a summary table such as Table 1 would be attractive to include in the body of the manuscript as well. However, we found Table 1 to be too sparse in the information presented. For example, unless all studies evaluated the same endpoints at the same time points, you should include a summary of the outcomes each study evaluated. A brief summary of trial findings and citations for each trial would also be useful additions to this table.

8. The bulk of available data come from studies of rhBMP-2 in the context of lower back procedures, but you combine these findings with result from the few studies of cervical procedures. It seems possible that the location of the surgery could in itself influence outcomes. Please consider separate analyses by surgical site or provide a sound rationale for combining these data.

9. You identify two trials that Medtronic did not fund and were able to obtain patient-level data for one of them. Please tell the reader more about this trial and why the investigators for the second trial were unwilling or unable to provide you with the data. Who funded this study? Which interventions and outcomes did it examine? What did it find?

10. Please revise your introduction to more objectively set up the purpose of this review, which as stated in the current abstract appears to be to use patient-level data to evaluate available evidence on effectiveness and safety of spinal fusion using rhBMP-2 vs. other common spinal fusion procedures that do not involve rhBMP-2. Refrain from implying that previous work in this area is biased, unfair, or unreliable. We also think that the description of YODA would be more appropriate to mention in the Methods when you describe how you obtained the data rather than in the Introduction.

11. You state in the methods that, “data extraction from published reports was undertaken by two researchers independently with any differences resolved by discussion.” Did the same two researchers extract all of the data or did different duos evaluate different subsets of included studies? Also, did the discussion to resolve differences involve the entire research group or just the two members of the group who extracted the data?

12. You use the term “efficacy” throughout the manuscript. We think that the term “effectiveness” would be more appropriate.

13. A major purpose of this review was to evaluate safety, so you should include all adverse event findings in the body of the manuscript rather than relegating it to appendices.

14. Under “Study Selection” in the Abstract, you refer to “comparative studies,” but we think “observational studies” would be a clearer term to use in this context.
15. Please follow Annals format for structured abstracts. The Abstract should include no more than 275 words organized according to the following sections: Background, Purpose, Data Sources, Study selection, data Extraction, Data Synthesis, Limitations, and Conclusions.

16. Please provide legends for all figures.

Please send your revised manuscript and cover letter to us within four weeks of receiving this letter. Number the lines of your revised manuscript consecutively throughout. In the cover letter, group and number your responses to correspond to the comments from the Editor, Statistician, and the production editor who asked for technical revisions. Please restate each comment and follow it with your response indicating what you did, why you did it, and the line number of the revised manuscript where the changes may be found. Submit your revised manuscript and cover letter at https://www.acponline.org/authors/ by clicking the "Revise Paper" link listed below your "Tasks."

Annals now requires authors to complete the International Committee of Medical Journal Editors (ICMJE) conflict of interest disclosure form. This form, which is discussed further at www.icmje.org/format.pdf, is intended to facilitate detailed reporting of conflicts of interest and standardize the format of reporting across ICMJE member journals. Upon final acceptance, each author will receive an email with a link to access, and upload, the ICMJE form. Individual forms for an article are published online on the day of publication for readers to access disclosure information.

Please keep all editorial correspondence confidential, and refrain from sharing either the correspondence itself or the essence of its content with individuals who are not your collaborators. Doing so helps ensure we can offer you advice that is in the best interests of your paper, without concern for how it might be considered or used by others.

We look forward to receiving your revised manuscript.

Sincerely,

Christine Laine, MD, MPH

Editor
7. Authors: Response Letter 1 (21 December 2012)

To the editor,
On behalf of my co-authors I have re-submitted the manuscript “The safety and efficacy of recombinant human bone morphogenetic protein-2 (rhBMP-2) for spinal fusion: An individual patient data meta-analysis” with revisions as requested by the reviewers, for consideration by Annals of Internal Medicine.
Below is a full response to all reviewers’ comments.
Yours faithfully,
Dr Mark Simmonds,
on behalf of the authors.

Response to Reviewers’ Comments

We thank the editors and reviewers for their helpful comments on our manuscript “The safety and efficacy of recombinant bone morphogenetic protein-2 (rhBMP-2) for spinal fusion”. The submitted manuscript was, of necessity, a very condensed version of a much larger report which will be made publicly available. For our original submission we struggled to fit our results within the standard word limit and so omitted considerable explanatory details, which, with the benefit of hindsight, are necessary for readers to understand what data were and were not available to us, and what we did and why. We agree that the additional information requested would improve clarity. We are grateful to the reviews for pointing out areas where this is necessary and have sought to add clarifications to the manuscript. This increases the length of the manuscript quite considerably and we thank the editors for agreeing to an increase in length. We have sought to minimise the extra length by adding only what was necessary and by including new appendices to cover statistical methods, quality assessment, searching, and to include the original protocol.
We respond to all the specific comments from the editors and reviewers below.

Comments from Editors

1. In reporting the findings of your systematic review and meta-analysis of the trials and of the observational studies, take care to follow the PRISMA and MOOSE recommendations, respectively (www.equator-network.org).

We have extended the methods section to explain the search, study selection, quality assessment and data extraction processes, following PRISMA guidelines, and have added a PRISMA-style flowchart (figure 1). Fuller details are also provided in the methods appendix.

2. While you have the great advantage of having individual patient-level data, many of your analyses reflect conventional meta-analysis. Similarly, you do not take full advantage of the longitudinal data that you have for many outcomes. In addition, the majority of the results...
section is devoted to the presentation of subset analyses while falling short of a clear, complete evaluation of the findings of the primary analyses. Refer to the statistical editor’s comments for further comment on these issues.

We reply to the relevant issues around longitudinal analyses in our reply to the statistical review below. We are unclear what is meant by “the majority of the results section is devoted to the presentation of subset analyses while falling short of a clear, complete evaluation of the findings of the primary analyses”: there is only one short section “Investigation by surgical approach” that reports subgroup analyses. While there were 17 trials provided by Medtronic only 11 of these fitted our inclusion criteria, namely, trials comparing rhBMP-2 to ICBG surgery.

3. One of our major concerns relates to a lack of clarity about the extent of the patient-level data that you had available to you for these analyses. How does the data you have overlap with the data Carragee and colleagues report in reference 6? Given that determining the safety of rhBMP-2 was a major objective of this review, it is of utmost importance that you include all available data on adverse events including data that was reported to the FDA even if it was not reported in any of the included trial or observational study reports. If you do not currently include such data, it is imperative that you do so.

We have added a new appendix table to clarify how the IPD link to publications of the Medtronic trials (appendix table 1). We had full IPD from all Medtronic trials and one further trial (Glassman). The Carragee review was limited to the published trials and not all Medtronic trials have been published. Carragee et al also based their conclusions on documents published by the FDA, however, none of these contained original data and were therefore not included in our analyses. Caragee did not have access to IPD.

To our knowledge the IPD relating to adverse events provided by Medtronic, and analysed in our paper, represents all adverse events that occurred during the course of the trials, whether or not it was reported to the FDA. We have commented on this in the first paragraph of the safety analysis section. Our full report also analyses adverse event data reported via “MedWatch” forms from rhBMP-2 use outside of the trials, but as this was not trial data, and we had no equivalent data for ICBG surgery. It was therefore omitted from this paper of the most important clinical aspects of the project.

4. Following on the above, you state that, “For evaluation of safety, we examined numbers of adverse events, including cancer, but were limited by classifications of adverse events produced by Medtronic.” Please provide more detail about these classifications so that the reader can better understand how they might influence your findings. Also, our understanding is that Medtronic was required to provide patient level data. As such, it seems
that you should have the ability to look at the primary adverse event reports and reclassify adverse events in a matter that was less limiting.

We have modified the relevant text in the safety analysis section (paragraph 1). Our checks of the adverse report forms provided with the IPD suggest that the simple classifications (eg. “back pain”, “arthritis/bursitis”) were appropriate. It was beyond the scope of our project to extract detailed data from adverse event report forms, so only the data provided in the IPD data sets were analysed.

5. The manuscript lacks clarity about how the data that Medtronic provided overlapped with the studies you identified in your own literature search. Please, describe the overlap more clearly.

We clarify this in paragraph 3 of the methods section and paragraph 1 of the result and have included a PRISMA flow chart. We have also included an appendix table that shows which publications relate to which Medtronic trials.

4. Please include a figure that summarizes your search and selection process. Did you search clinical trials registries as well as publication databases when trying to identify all relevant studies?

See new figure 1 and clarifications in paragraph 3 of the methods section. We searched clinicaltrials.gov and found one ongoing (but now terminated) trial.

5. The description of included outcomes lacks sufficient detail and clarity.

• Please clearly identify in both the abstract and the text the outcomes you sought to examine including the time points for these outcomes.

See under “data extraction” in the abstract and the effectiveness analysis section (paragraph 1) for clarification of these issues.

• What methods did included studies use to measure pain?
Only ODI (or NDI for cervical trials), SF-36 and Medtronic’s own back and leg pain scores were used. See the effectiveness analysis section (paragraph 1) for clarification.

- In some places in the manuscript you refer to pain and disability/function as separate outcomes, but in other places you seem to be referring to them as a composite outcome, making the findings about effectiveness of rhBMP-2 difficult to follow.

We have sought to simplify this throughout by referring to pain only. SF-36 PCS includes function components, but the score is dominated by pain components.

- It also appears that you treat pain as both an efficacy outcome and an adverse event, which we find problematic.

See the safety results for a short discussion of this point (paragraph 3), and also the discussion. We agree that this may be confusing, and discuss this in the main report. However that was how the data were provided by Medtronic, and for a complete analysis it was necessary to analyse all pain data. In particular pain immediately post-surgery was not available as an effectiveness outcome, only as an adverse event, so analysing this was particularly important.

- On page 5 you state, “Data on these outcomes were analyzed at 6 weeks, 3, 6, 12, and 24 months after surgery.” Did all included studies measure all outcomes and measure them at all of these time points? If not, the results must include details about which studies measured what and when. Missing outcome data is another issue on which you need to report greater detail.

All Medtronic trials provided data for the outcomes at all time points. See paragraph 4 of the results for clarification. We briefly discuss sensitivity analyses for missing data in the statistical methods section (paragraph 1), and in the methods appendix.

- Although you report changes in pain and disability and the statistical significance of these changes you do not say anything about the clinical significance of these changes. For
example, ODUI improved by 25 percentage points with ICBG vs. 30 percentage points with rhBMP-2. Are these findings clinically significant?

We now discuss clinical significance in more detail in the discussion.

6. We found that the report lacked sufficient detail about your assessment of the quality of included studies and risk.

We have provided brief details of the quality assessment in paragraph 5 of the methods section; full details are provided in the methods appendix. Results of the risk of bias assessment for the trials are given in paragraph 4 of the results and for the observational studies in paragraph 2 of the “Adverse events reported in the wider literature” section.

Quality was assessed using the Cochrane risk of bias tool for trials and the Newcastle-Ottawa scale for observational studies. As all Medtronic trials followed essentially the same protocol, the results were essentially identical for all trials, so we do not present a detailed trial-by-trial assessment of quality.

7. Please include evidence tables that summarize key characteristics of each included study. Complete tables will likely need to be presented as appendices, so a summary table such as Table 1 would be attractive to include in the body of the manuscript as well. However, we found Table 1 to be too sparse in the information presented. For example, unless all studies evaluated the same endpoints at the same time points, you should include a summary of the outcomes each study evaluated. A brief summary of trial findings and citations for each trial would also be useful additions to this table.

We have revised Table 1 to present more detail. The new appendix table 1 links trials to publications. We note that because this is an IPD analysis main Table 1 summarises the actual trials and their IPD, not publications (some trials have not been published). For this reason we think neither citations, nor a summary of trial findings, are appropriate in this table. As stated above, all Medtronic trials provided data for the outcomes at all time points, so presenting outcomes reported in not necessary in this table.

8. The bulk of available data come from studies of rhBMP-2 in the context of lower back procedures, but you combine these findings with result from the few studies of cervical procedures. It seems possible that the location of the surgery could in itself influence outcomes. Please consider separate analyses by surgical site or provide a sound rationale for combining these data.
Only one small trial (Cornerstone) of cervical surgery was included. An analysis by surgical site and approach is provided in figure 5 of the paper. For ODI there was some evidence that the cervical trial had different results from the lumbar trials, but removing the trial would not substantially alter the results. For all other outcomes there were no differences in effect between cervical and lumbar trials. Since our aim was to analyse all rhBMP-2 trials in spinal surgery, we believe the cervical trial data should be included.

9. You identify two trials that Medtronic did not fund and were able to obtain patient-level data for one of them. Please tell the reader more about this trial and why the investigators for the second trial were unwilling or unable to provide you with the data. Who funded this study? Which interventions and outcomes did it examine? What did it find?

We briefly discuss this trial at the start of the results section.

The trial (Shapiro et al) evaluated local bone and cortical wedges +/- rhBMP-2 in 40 patients undergoing single-level bilateral PLIF. It reported 100% fusion in both groups and no difference in back pain or ODI at 12 months. It was independently funded, and the author had “no interest and no time” to share data.

10. Please revise your introduction to more objectively set up the purpose of this review, which as stated in the current abstract appears to be to use patient-level data to evaluate available evidence on effectiveness and safety of spinal fusion using rhBMP-2 vs. other common spinal fusion procedures that do not involve rhBMP-2. Refrain from implying that previous work in this area is biased, unfair, or unreliable. We also think that the description of YODA would be more appropriate to mention in the Methods when you describe how you obtained the data rather than in the Introduction.

We have modified our introduction accordingly. Given the nature of this project, and the paper submitted by the Oregon group, some mention of YODA and the reasons for conducting this analysis in the introduction is necessary.

11. You state in the methods that, “data extraction from published reports was undertaken by two researchers independently with any differences resolved by discussion.” Did the same two researchers extract all of the data or did different duos evaluate different subsets of included studies? Also, did the discussion to resolve differences involve the entire research group or just the two members of the group who extracted the data?
This is now clarified in paragraph 5 of the methods section. Three researchers extracted data in pairs. For each study, disagreements were resolved by discussion within the allocated pair, or were referred to the third researcher where necessary.

12. You use the term “efficacy” throughout the manuscript. We think that the term “effectiveness” would be more appropriate.

Changed throughout.

13. A major purpose of this review was to evaluate safety, so you should include all adverse event findings in the body of the manuscript rather than relegating it to appendices.

We have moved adverse event forest plots from the wider literature into the main body of the paper (Figures 10 to 12). We had originally placed them in the appendix to reduce the number of main figures, we did not intend this to be seen as relegating these analyses.

14. Under “Study Selection” in the Abstract, you refer to “comparative studies,” but we think “observational studies” would be a clearer term to use in this context.

We have changed this accordingly.

15. Please follow Annals format for structured abstracts. The Abstract should include no more than 275 words organized according to the following sections: Background, Purpose, Data Sources, Study selection, data Extraction, Data Synthesis, Limitations, and Conclusions.

We have rewritten the abstract.

16. Please provide legends for all figures.

Legends have been provided
Reviewer 1

**Major Comments:**

1) *Page 5- bias assessment method is noted here but not provided in results*

We have slightly amended our reporting of risk of bias (paragraph 4 of results). All trials used essentially the same protocol, so detailed reporting of bias for individual trials was not deemed appropriate (see also comment 6 of the editor’s comments).

2) *Page 13- Need citation for rhBMP-2 and cancer risk with more discussion on biologic mechanisms. Unclear what last sentence in second paragraph means?*

We now give a citation on cancer risk in the discussion. The discussion has been revised to remove the confusing sentence.

3) *Were the results sensitive to including the Medtronic trial which did not include ICBG as comparator? Where other sensitivity analyses performed?*

Most Medtronic trials without an ICBG group were single-arm trials, so could not be included in any sensitivity analyses. One trial used the MAVERICK disk system as the comparator. The aim of our project was to compare rhBMP-2 to ICBG only so this trial arm was out of scope and was not analysed. As it is unclear whether MAVERICK can be considered similar to ICBG we believe any sensitivity analysis including this trial could be misleading.

4) *Cancer conclusion is too strong given lack of statistical significance. Agree trend is concerning. But need to qualify*

We have modified the discussion on this issue where cancer conclusions are discussed.

5) *Mention increased fusion rate in first paragraph of discussion.*

We have revised our discussion. We now discuss fusion rates before pain reduction.

*Figure 6 does not contribute very much. Are there a priori hypotheses for these associations? Some are likely fully unrelated to BMP or ICBG*

We think what was Figure 6 (now figure 7, adverse events forest plot) is important because it clearly shows the increased risk of certain adverse events when using rhBMP-2. While we agree that some adverse events may be unrelated to spinal surgery or treatment received, we think making a priori assumptions about which events are or are not related is inappropriate, and a full report of all adverse events is the correct approach for this paper.

**Additional Comments:**

1) *Figures are mis-numbered and hard to follow.*

Figure numbers have been corrected.

2) *Many odd acronyms used, need to reduce.*

We have sought to reduce acronym use where possible. However all the acronyms used are commonly used by spinal surgery specialists and are required to avoid considerable repetition of long medical terms (eg. ICBG – iliac crest bone graft).

3) *Need to describe direction of change of pain scores and functional measures and what these scales mean*

See the effectiveness analysis section (paragraph 1) for clarification of the pain scores.

4) *Introduction aim to provide “robust and fair assessment” -- tone down, let the reader decide.*
We have removed this phrase and rewritten this part of the introduction.

5) Table 1 - needs citation to trials references

We now provide details of publications in Appendix table 1. Table 1 in the main paper is intended to describe all the trials for which the IPD were provided. Data from these trials is taken from the IPD, not publications. Not all trials have been published, many have multiple publications, and some publications report on multiple trials. Therefore we feel including citations in this table would not be helpful.

6) Table 2 - needs to better define “at or shortly after surgery” in table

We have done this (i.e up to four weeks after surgery)

Reviewer 2

Major issues:

1. Because this is not a simple standard linear progression of SR leading to data extracted leading to MA, the overall description of the study must be clear. It is not clear whether a SR resulted in 11 Medtronic studies + 1 non Medtronic study or if these were the a priori RCTs already known to exist. Without this knowledge and without knowing what studies were excluded (if a SR was indeed performed), the reader cannot tell if these are the best data (because they are complete IPD) or the most comprehensive.

We have sought to clarify our aims and searches throughout. See paragraph 3 of the methods section, for example. Our aim was to be comprehensive by identifying all trials, and also to obtain IPD from as many RCTs as possible. We have sought to clarify this in the introduction and methods. We have sought to clarify that the comprehensive systematic literature searches were intended to identify all trials, particularly non-Medtronic trials, observational studies, and any Medtronic trials that might not have been provided. Medtronic trials (but not all their associated publications) had already been identified through provision by Medtronic. Related to item 1, as a MA/SR, the overall methods must be clear and transparent such that the study could be independently replicated like an experiment. An examination of the CDR entry in PROSPERO did not clarify the search strategy sufficiently for this to occur. Inclusion and exclusion criteria should be explicitly stated. In fact, all search details should be included along with a flow diagram. In addition, was a manual bibliography search performed. SRs reported in the orthopaedic literature—to which this study pertains—indicate several instances where relevant literature was found in bibliographic searches.

We have sought to clarify the details of the search for effectiveness trials and observational studies of safety in the methods section. A full search strategy is now provided as a web appendix. A flow diagram of the search and study selection process is now provided in figure 1. The PROSPERO entry was limited for confidentiality, because of the need to retain strict separation from the other group (in Oregon) also working on this project. Our full protocol was lodged with YODA at the outset of the project.

Although the recently reported adverse events are of public concern—in fact, I too am concerned and biased toward the same conclusions—the authors are encouraged to take a more neutral scientific stance for this MA/SR. Or to provide more compelling reasons from the literature that make it reasonable to start with a skewed stance. This bias is also subtly present in the study design where efficacy is meta-analyzed using IPD but the adverse events are not.
We have amended the discussion. We note that most of the adverse event analyses are based on the Medtronic IPD. It is only those for dysphagia, radiculitis and heterotopic bone formation that are not, where Medtronic did not collect these data.

**Minor issues:**

1. *In the methods, please report the kappa agreement between the 2 reviewers in reviewing and agreeing on inclusion of studies.*

We have not included kappa statistics. As they would only be of relevance to the non-IPD section of the review, relating to adverse event studies and safety analysis, we do not think they would be useful to readers.

2. *In the methods, please describe the definition of fusion according the Medtronic and comment on its appropriateness.*

The definition of fusion has been added to the effectiveness analysis section (paragraph 1).

3. *In the methods, please clarify the use of linear versus logistic regression: eg why not logistic regression for fusion versus pain and function outcomes?*

We have now added a methods appendix to clarify these issues. Note that the fusion versus pain analysis has been removed.

4. *In the results, please utilize an accepted scoring system for study quality.*

We consider the Cochrane risk of bias tool to be an accepted system for assessing quality. It does not generate a score but instead assesses six domains relating to trial conduct (sequence generation, allocation concealment, blinding, incomplete outcome data, selective outcome reporting and ‘other issues’). We discuss how these assessments might influence the potential for bias for specific outcomes among the included studies.

5. *In the results, please report data in the text consistently throughout for clarity and readability; ie p values (instead of “statistically significant” and RR with CI when relevant)*

We have edited the results section with the aim of consistently reporting effect estimates and confidence intervals throughout. We have removed all references to “statistical significance”. As confidence intervals are provided we do not think p-values for effect estimates are generally necessary, but some have been included for clarity.

6. *In the results and discussion, it should be made clear that adverse events including rates of cancer are associations and are not derived from the same quality of data (eg IPD) as the MA of outcomes.*

The adverse events analyses for the Medtronic trials, and in particular the cancer analyses, are based on IPD provided by Medtronic, as clarified in the safety analysis section (paragraph 1).

**Statistical review**

As a general reply to this review we note that this submitted paper is a greatly condensed adaptation of the much longer full report. In seeking to keep the paper to an appropriate length we omitted much detail on the methods involved. We thank the reviewers for suggesting a statistical appendix. We hope that including this appendix will address many of the requests for greater detail on the statistical methods, without over-burdening the main text of the paper.
General comments

1. ... The following methodological issues came up

2. Literature search. Please provide a flow diagram of the study search process, indicating the number of studies that were included and excluded along the process and the final number of studies included in the individual patient meta-analysis and in the additional meta-analysis with published studies, separately for efficacy and safety endpoints. Please also complete a search of RCT registries to identify potentially completed but unpublished studies.

A flow diagram is provided in Figure 1
The clinicaltrials.gov registry was searched and one ongoing trial (but now suspended) was identified (see Results paragraph 1).

3. Presentation of which studies contribute to different endpoints. Throughout the Results section, it was difficult to follow all the different endpoints that were being evaluated as well as which studies contributed to each endpoint. Please provide a table or a figure (either in the main text or in an appendix) with a grid showing which studies contribute to each endpoint (including efficacy and safety endpoints). For efficacy endpoints, please detail availability by time period (6 weeks, 3, 6, 12 and 24 months). In addition to this Table, at least for the primary efficacy endpoints that are based on scale measurements, please provide some guidance (either in the Methods section or in the caption of for the reader as to what is the range of the scale and what are considered clinically meaningful differences.

Generally all trials contributed to all outcomes and at all time points, so we do not think the suggested table is necessary. We have sought to clarify in the text in paragraph 3 of the results section. Exceptions to this (eg. Glassman trial did not contribute to adverse event analyses) are now stated.
We have added descriptions of the outcome scales to the efficacy analysis section (paragraph 1). We do not discuss clinically meaningful differences in the methods since these are matters of opinion rather than fact. We instead discuss clinical significance of results in the discussion.

4. Description of statistical methods. In general, we found the description of statistical methods insufficient for an informed reader to reproduce the analysis. For both efficacy and safety outcomes, please provide detail methods with respect to each of the following four aspects:

- Analysis of individual studies with individual patient data.
- Combination of studies with individual patient data.
- Combination of studies with individual patient data with studies with data available only from study reports (including more detailed description of the approach to incorporating non-randomized studies for safety endpoints).
- Sensitivity analyses.

We have provided a methods appendix to cover these issues as well as adding to the main text as described below.
There was only one eligible RCT that did not provide IPD. This lacked sufficient detail in its published form (abstract only) to be included, hence there was no need for analyses combining IPD with aggregate study report data.
Each of these items does not necessarily need a separate section, but should be clearly described in a structured way (please consider preparing a detailed Statistical Appendix with a more detailed description than that provided in the main text). With respect to the statistical methods, please:

- Specify exactly what is meant by one-stage and two-stage methods.

We now do this in the methods appendix to avoid excessive length in the methods section.

- When describing mixed models, please provide more detailed descriptions of the methods: what were the fixed and the random effects in the models, where there random intercepts only or also random coefficients, what were the estimation methods (for both linear and logistic mixed models, although this is particularly important for mixed logistic models)?

Likewise, please refer to the methods appendix.

- Please also provide a detailed description of the analytical methods for each individual study and use methods consistent with appropriate modern standards. For instance, for continuous outcomes it seems that the authors simply calculated the difference in outcomes at each point in time with respect to baseline, while we would expect a more efficient use of the data from each individual study by fitting a mixed model for longitudinal data from each study (for both continuous and dichotomous endpoints, although alternative survival-type models may be considered for longitudinal dichotomous endpoints – see below). Notice that we also expect that the authors consider the elements of each individual study with the same degree of granularity as we would expect in the report of an individual study. This includes, for instance, indicating the extent and analytical considerations of missing data and losses to follow-up in longitudinal models (what was the approach used in the analysis?), indicating the approach taken about study centers in each individual study, specifying if the estimates calculated are marginal or population-averaged, specifying the approach to subgroup analyses (notice that with the current description, it is unclear how the subgroup analyses were actually done and tested for), etc.

While we agree that a longitudinal model may be the best statistical analysis our aim was to present an analysis that may be clearly and directly understood by non-statistical readers. Also, the time available to perform this very large IPD analysis was limited, so there was little scope to develop complex models. For both these reasons we selected a two-stage meta-analysis conducted independently at each time point, accepting that this may not be statistically optimal. A longitudinal model may also require assumptions about pain trends over time, such as assuming linear change in pain over time that would be difficult to create a priori and may not be justifiable.

Missing data were minimal in these IPD. The analysis presented in the paper was therefore a complete-case analysis. We also performed a last observation carried forward analysis (see the methods appendix and statistical methods sections) to investigate the effect of missing data and found no difference in results from the complete case analysis, so this was not reported in the original submission. This is now discussed in the results at the end of the pain and fusion results sections.
The original Medtronic data did not include sufficient information to reliably take study centres into account in any models.

- For the endpoint of spinal fusion, the investigators used relative risks at different points in time, but it is unclear if this is the most appropriate model given that the endpoint is theoretically permanent and the likely occurrence of losses to follow-up. It seems that a survival-type model that takes into account the fact that fusion occurs between two visits (and not exactly at each visit) seems much more appropriate than the model considered in the manuscript. Please base the analysis for this endpoint in a survival-type model or provide a solid argument justifying the current approach.

We note that the IPD supplied by Medtronic did not include any dates of fusion, or of any other events or observations, hence a full survival analysis was not possible, as we had planned. An interval censored analysis is possible (Simmonds et al. Research Synthesis Methods 2011; 2:139-49), and we did attempt this as an exploratory analysis. There are problems with such a model, however, particularly that it is based on just three time intervals (0-6 months, 6-12, 12-24) so its estimation of the baseline hazard is very approximate, and the results cannot therefore considered to be reliable as it is a very poor approximation to a standard proportional hazards model. It is also uncertain whether a proportional hazards assumption is appropriate in this context. We also think that a summary hazard ratio would be more difficult to interpret and understand than the analysis at specific times we have presented. For these reasons we do not present such an analysis in the paper.

- For most of the analyses presented in the manuscript, the investigators had the option of providing one step or two step models. For the analysis of efficacy endpoints, the investigators selected two stage models. As sensitivity analyses, please use one-step models for efficacy endpoints.

We now briefly mention these models in the statistical methods, with details in the methods appendix, and note that the results were the same as the two-stage methods. In our report to YODA one-stage meta-analyses were used as sensitivity analyses throughout. As the results were almost identical to the two-stage methods we did not originally include them in this paper for brevity and clarity.

5. Presentation of Results: Please consider the following issues:

- We considered the presentation of the characteristics of individual studies insufficient to provide enough context for the readers of the paper. Please provide more detailed evidence tables with key characteristics of individual studies (for instance, it was unclear if we were dealing with similar studies in terms of patient characteristics and pre-intervention pain, or if we were dealing with very different types of studies). In addition, since this is an individual patient meta-analysis, the investigators have the opportunity of presenting much more detailed descriptive and analytical data for each individual study. Please take advantage of this when preparing evidence and descriptive tables of participant characteristics.

We have modified Table 1 to provide greater detail about the trials
With respect to the Figures, it seems that the numeration of the Figures did not correspond to the description in the text. In the revised version, please check the figures and tables and consider limiting the number of figures in the main text and making more effective use of figures. Given the lack of synchrony between text and figures, we found it difficult at this time to provide detailed guidance on each figure, but we ask the investigators that they carefully consider the role of each figure in the revised manuscript and provide informative and carefully designed figures to support the Results section.

Figure numbers have been corrected.

6. **Relationship between successful fusion and improvement in pain and function scores.**

   In the paper, the investigators present the impact of fusion (a post randomization event) on improvement in pain and fusion. While this type of mediation analysis is interesting, it is subject to potentially complex biases and does not add to the main aims of the paper. The authors considered that this analysis is best presented in a different paper.

   While we believe that this analysis provides useful information, suggesting that pain reduction may be unrelated to fusion, we agree that such within-group comparisons are not randomised and may be subject to bias. We have therefore removed this analysis and figure. We have replaced it with an alternative figure from our full report which plots, for each trial, the mean difference in pain score against the relative risk of fusion for the four pain outcomes (figure 6). This shows no correlation between improvements in pain and improvements in fusion. In particular rbMP-2 appears to reduce pain even when it does not improve fusion (relative risk = 1), suggesting that the reduction in pain on rbMP-2 may be due to interpretation bias in pain assessment by the patients. We recognize that this figure is essentially a meta-regression, and so has potential for bias.

7. **Discussion.** Overall, the editors found that the discussion was a bit too mechanical and did not integrate study characteristics and biological mechanisms with the findings in the analysis. In the revision of the paper, please provide a more integrated Discussion of the findings.

   We have re-written the discussion.
8. **Authors: Revised Manuscript 1 (21 December 2012)**
The safety and effectiveness of recombinant human bone morphogenetic protein-2 (rhBMP-2) for spinal fusion

An individual participant data meta-analysis

Running title:

Bone morphogenetic protein-2 in spinal fusion

Mark C Simmonds¹
Jennifer VE Brown¹
Morag K Heirs¹
Julian PT Higgins¹²
Richard J Mannion³
Mark A Rodgers¹
Lesley A Stewart ¹ *

1 Centre for Reviews and Dissemination, University of York, UK
2 MRC Biostatistics Unit, Cambridge, UK
3 Addenbrookes Hospital, Cambridge, UK

* Corresponding author
Centre for Reviews and Dissemination
University of York, UK
YO10 5DD
Tel: 01904 321040
Email: lesley.stewart@york.ac.uk

Project funded by a competitive grant from the Yale Open Data Access (YODA) project.

Word count: 4379 words
Abstract

Background

Recombinant human bone morphogenetic protein-2 (rhBMP-2) is used to promote fusion in spinal surgery. Controversy around safety requires thorough evaluation of the evidence supporting its use.

Purpose

We aimed to evaluate the effectiveness and safety of rhBMP-2 in spinal surgery, principally through systematic review and meta-analysis of individual participant data (IPD).

Data sources

IPD were available via the Yale Open Data Access project and obtained from one further trial. Supplementary adverse event data were extracted from publications.

Study selection

Randomised controlled trials of rhBMP-2 versus iliac crest bone graft (ICBG) in patients undergoing spinal fusion surgery for degenerative disc disease and related conditions were eligible. Observational studies in similar participants were eligible for evaluation of safety.

Data extraction

Primary outcomes were: pain, Oswestry disability index (ODI), SF-36, successful fusion and adverse events. Supplied IPD were collated. Some additional data on adverse events were extracted from publications.

Data synthesis

IPD from 1411 participants were analysed. At two years rhBMP-2 patients had ODI scores 3.5% lower than ICBG patients (95% CI: 0.49 to 6.47); radiographic fusion was 14% more common (95% CI:
3% to 25%). At or shortly after surgery, several adverse events were more common among rhBMP-2 patients. Cancer was almost twice as common (RR 1.84, 95% CI: 0.81 to 4.16).

Limitations

Observational studies in the supporting evaluation of safety were diverse and many may be at risk of bias.

Conclusions

There is clear evidence that rhBMP-2 compared to ICBG increases fusion rates and results in additional, but unlikely to be clinically meaningful reductions in pain. There is evidence that rhBMP-2 increases early post-surgical pain, and inconclusive evidence that it may increase cancer incidence.

(274 words)
Recombinant human bone morphogenetic protein-2 (rhBMP-2) is widely used as an alternative to iliac crest bone graft (ICBG) to promote fusion in spinal surgery (1) (2). Following Food and Drug Administration approval of rhBMP-2 in anterior lumbar interbody fusion surgery (3), numbers of spinal procedures using rhBMP-2 have grown rapidly, including off-label indications (2) (4). Recently, concerns have grown about adverse events associated with rhBMP-2, with a review of publicly available data suggesting that the risk of adverse events is 10 to 50 times higher than reported in trial publications (5).

The Yale Open Data Access (YODA) project invited the manufacturer, Medtronic, to provide full data from all their trials of rhBMP-2, to allow independent re-analysis. YODA subsequently funded the Centre for Reviews and Dissemination and one other group to do so.

This provided an opportunity to undertake an individual participant data (IPD) meta-analysis, which is regarded as a ‘gold standard’ approach to synthesis, and allows more in-depth analysis than is possible in a systematic review of the published literature.

We embedded our meta-analysis of IPD within a wider systematic review and sought to examine all available research evidence. This included the IPD made available by Medtronic from all their trials of rhBMP-2 (randomised and non-randomised), as well as from one additional RCT. To investigate the safety of rhBMP-2 more widely, we sought all observational studies of rhBMP-2 in spinal surgery that reported adverse events.

**Methods**

**Eligibility, search, data collection and critical assessment**

Methods were, in advance of knowledge of the IPD to be provided, pre-specified in a protocol, registered in PROSPERO in February 2012 (registration number CRD42012001907)(6) which was also lodged with the YODA team at the outset of the project.
Eligible for inclusion in our principal analysis were all randomised controlled trials comparing rhBMP-2 with ICBG in spinal fusion surgery, irrespective of spinal level or surgical approach. We included trials of the licensed INFUSE® rhBMP-2 formulation and unlicensed AMPLIFY™ and BCP formulations. For supporting analyses of safety, all studies comparing rhBMP-2 with any other spinal fusion technique and reporting adverse events were eligible for inclusion.

As our re-analysis of the Medtronic data was done in the context of a full systematic review, we performed a systematic literature search to identify all eligible studies, particularly those not provided by Medtronic. This involved searching multiple bibliographic databases including CENTRAL, MEDLINE, EMBASE and Science Citation Index in January 2012, and automated ‘current awareness’ searches up to June 2012 (see web appendix for detailed search strategies). We also searched clinicaltrials.gov to identify ongoing or unpublished randomised trials and published a call for evidence (see web appendix for details).

The YODA project team provided us with IPD from 17 trials undertaken by Medtronic. Eleven of these were RCTs comparing rhBMP-2 to ICBG surgery and were eligible for inclusion in our principal evaluation of effectiveness. Of the others, four were single-arm trials of rhBMP-2 and one (19) used a different comparator. These were included in our supporting consideration of adverse events. One trial was stopped early after recruiting only three patients and not considered further.

One researcher collated IPD from across the multiple SAS data files provided by Medtronic, and a second researcher checked the process. The IPD were checked for completeness, internal consistency, improbable values and balance of patient characteristics across treatment arms (see methods appendix for details). The study selection and data extraction from published reports was undertaken by three researchers independently working in pairs, with any differences resolved by discussion or by referring to the third researcher. We assessed risk of bias in the RCTs using the Cochrane Collaboration’s ‘risk of bias’ tool, and using a modified form of the Newcastle-Ottawa scale.
for non-randomised studies (7, 8) (see methods appendix for details). Risk of bias was assessed by at least two researchers independently, with any differences resolved by discussion.

**Effectiveness analysis**

We pre-specified our primary outcomes as those that we considered likely to be important to patients. The Oswestry Disability Index (ODI) and Neck Disability Index (NDI) for cervical spinal surgery measure, respectively, lower back and neck pain on a scale from 0 (no pain) to 100% (extreme pain). The SF-36 Physical Component Score (SF-36 PCS) assesses both pain and physical function on a scale from 0 (worst) to 100% (best). Medtronic also provided data on back and leg pain, which were both measured and supplied on a scale from 0 (no pain) to 20 (extreme pain). Spinal fusion was supplied as success or failure as defined radiographically within the trials according to Medtronic criteria that required all of the following: evidence of bridging trabeculae, no evidence of motion (<3mm difference in translation, less than 5° difference in angular motion) and no evidence of radiolucency. We analysed these outcomes at 6 weeks, 3, 6, 12 and 24 months after surgery. We did not analyse the limited data that were available for longer follow-up times because these were available for only a minority of the participants.

We also considered four secondary outcomes: duration of hospital stay, operating time, successful return to work / usual activity and use of pain-relief medication.

**Statistical methods**

Full details of statistical methods and analyses are provided in the methods appendix. In the analyses of effectiveness for continuously distributed outcomes (such as ODI) we calculated mean differences between treatment arms in the change in score from pre-operative values. For dichotomous outcomes (such as successful fusion) we calculated relative risks. These were
calculated separately for every trial at every time point. We then used standard random-effects meta-analytic techniques (9) (10) to combine effect estimates across trials. Separate meta-analyses were performed for each of the specified time points. Linear and logistic random-effects regression models were used to combine all data from all trials in “one-stage” meta-analyses as sensitivity analyses.(11, 12) To explore the influence of missing observations we used the last observation carried forward method to impute data which were missing in sensitivity analyses.

Heterogeneity was assessed in all meta-analyses using the Higgins’ I² statistic (13) and Cochran’s Q test. We examined whether effects varied according to the type of spinal surgery, or by rhBMP-2 formulation (INFUSE® or AMPLIFY ™) using subgroup analysis (stratified by trial). We investigated whether patient-level factors (age, sex, smoking, alcohol consumption, body mass index, diabetic status, and history of spinal surgery for back pain) were associated with the effectiveness of rhBMP-2 surgery using a one-stage random-effects regression model (11) including interaction terms between patient-level factors and treatment.

Safety analysis

For evaluation of safety, we examined numbers of adverse events, including cancer, provided in the IPD according to the classifications of adverse events produced by Medtronic. Case report forms did not form part of the Medtronic submission, but summaries of cases were provided for the majority of patients in the clinical study reports. A full clinical assessment of these narratives was beyond the scope of this time-limited project, but checks in one trial (Interfix PLIF) showed that the Medtronic classifications of events seemed appropriate. To our knowledge these data represent all adverse events that occurred during the follow-up periods of these trials. Because the numbers of specific adverse events in most trials were generally small, we used one-stage random-effects logistic regression meta-analysis models (12) to analyse these data (see methods appendix for model details). Results of these analyses are presented as odds ratios.
From other published studies that compared spinal fusion surgery using rhBMP-2 with ICBG and other comparators, we extracted data on numbers of adverse events as specified in our protocol. This was restricted to comparative studies including more than ten adult participants.

Our full report to YODA is publicly available and examines a wide range of issues around the effectiveness and safety of rhBMP-2 in spinal surgery. It also includes an examination of the reliability of the literature published in this field (which will be submitted for publication separately) (14) (15).

**Results**

Figure 1 shows the result of our systematic search. We identified 41 publications relating to eligible RCTs. Among these were publications of ten Medtronic trials, and three trials that did not use ICBG as a comparator, so were included only in our safety analyses. We identified two eligible randomised trials not conducted by Medtronic which compared rhBMP-2 with ICBG surgery. We requested IPD from the authors, and obtained these from one (Glassman) (16). IPD from the other trial in 40 patients undergoing single-level bilateral posterior lateral interbody fusion were not made available by the trial author (17). This trial reported 100% fusion in both groups and no difference in back pain or ODI at 12 months, and therefore could not contribute further to our analyses. One ongoing trial was identified, for which recruitment had been suspended, but as it had not been closed it could not be included (18).

Our IPD meta-analysis was therefore based on data from 1,411 patients included in 11 Medtronic RCTs that compared rhBMP-2 with ICBG surgery plus data from 106 participants in one RCT not sponsored by Medtronic (Glassman) (16). Data from one Medtronic RCT (19) were not included because the comparator arm was not ICBG surgery. Table 1 summarises the 12 included trials.

Details of publications associated with each the Medtronic trials are provided in Appendix table 1.

Note however that our analyses used the supplied IPD and not data reported in these publications.
For all the included trials, data on all pain outcomes were available at all time points from six weeks to 24 months. Similarly data on spinal fusion were available for all trials (except the LT-Cage pilot trial and the Glassman trial) at all times from six months onwards.

As the trial protocols for all the Medtronic trials were very similar, our assessment of risk of bias was the same for all trials. We found that the randomisation and allocation concealment procedures were adequate for all trials. Neither patients nor physicians were blinded to the treatment received, and all pain and function outcomes were patient assessed, so there was a potential for bias in these outcomes. Assessment of successful fusion was performed blind to treatment received, minimising bias.

**Effectiveness**

**Pain**

Figure 2 illustrates meta-analyses across the 12 RCTs for four pain outcomes (SF-36 PCS also incorporates physical function assessment). Points on the plot represent mean differences in changes in scores (from pre-operative values) line between rhBMP-2 and ICBG at each time point, with vertical lines showing the 95% confidence intervals for these. For the first three outcomes, points below the horizontal zero line indicate a benefit of rhBMP-2; for SF-36 PCS, points above the line indicate a benefit of rhBMP-2.

The use of rhBMP-2 generally achieved greater pain reduction (from pre-operative values) than did ICBG from six months after surgery onwards. At 24 months after surgery the ODI was approximately 3.5 percentage points better (MD -3.48, 95%CI: -6.47 to -0.49, $I^2 = 38\%$) and back pain was similarly better among rhBMP-2 patients, by more than one point on the 20-point scale used (MD -1.58, 95%CI: -2.65 to -0.51, $I^2 = 44\%$). SF-36 PCS was approximately two percentage points higher for patients receiving rhBMP-2 (MD 1.93, 95%CI: 0.63 to 3.22, $I^2 = 0$) at 24 months. Improvements in pain and function were small. There was no evidence of a difference in leg pain reduction between treatment groups (MD -0.59, 95%CI: -1.27 to 0.09, $I^2 = 0$). Patients in both groups improved
considerably over time such that the extra benefit of rhBMP-2 over ICBG surgery was small in comparison (appendix figure 1). ODI improved by approximately 26 percentage points at 24 months for ICBG patients and by 30 percentage points for rhBMP-2 patients.

Results of the one-stage linear regression meta-analysis models of pain outcomes were almost identical to the results presented above (see appendix figure 2). Similarly, using the last observation carried forward where pain data were missing gave almost identical results.

**Fusion**

Data on fusion were not available for the LT-Cage pilot and Glassman trials (total 120 patients). For the remaining trials, meta-analyses of success of spinal fusion at 6, 12 and 24 months are illustrated in Figure 3. This shows that rhBMP-2 increased the rate of spinal fusion by 10 to 20% (where 69% of ICBG patients achieved fusion within 24 months). Figure 4 shows a forest plot for successful fusion 24 months after surgery, where rhBMP-2 increased fusion rates by 14% (RR 1.14, 95%CI: 1.03 to 1.25). There was, however, substantial heterogeneity in the relative risk of successful fusion across trials at earlier time points, with $I^2$ at six, 12 and 24 months being 97%, 80% and 76% respectively. This heterogeneity is apparent in Figure 4.

Results from the one-stage logistic regression meta-analysis model for successful fusion were almost identical to the results above. Similarly, using the last observation carried forward where fusion data were missing gave almost identical results.

**Investigation by surgical approach**

We investigated whether the effectiveness of rhBMP-2 varied across anterior lumbar fusion, posterior lumbar fusion, and anterior cervical fusion (see Table 1) using subgroup analysis. Figure 5 shows the results of these subgroup analyses for ODI and successful fusion 24 months after surgery. A test for heterogeneity showed that there was moderate evidence of a difference between surgery types for ODI (p-value 0.065), but this was primarily due to the very large benefit of rhBMP-2 on the
NDI observed in the single, small cervical surgery trial (23 patients). Excluding this trial resulted in no clear difference in the effectiveness of rhBMP-2 between anterior or posterior approaches ($p = 0.17$). There was no evidence of a difference in the relative risks of successful fusion at two years across surgery types ($p = 0.88$). Nor was there evidence of difference between surgery types for any other outcome.

We also considered whether a range of patient-level factors were associated with the effectiveness of rhBMP-2. There was generally no evidence of interactions between rhBMP-2 and the patient-level factors (age, sex, smoking, alcohol consumption, body mass index, diabetic status, and history of spinal surgery). One possible exception was that, for people with a history of previous spinal surgery, there was no difference in the effectiveness of rhBMP-2 and ICBG, either at reducing ODI or improving fusion rates (whereas there was an overall benefit of rhBMP-2 across all patients). Given the number of analyses performed, this result may be a chance finding. We do not present these results here. They are available in our full report.

Further outcomes

Among secondary outcomes there was no overall evidence of any difference in duration of hospital stay (mean difference -0.15 days, 95%CI: -1.35 to 1.06 days) or that rhBMP-2 surgery increased the probability of returning to work or usual activity earlier when compared with receiving ICBG (for example, at two years: RR 1.01, 95%CI: 0.88 to 1.17). Using rhBMP2 shortened operating times, by 21 minutes (95% CI: 15 to 27, see Appendix figure 3), from an average of 135 minutes. There was no evidence that use of pain-relief differed between rhBMP-2 and ICBG arms at any time (see Appendix figure 4).

Figure 6 shows, for each of the four pain outcomes (ODI, SF-36 PCS, back and leg pain), the mean difference in pain 24 months after surgery plotted against the relative risk of successful fusion 24 months after surgery. Each point in the figure represents the results for one Medtronic trial. This
figure does not indicate a consistent relationship between improvements in fusion due to rhBMP-2 and improvements in pain or function. If successful fusion resulted in reduced pain we might expect trials with higher fusion rates on rhBMP-2 to show greater improvement in pain scores, but this does not appear to be the case. In particular those trials where fusion was less common in the rhBMP-2 patients (INTER FIX ALIF pilot and BCP US), still showed a benefit of rhBMP-2 in terms of improved SF-36 PCS. The BCP US trial also showed a benefit of rhBMP-2 on ODI and back pain. The apparent benefits of rhBMP-2 in pain reduction do not, therefore, appear to be due to increased fusion rates.

Safety

The numbers of adverse events in the 11 Medtronic RCTs are given in Table 2.

All Medtronic trials provided data on adverse events at all the specified time points and also at or shortly after (up to four weeks) surgery. Reporting of adverse events in the Glassman trial was not consistent with the Medtronic trials so this trial was not included in these analyses. The numbers of adverse events in the 11 Medtronic RCTs are given in Table 2.

We note that pain was reported as an adverse event in the Medtronic IPD as well as being assessed as an effectiveness outcome using the pain scales discussed earlier. The reasons for this were not clear, but for completeness we analyse pain reported as an adverse event here, particularly because pain immediately after surgery was not recorded on the pain scales, only as an adverse event (recorded as present or absent).

Figure 7 illustrates the results of the one-stage meta-analyses for adverse events across these 11 Medtronic RCTs at, or shortly after (up to four weeks), surgery. Some types of adverse event were more common among rhBMP-2 patients: arthritis/bursitis, implant-related events, neurological events, other pain, retrograde ejaculation, wound complications and vascular events all had at least
a 50% increase in risk of occurring. Because there were few events, confidence intervals were wide and findings inconclusive. For back and leg pain there was clear evidence of a higher incidence among rhBMP-2 patients (OR 1.92, 95% CI: 1.14 to 3.25, \( p = 0.004 \)).

Figure 8 shows the results of meta-analyses for four key categories of adverse event (implant related, infections, neurological, any pain) across all time periods. Again, as events were uncommon the evidence is generally inconclusive. There was no evidence of a difference in the risk of adverse events from three months after surgery onwards. At, or shortly after, surgery there was inconclusive evidence that implant related events were more common in rhBMP-2 patients (OR 1.87, 95% CI: 0.67 to 5.21), as were neurological events (OR 1.58, 95% CI: 0.88 to 2.83). Only for pain was there clear evidence that this was more common in the rhBMP-2 patients at or shortly after surgery (OR 1.78, 95% CI: 1.06 to 2.95, \( p = 0.007 \)). This contrasts with the results seen in the analyses of ODI, SF-36 PCS and back pain in the effectiveness analyses where pain reduction was greater in the rhBMP-2 patients from three months after surgery onwards.

**Cancer**

Table 3 summarises the cancers observed in the 11 Medtronic RCTs, of which five observed at least one cancer case. It excludes pre-existing cancers. It includes three cases of cancer in the INFUSE®/LT-CAGE® open pivotal trial that were identified through extended follow-up of only the rhBMP-2 patients. However, these three cancers were not included in the quantitative analyses, as this would have been unfair against rhBMP-2 since any equivalent cancers occurring in the ICBG had not been sought. A one-stage random-effects meta-analysis model found that cancer was nearly twice as common among rhBMP-2 patients (RR 1.98, 95% CI 0.86 to 4.54), but the 95% confidence intervals were consistent with risk in rhBMP-2 patients being anywhere from 14% lower to 454% higher. The absolute risk of cancer was low (3% in rhBMP-2 patients). A forest plot for the equivalent two-stage analysis is given in Figure 9 (RR 1.84, 95% CI: 0.81 to 4.16). The relative risk of cancer was similar across trials. In particular, the relative risk of cancer in the AMPLIFY™ trial, which
used a different preparation of rhBMP-2 at a higher dose, was no greater than in those trials which used INFUSE ® (p = 0.82). We note that as well as the three additional cancers identified during additional follow up, there were also three cancers among rhBMP-2 patients in the MAVERICK™ trial and two in a single-arm Medtronic trial; these were not included in the analyses here as there was no ICBG comparator.

Adverse events reported in the wider literature
We also investigated adverse events data reported in publications of randomised trials with comparators other than ICBG and in non-randomised studies. We identified 35 observational studies (in 43 publications) that reported adverse effects of rhBMP-2 for at least ten adult patients. There were 14 studies of posterior lumbar fusion, five in anterior lumbar fusion, ten in cervical fusion, and eight using multiple spinal fusion procedures. These studies used a variety of spinal fusion techniques as a control, including ICBG, local bone graft, allograft and bone marrow aspirates. Other than for the Medtronic Maverick trial, we did not have IPD for any of these studies. These studies are summarised in Appendix table 2. Quality assessment found that all included a cohort of patients representative of those likely to receive treatment in practice, and exposure to treatment was clearly established. However, most made no attempt to match or control for potential confounding factors and data on the comparability of the treatment groups was generally limited or not reported.

Given the diversity in the nature of these studies and their potential for bias (as a body of evidence) we did not combine studies in meta-analyses.

Despite the methodological issues there is some suggestive evidence of higher rates of particular adverse events among rhBMP-2 patients. In particular, heterotopic bone formation (reported in nine studies) was more common among rhBMP-2 patients, though whether this led to any clinical consequences for those patients was unclear (see Figure 10). Leg pain and radiculitis (five studies) appeared to be more common among rhBMP-2 patients, as had been observed in the Medtronic trials (see Figure 11). Osteolysis was more common among rhBMP-2 patients, but only two studies
reported on this event. Dysphagia (eight studies) appeared to be more common for rhBMP-2 patients in cervical spinal surgery, although there were some inconsistencies in the results of these studies (see Figure 12). Comer et al (20) compared four consecutive-patient cohorts undergoing ALIF with or without rhBMP-2, reporting a higher rate of retrograde ejaculation among rhBMP-2 treated patients (6.3% vs 0.9%, p=0.0012).

**Discussion**

The analyses in this systematic review are based on data from 1411 individual participants in 12 eligible RCTs, including all trials conducted by Medtronic (both published and unpublished) and one further trial. Data on all measured outcomes were provided. We found the randomisation procedures to be adequate in all trials, but participants were not blind to the treatment received. While assessment of some outcomes such as radiologic assessment of fusion was blinded, patient reported outcomes relating to pain were not. Follow up was reasonably complete up to our final analysis time point of 24 months. Although there are some caveats, in general we consider the body of evidence for comparative effectiveness to be strong.

We found clear evidence that rhBMP-2 when compared with ICBG, improves rates of fusion, according to Medtronic definitions which may have been stringent given that only 69% of ICBG patients achieved fusion within two years, which is lower than would be expected generally. There were high levels of inconsistency across trials, with large $I^2$ values at all time points.

We also found that rhBMP-2 improves back pain and quality of life when compared with ICBG at between six months and two years after surgery. These improvements in pain are modest, and fall below the thresholds of what are regarded to be clinically meaningful improvements in pain (estimated as being between 4 and 17 percentage points for ODI, and at least 5.4 points for SF-36 PCS (21) (22)). Thus it does not seem that improved rates of fusion lead to clinically meaningful improvements in the level of long-term pain.
In general there does not appear to be a strong correlation between successful fusion and pain reduction. Those trials with higher fusion rates on rhBMP-2 did not also achieve greater pain reduction. In the trials where fusion rates were worse among rhBMP-2 patients there was still a greater pain reduction in rhBMP-2 patients. It therefore appears that, either rhBMP-2 surgery has an effect on pain over and above that brought about by fusion, which seems medically unlikely, or there was some bias in the interpretation of pain. As participants were neither blinded to treatment received nor to their fusion status, they may have reported exaggerated benefits on the 'new' treatment and biased pain assessment in favour of rhBMP-2.

In contrast, the analysis of adverse events reported in the IPD showed an increased risk of pain associated with rhBMP-2 in the immediate post-surgical period. Although this might appear to contradict the finding that rhBMP-2 reduces pain from six months onward, it may be that rhBMP-2 surgery leads to increased pain shortly after surgery but reduces pain in the longer term.

The IPD also indicate a possible increase in the risk of cancer associated with rhBMP-2 with nearly double the number of new cancers occurrences when compared to ICBG patients. The overall absolute risk of cancer is low in both groups, however, and so whether this increased risk is genuine is uncertain, but it is consistent with the literature suggesting a possible link between BMP and cancer.(23)

This review included both the licensed INFUSE® preparation of rhBMP-2 and unlicensed AMPLIFY™ and BCP preparations, which use a higher dose or rhBMP-2. We found no evidence of any difference in effectiveness, safety or cancer risk between licensed and unlicensed preparations.

Data on adverse events in the wider literature raise concerns that rhBMP-2 may increase the risk of heterotopic bone formation, osteolysis, radiculitis and retrograde ejaculation. However, these are based on only the published literature, and observations are from non-randomised studies most of
which provided little information about the comparability of groups. These findings should therefore be interpreted cautiously.

Conclusions

The use of rhBMP-2 in spinal fusion surgery increases the likelihood of successful fusion at up to two years, but this does not seem to translate into clinically significant reduction in pain. The small improvements in fusion and in the level of pain reduction, which manifest after six months, also seem to come at the expense of more frequent presence of pain in the immediate post-operative period, and a possibly increased number of cancer cases. We suggest that it is important that these findings are explained clearly to patients so that they are able to make informed choices about the type of surgery they would prefer.

Acknowledgements

We thank Leah Carreon for providing the IPD for the Glassman trial, Kath Wright for performing the literature search and Charlotte Seneschall for assistance in extracting data for the safety evaluation.

References


<table>
<thead>
<tr>
<th>Trial name</th>
<th>Date of 1st patient enrolment</th>
<th>Surgical Approach</th>
<th>Intervention/Control</th>
<th>Patients randomised/enrolled</th>
<th>Patients excluded</th>
<th>Patients evaluated</th>
<th>Mean age</th>
<th>Percentage male</th>
<th>Previous spinal surgery (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT-CAGE pilot</td>
<td>January 1997</td>
<td>Anterior lumbar interbody fusion (ALIF) Single level open/laparoscopic</td>
<td>rhBMP-2/ACS + LT-CAGE (1.5mg/ml)</td>
<td>11</td>
<td>0</td>
<td>11</td>
<td>42</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICBG + LT-CAGE</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>40</td>
<td>67</td>
<td>0</td>
</tr>
<tr>
<td>LT-CAGE open</td>
<td>August 1998</td>
<td>ALIF Single level open</td>
<td>rhBMP-2/ACS + LT-CAGE (1.5mg/ml)</td>
<td>145</td>
<td>2</td>
<td>143</td>
<td>43</td>
<td>55</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICBG + LT-CAGE</td>
<td>137</td>
<td>1</td>
<td>136</td>
<td>42</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>Bone dowel pilot</td>
<td>April 1998</td>
<td>ALIF Single level open</td>
<td>rhBMP-2/ACS + cortical bone dowels (1.5mg/ml)</td>
<td>24</td>
<td>0</td>
<td>24</td>
<td>41</td>
<td>33</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICBG + cortical bone dowels</td>
<td>23</td>
<td>1</td>
<td>22</td>
<td>45</td>
<td>45</td>
<td>32</td>
</tr>
<tr>
<td>Bone dowel pivotal</td>
<td>September 2000</td>
<td>ALIF Single level open</td>
<td>rhBMP-2/ACS + cortical bone dowels (1.5mg/ml)</td>
<td>64</td>
<td>9</td>
<td>55</td>
<td>39</td>
<td>44</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICBG + cortical bone dowels</td>
<td>31</td>
<td>1</td>
<td>30</td>
<td>42</td>
<td>30</td>
<td>33</td>
</tr>
<tr>
<td>Interfix PLIF</td>
<td>March 1999</td>
<td>Posterior lumbar interbody fusion (PLIF) Single level</td>
<td>rhBMP-2/ACS + INTER FIX cages (1.5mg/ml)</td>
<td>35</td>
<td>1</td>
<td>34</td>
<td>46</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICBG + INTER FIX cage</td>
<td>36</td>
<td>3</td>
<td>33</td>
<td>46</td>
<td>45</td>
<td>39</td>
</tr>
<tr>
<td>Cornerstone pilot</td>
<td>September 1999</td>
<td>One or two level anterior cervical interbody fusion</td>
<td>rhBMP-2/ACS + Cornerstone implant + Atlantis plate (1.5mg/ml)</td>
<td>18</td>
<td>0</td>
<td>18 10/18 had single level surgery</td>
<td>51</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICBG/Cornerstone implant + Atlantis plate</td>
<td>15</td>
<td>0</td>
<td>15 8/15 had single level surgery</td>
<td>47</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>Mastergraft pilot</td>
<td>April 2003</td>
<td>Posterior lumbar fusion (PLF) Single level</td>
<td>rhBMP-2/ACS/MASTERGRAFT/CD HORIZON (12mg)</td>
<td>27</td>
<td>2</td>
<td>25</td>
<td>56</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Trial name</td>
<td>Surgical Approach</td>
<td>Intervention/Control</td>
<td>Patients randomised/enrolled</td>
<td>Patients excluded</td>
<td>Patients evaluated</td>
<td>Mean age</td>
<td>Percentage male</td>
<td>Percentage previous spinal surgery (%)</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------------</td>
<td>-----------------------------------------------------------</td>
<td>-----------------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>----------</td>
<td>----------------</td>
<td>----------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Interfix ALIF pilot</td>
<td>ALIF Single-level open</td>
<td>rhBMP-2/ACS+NOVUS LC Device</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td>45</td>
<td>44</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>February 1999</td>
<td></td>
<td>ICBG+NOVUS LC</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>45</td>
<td>45</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>BCP US</td>
<td>PLF Single level</td>
<td>rhBMP-2/BCP (3mg/ml)</td>
<td>22</td>
<td>0</td>
<td>11</td>
<td>50</td>
<td>55</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>May 1999</td>
<td></td>
<td>rhBMP-2/BCP + TSRH (2mg/ml)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ICBG + TSRH</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>52</td>
<td>40</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>BCP Canada</td>
<td>PLF One or two level</td>
<td>rhBMP-2/BCP + CD HORIZON or rhBMP-2/BCP + TSRH</td>
<td>102</td>
<td>4</td>
<td>98</td>
<td>53</td>
<td>36</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>September 1999</td>
<td></td>
<td>ICBG + CD HORIZON or ICBG + TSRH</td>
<td>105</td>
<td>6</td>
<td>99</td>
<td>53</td>
<td>48</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Amplify</td>
<td>PLF Open bilateral</td>
<td>rhBMP-2/CRM/CD Horizon (2mg/ml)</td>
<td>262</td>
<td>23</td>
<td>239</td>
<td>53</td>
<td>45</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>March 2002</td>
<td></td>
<td>ICBG + CD Horizon</td>
<td>256</td>
<td>32</td>
<td>224</td>
<td>52</td>
<td>42</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Glassman (not Medtronic)</td>
<td>PLF Single or multilevel</td>
<td>rhBMP-2/ACS + fixation</td>
<td>n/a</td>
<td>n/a</td>
<td>52</td>
<td>69</td>
<td>30</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>May 2004</td>
<td></td>
<td>ICBG + fixation</td>
<td>n/a</td>
<td>n/a</td>
<td>54</td>
<td>70</td>
<td>33</td>
<td>37</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Incidence of adverse events in 11 Medtronic RCTs

<table>
<thead>
<tr>
<th>Nature of adverse event</th>
<th>At all times</th>
<th>At or shortly after surgery (up to 4 weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rhBMP-2</td>
<td>ICBG</td>
</tr>
<tr>
<td>Total number of patients</td>
<td>672</td>
<td>609</td>
</tr>
<tr>
<td>Arm and neck (upper extremity) pain</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Arthritis or bursitis</td>
<td>24</td>
<td>16</td>
</tr>
<tr>
<td>Back and leg (lower extremity) pain</td>
<td>268</td>
<td>207</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>93</td>
<td>99</td>
</tr>
<tr>
<td>Death</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Dural injury</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>117</td>
<td>115</td>
</tr>
<tr>
<td>Graft related bone formation</td>
<td>---</td>
<td>45</td>
</tr>
<tr>
<td>Heterotopic bone formation</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>Implant</td>
<td>22</td>
<td>10</td>
</tr>
<tr>
<td>Infection</td>
<td>104</td>
<td>90</td>
</tr>
<tr>
<td>Neurological</td>
<td>152</td>
<td>119</td>
</tr>
<tr>
<td>Osteolysis **</td>
<td>12</td>
<td>---</td>
</tr>
<tr>
<td>Other pain</td>
<td>97</td>
<td>63</td>
</tr>
<tr>
<td>Respiratory</td>
<td>34</td>
<td>27</td>
</tr>
<tr>
<td>Retrograde ejaculation</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Spinal</td>
<td>70</td>
<td>63</td>
</tr>
<tr>
<td>Trauma</td>
<td>153</td>
<td>137</td>
</tr>
<tr>
<td>Urogenital</td>
<td>81</td>
<td>63</td>
</tr>
<tr>
<td>Vascular</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Vertebral fracture</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Wound complication</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

* Only reported for Interfix PLIF trial as additional data beyond the protocol

** Only reported in rhBMP-2 arms of Bone Dowel trials as additional data beyond the protocol
Table 3: Incidence of cancer in the Medtronic RCTs.

<table>
<thead>
<tr>
<th>Cancer</th>
<th>rh-BMP-2</th>
<th>ICBG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skin cancers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanoma</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Basal cell carcinoma</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Breast</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Colon</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Larynx</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Leukaemia</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lung</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Non-Hodgkin’s lymphoma</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ovarian</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Pancreatic</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Prostate</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Stomach</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Testicular</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Thyroid</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Total cancers</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Total patients</td>
<td>672</td>
<td>609</td>
</tr>
</tbody>
</table>
Figure legends

Figure 1: Flowchart displaying the flow of materials through the review process

Figure 2: Meta-analyses of pain and function outcomes at 6 weeks, 3, 6, 12 and 24 months after surgery.

Figure 3: Meta-analyses of successful spinal fusion at 6, 12 and 24 months after surgery.

Figure 4: Forest plot of relative risk of successful fusion 24 months after surgery.

Figure 5: Subgroup analyses by surgical approach for ODI and successful fusion.

Figure 6: Relationship between relative risk of fusion and mean difference in pain outcomes across trials 24 months after surgery.

Figure 7: Meta-analysis of adverse events by category at or shortly after time of surgery in 11 Medtronic RCTs

Figure 8: Meta-analyses of adverse events according to adverse event category.

Figure 9: Forest plot of cancer incidence in the Medtronic RCTs.

Figure 10: Heterotopic bone formation in five non-randomised studies

Figure 11: Leg pain or radiculitis in four non-randomised studies

Figure 12: Dysphagia in six non-randomised studies of cervical surgery

Appendix figure 1: Reduction in pain scores from pre-operative results by treatment received.

Appendix figure 2: Results of the one-stage meta-analyses of pain and function outcomes at 6 weeks, 3, 6, 12 and 24 months after surgery.

Appendix figure 3: Mean difference in operating time between rhBMP-2 and ICBG in the Medtronic trials
Appendix figure 4: Meta-analysis of use of four types of pain-relief medication in the Medtronic trials

Appendix tables

Appendix table 1: Details of publications of Medtronic trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Publications</th>
<th>Full paper</th>
<th>Abstract only</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT-CAGE pilot RCT</td>
<td>Boden et al 2000</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LT-CAGE open RCT</td>
<td>Dickman 2001</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gornet et al 2001</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2002</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2003a</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2003b</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chhabra et al 2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chhabra et al 2007</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Van Genugten et al 2008</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alt et al 2009</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2009</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2011</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LT-CAGE laparoscopic Single-arm trial</td>
<td>Kleeman et al 2001</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zdeblick et al 2001</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2003a</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chhabra et al 2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chhabra et al 2007</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Van Genugten et al 2008</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alt et al 2009</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2009</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Bone dowel pilot RCT</td>
<td>Burkus et al 2001</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2002</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2005</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Bone dowel pivotal RCT</td>
<td>Burkus et al 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus 2005</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>INTER FIX™ PLIF RCT</td>
<td>Alexander et al 2002</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Haid et al 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>CORNERSTONE® pilot RCT</td>
<td>Baskin et al 2003</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MASTERGRAFT® pilot RCT</td>
<td>Bae et al 2007</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MAVERICK™ Disc pivotal RCT (not ICBG)</td>
<td>Gornet et al 2007</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gornet et al 2011</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>BCP US RCT</td>
<td>Boden et al 2002</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>BCP Canada</td>
<td>Assiri 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Trial</td>
<td>Publications</td>
<td>Full paper</td>
<td>Abstract only</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>RCT</td>
<td>Alexander et al 2007</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abraham et al 2008</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abraham et al 2010</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AMPLIFY™ RCT</td>
<td>Glassman et al 2005</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dimar et al 2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dimar et al 2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glassman et al 2007</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dimar et al 2009</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>McInnis 2010</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>INTER FIX™ ALIF pilot RCT</td>
<td>Not published</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TELAMON PEEK instrumented PLIF pilot Single-arm trial</td>
<td>Not published</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRM Two-level pilot Single-arm trial</td>
<td>Not published</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCP Mexico pilot Single-arm trial</td>
<td>Not published</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CORNERSTONE® ACDF pivotal RCT</td>
<td>Not published (terminated after recruitment of 3 patients)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix table 2: Details of published studies reporting on the safety of rhBMP-2 surgery

<table>
<thead>
<tr>
<th>Comparator</th>
<th>Surgical approach</th>
<th>Author</th>
<th>rhBMP-2 No. Patients</th>
<th>Comparator</th>
<th>Adverse events reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICBG</td>
<td>Cervical</td>
<td>Butterman (2008) 24</td>
<td>30</td>
<td>RhBMP-2</td>
<td>Wound complications</td>
</tr>
<tr>
<td>ICBG</td>
<td>Cervical</td>
<td>Crawford (2009) 25</td>
<td>41</td>
<td>RhBMP-2</td>
<td>Wound complications</td>
</tr>
<tr>
<td>ICBG</td>
<td>Anterior lumbar</td>
<td>Carragee (2012) 20</td>
<td>239</td>
<td>RhBMP-2</td>
<td>Retrograde ejaculation, infection, urinary retention</td>
</tr>
<tr>
<td>ICBG</td>
<td>Anterior lumbar</td>
<td>Pradhan (2006) 26</td>
<td>9</td>
<td>RhBMP-2</td>
<td>Wound complications</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Anand (2006) 27</td>
<td>85</td>
<td>RhBMP-2</td>
<td>Neurologic events, hardware failure, leg pain/radiculitis</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Glassman (2007) 28</td>
<td>91</td>
<td>RhBMP-2</td>
<td>Hardware failure</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Lee (2010) 29</td>
<td>86</td>
<td>RhBMP-2</td>
<td>Infection, wound complications</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Mannion (2010) 30</td>
<td>17</td>
<td>RhBMP-2</td>
<td>Heterotopic bone formation</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Rihn (2009) 32</td>
<td>86</td>
<td>RhBMP-2</td>
<td>Heterotopic bone formation, osteolysis, infection, hardware failure, wound complication, leg pain/radiculitis, inflammatory cyst formation</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Singh (2006) 33</td>
<td>41</td>
<td>RhBMP-2</td>
<td>Heterotopic bone formation</td>
</tr>
<tr>
<td>Allograft</td>
<td>Cervical</td>
<td>Lu (2010) 34</td>
<td>100</td>
<td>RhBMP-2</td>
<td>Dysphagia</td>
</tr>
<tr>
<td>Allograft</td>
<td>Anterior lumbar</td>
<td>Slosar (2007) 35</td>
<td>45</td>
<td>RhBMP-2</td>
<td>Heterotopic bone formation, wound complication</td>
</tr>
<tr>
<td>Allograft</td>
<td>Various</td>
<td>Vaidya (2007b) 36</td>
<td>36</td>
<td>RhBMP-2</td>
<td>Hardware failure, dysphagia</td>
</tr>
<tr>
<td>Allograft / autograft</td>
<td>Posterior lumbar</td>
<td>Mindea (2009) 37</td>
<td>35</td>
<td>RhBMP-2</td>
<td>Leg pain/radiculitis</td>
</tr>
<tr>
<td>Autologous bone</td>
<td>Posterior lumbar</td>
<td>Gray (2010) 38</td>
<td>82</td>
<td>RhBMP-2</td>
<td>Heterotopic bone formation, osteolysis, hardware failure, leg pain/radiculitis</td>
</tr>
<tr>
<td>Autologous bone</td>
<td>Posterior lumbar</td>
<td>Joseph (2007) 39</td>
<td></td>
<td>RhBMP-2</td>
<td>Heterotopic bone formation</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Cervical</td>
<td>Longley (2009) 40</td>
<td>84</td>
<td>RhBMP-2</td>
<td>Dysphagia</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Cervical</td>
<td>Smucker (2006) 41</td>
<td>69</td>
<td>RhBMP-2</td>
<td>Dysphagia</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Cervical</td>
<td>Vaidya (2007a) 42</td>
<td>22</td>
<td>RhBMP-2</td>
<td>Dysphagia</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Cervical</td>
<td>Williams (2011) 43</td>
<td>4532</td>
<td>RhBMP-2</td>
<td>Dysphagia, wound complications</td>
</tr>
<tr>
<td>Group</td>
<td>Region</td>
<td>Study</td>
<td>N1</td>
<td>N2</td>
<td>Adverse Outcomes</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>-----</td>
<td>-----</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Cervical</td>
<td>Xu (2011) (44)</td>
<td>48</td>
<td>156</td>
<td>Heterotopic bone formation, infection, hardware failure, dysphagia, recurrent laryngeal neck palsy, wound complications</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Posterior lumbar</td>
<td>Rowan (2011) (46)</td>
<td>64</td>
<td>40</td>
<td>Leg pain/radiculitis</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Posterior lumbar</td>
<td>Williams (2011) (43)</td>
<td>6049</td>
<td>18267</td>
<td>Wound complications, inflammatory cyst formation</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Various</td>
<td>Cahill (2009) (2)</td>
<td>17626</td>
<td>53026</td>
<td>Dysphagia, wound complications</td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Various</td>
<td>Williams (2011) (43)</td>
<td>11933</td>
<td>43929</td>
<td>Wound complications</td>
</tr>
<tr>
<td>Bone marrow aspirate</td>
<td>Anterior lumbar</td>
<td>Gerszten (2011) (47)</td>
<td>45</td>
<td>54</td>
<td>Heterotopic bone formation, neurologic events, hardware failure, wound complications, vascular events</td>
</tr>
<tr>
<td>Bone marrow aspirate</td>
<td>Posterior lumbar</td>
<td>Taghavi (2010) (48)</td>
<td>24</td>
<td>38</td>
<td>Hardware failure</td>
</tr>
<tr>
<td>Silicated Calcium Phosphate</td>
<td>Posterior lumbar</td>
<td>Pimenta (2011) (49)</td>
<td>15</td>
<td>15</td>
<td>Heterotopic bone formation</td>
</tr>
<tr>
<td>Mixed</td>
<td>Cervical</td>
<td>Hiremath (2009) (50)</td>
<td>16</td>
<td>67</td>
<td>Neurologic events, hardware failure, dysphagia, recurrent neck palsy, wound complications</td>
</tr>
</tbody>
</table>
References identified in database search: 6,807

Additional records identified from bibliography hand-searches and electronic update searches: 103

Excluded during de-duplication: 4,073

Screening of titles and abstracts: 2,837

Excluded during screening of titles and abstracts: 2,521

Materials provided by Medtronic:
- IPD for 17 trials: Efficacy and safety data
- Protocols and publications for the trials

Full-text screening: 311

Excluded Medtronic trials:
- 4 single-arm trials
- 1 trial with no ICBG arm
- 1 trial of 3 patients

Full text could not be retrieved: 3

Excluded during full-text screening due to unmet inclusion criteria: 233

Effectiveness analysis
- IPD from 11 Medtronic RCTs
- IPD from 1 additional RCT (Glassman 2008)
- 1 further RCT (Shapiro 2005) eligible but IPD not provided

Safety analyses
- IPD from 11 Medtronic RCTs
- 43 publications reporting on 35 additional controlled adverse events studies (not IPD)
### Estimates with 95% confidence intervals

<table>
<thead>
<tr>
<th>Trial</th>
<th>rhBMP Fused/Total</th>
<th>ICBG Fused/Total</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT Cage Open</td>
<td>123/130</td>
<td>106/119</td>
<td>1.06 (0.99, 1.15)</td>
</tr>
<tr>
<td>Bone Dowel Pilot</td>
<td>24/24</td>
<td>13/19</td>
<td>1.46 (1.08, 1.98)</td>
</tr>
<tr>
<td>Bone Dowel Pivotal</td>
<td>43/44</td>
<td>22/27</td>
<td>1.20 (1.00, 1.44)</td>
</tr>
<tr>
<td>Interfix ALIF Pilot</td>
<td>16/20</td>
<td>13/14</td>
<td>0.86 (0.66, 1.12)</td>
</tr>
<tr>
<td>Cornerstone Pilot</td>
<td>11/11</td>
<td>11/11</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Interfix PLIF</td>
<td>24/26</td>
<td>21/27</td>
<td>1.19 (0.94, 1.49)</td>
</tr>
<tr>
<td>Mastergraft Pilot</td>
<td>18/19</td>
<td>14/20</td>
<td>1.35 (1.00, 1.84)</td>
</tr>
<tr>
<td>BCP US</td>
<td>7/10</td>
<td>3/3</td>
<td>0.70 (0.47, 1.05)</td>
</tr>
<tr>
<td>BCP Canada</td>
<td>86/89</td>
<td>64/91</td>
<td>1.37 (1.20, 1.58)</td>
</tr>
<tr>
<td>Amplify</td>
<td>186/194</td>
<td>151/169</td>
<td>1.07 (1.01, 1.14)</td>
</tr>
</tbody>
</table>

**Pooled**

1.14 (1.03, 1.25)

Relative risk of fusion (rhBMP-2 vs ICBG)

ICBG better ← → rhBMP-2 better
Oswestry Disability Index

Successful fusion
Estimates with 95% confidence intervals

<table>
<thead>
<tr>
<th>Type of adverse event</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arm and neck pain</td>
<td>0.65 (0.06, 7.24)</td>
</tr>
<tr>
<td>Arthritis/bursitis</td>
<td>2.84 (0.29, 27.51)</td>
</tr>
<tr>
<td>Back and leg pain</td>
<td>1.92 (1.13, 3.25)</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>0.99 (0.68, 1.44)</td>
</tr>
<tr>
<td>Dural injury</td>
<td>1.49 (0.62, 3.62)</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>0.30 (0.02, 3.59)</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>1.04 (0.73, 1.48)</td>
</tr>
<tr>
<td>Implant</td>
<td>2.11 (0.73, 6.07)</td>
</tr>
<tr>
<td>Infection</td>
<td>1.23 (0.82, 1.85)</td>
</tr>
<tr>
<td>Neurological</td>
<td>1.73 (0.96, 3.12)</td>
</tr>
<tr>
<td>Other pain</td>
<td>1.68 (0.69, 4.07)</td>
</tr>
<tr>
<td>Respiratory</td>
<td>1.27 (0.64, 2.51)</td>
</tr>
<tr>
<td>Retrograde ejaculation</td>
<td>3.00 (0.31, 29.15)</td>
</tr>
<tr>
<td>Spinal</td>
<td>0.53 (0.15, 1.85)</td>
</tr>
<tr>
<td>Trauma</td>
<td>0.93 (0.34, 2.55)</td>
</tr>
<tr>
<td>Urogenital</td>
<td>1.38 (0.86, 2.21)</td>
</tr>
<tr>
<td>Vascular</td>
<td>1.94 (0.53, 7.18)</td>
</tr>
<tr>
<td>Wound complication</td>
<td>1.76 (0.39, 8.06)</td>
</tr>
</tbody>
</table>

Odds ratio of adverse events (rhBMP–2 vs ICBG)

More common with ICBG ᢱ More common with rhBMP–2
### Estimates with 95% confidence intervals

<table>
<thead>
<tr>
<th>Trial</th>
<th>rhBMP Cancers/Total</th>
<th>rCBG Cancers/Total</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT Cage Open</td>
<td>27/43</td>
<td>1/137</td>
<td>1.92 (0.18, 20.89)</td>
</tr>
<tr>
<td>Bare Dowel Pilot</td>
<td>1/24</td>
<td>0/22</td>
<td>2.76 (0.12, 64.23)</td>
</tr>
<tr>
<td>Bare Dowel Posterior</td>
<td>1/55</td>
<td>0/30</td>
<td>1.65 (0.07, 39.25)</td>
</tr>
<tr>
<td>BCP Canada</td>
<td>108</td>
<td>2/26</td>
<td>0.51 (0.05, 5.48)</td>
</tr>
<tr>
<td>Amplify</td>
<td>132/40</td>
<td>5/226</td>
<td>2.25 (0.81, 6.29)</td>
</tr>
<tr>
<td>Pooled</td>
<td></td>
<td></td>
<td>1.84 (0.81, 4.16)</td>
</tr>
</tbody>
</table>

**Relative risk of cancer (rhBMP-2 vs rCBG)**

mBMP-2 better ← ➝ rCBG better
**Estimates with 95% confidence intervals**

<table>
<thead>
<tr>
<th>Study</th>
<th>Comparator</th>
<th>Surgery</th>
<th>RhBMP-2 events/total</th>
<th>Comparator events/total</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rihn</td>
<td>ICBG</td>
<td>TLIF</td>
<td>246</td>
<td>0/33</td>
<td>1.94 (0.10, 39.39)</td>
</tr>
<tr>
<td>Mannan</td>
<td>ICBG</td>
<td>PLF/TUF</td>
<td>2/17</td>
<td>0/19</td>
<td>5.57 (0.29, 109.27)</td>
</tr>
<tr>
<td>Gray</td>
<td>Autologous bone</td>
<td>PLF</td>
<td>4/92</td>
<td>10/99</td>
<td>2.05 (1.15, 3.63)</td>
</tr>
<tr>
<td>Joseph</td>
<td>Autologous bone</td>
<td>PLF/TUF</td>
<td>5/23</td>
<td>1/10</td>
<td>2.17 (0.29, 18.30)</td>
</tr>
<tr>
<td>Pimenta</td>
<td>SCP</td>
<td>LIF</td>
<td>1/15</td>
<td>0/15</td>
<td>3.00 (0.13, 69.09)</td>
</tr>
</tbody>
</table>

More common with comparator ←→ More common with rhBMP-2

Relative risk of heterotopic bone formation.
### Estimates with 95% confidence intervals

<table>
<thead>
<tr>
<th>Study</th>
<th>Comparator</th>
<th>Surgery</th>
<th>rhBMP-2 events/total</th>
<th>Comparator events/total</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARM</td>
<td>ICBG</td>
<td>TLIF</td>
<td>13865</td>
<td>103</td>
<td>4.60 (0.82, 34.03)</td>
</tr>
<tr>
<td>Gray</td>
<td>Autologous bone</td>
<td>PLIF</td>
<td>2402</td>
<td>199</td>
<td>2.39 (0.12, 49.69)</td>
</tr>
<tr>
<td>Minnea</td>
<td>Autologous bone</td>
<td>TLIF</td>
<td>4036</td>
<td>368</td>
<td>2.15 (0.19, 36.31)</td>
</tr>
<tr>
<td>Rowan</td>
<td>No rhBMP-2</td>
<td>TLIF</td>
<td>11864</td>
<td>5480</td>
<td>2.29 (0.68, 7.72)</td>
</tr>
</tbody>
</table>

More common with comparator ← → More common with rhBMP-2
February 04, 2013

Mark C. Simmonds Dr, MA PhD
Centre for Reviews and Dissemination
University of York
York, YO10 5DD
GBR

REF: M12-2603

Dear Dr. Simmonds:

Thank you for submitting your revised manuscript, “The safety and effectiveness of recombinant human bone morphogenetic protein-2 (rhBMP-2) for spinal fusion: An individual participant data meta-analysis,” to Annals of Internal Medicine. Our team of physician and statistical editors reviewed the manuscript and found it very responsive to the requests we sent to you after our initial review. I apologize for the delay in getting back to you, but the holidays and travel abroad of the lead statistical editor contributed to difficulty in scheduling a time when all members of the team could gather to discuss the revised manuscript.

There are a few remaining issues that we must ask you to address before publication. Most of these relate to article format, but the first few issues are more substantive.

First, thank you for explaining in your cover letter how the data you used in this analysis compares to the data Carragee and colleagues reported in reference 5. Please include a similar description in your Discussion. Readers will likely be interested in comparing these different reviews and will need a clear understanding of the data that each used. If feasible, you might also expand Appendix Table 1 to indicate whether Carragee and colleagues considered data from each of the listed trials in their analysis.

Second, we continue to be concerned about the description of missing data and the methods used to address this issue. With respect to the description of missing data, please provide additional information on the completeness of follow-up. We offer two suggestions for how you might achieve this. You could develop an appendix table that provides the number of observations per visit in each treatment group in each trial. This would describe the extent of missing data quite completely. Alternatively (or in addition), you could add the proportion of patients who attended the last scheduled visit in the trial to one of the existing descriptive tables of included studies. With respect to methods to address the impact of missing data, you currently use last observation carried forward (LOCF) as sensitivity analysis (compared to complete case analysis). Given the well-described limitations of LOCF methods, Annals now routinely asks authors to avoid LOCF methods and to instead use methods to handle missing data that that are less likely to be biased and that better incorporate statistical
uncertainty due to missing data. While we agree with using complete case analysis as the main analytical approach, please provide an alternative method for handling missing data in a sensitivity analysis that addresses this issue more in accordance with current biostatistical methods (see, e.g., Little et al., NEJM 2012;367:1355-60).

Third, we ask that you make the following editorial changes to the tables and figures to make the body of the report more manageable while retaining detail for those readers who wish to delve deeper:
- Figure 1: OK as is
- Figure 2: Add the number of patients in each group at each time point
- Figure 3: Delete this figure and incorporate results into the text
- Figure 4: OK as is
- Figure 5: Move this figure to an appendix
- Figure 6: Delete this figure and incorporate results into the text
- Figure 7: Consolidate the information provided in Figure 7 and Table 2 into a single figure or table
- Figure 8: Move this figure to an appendix
- Figure 9: Add number of patients and number of events per group at each time point.
- Figures 10, 11, and 12: Consolidate into a single figure

Fourth, the Abstract would benefit from some reorganization and inclusion of greater detail. We suggest the following revised Abstract for your approval:

Background: Recombinant human bone morphogenetic protein-2 (rhBMP-2) promotes fusion in spinal surgery, but its safety is debated.

Purpose: To evaluate the effectiveness and safety of rhBMP-2.

Data Sources: Individual patient data (IPD) from Medtronic-funded trials through the Yale Open Access Data Project; CENTRAL, MEDLINE, EMBASE, Science Citation Index, and ClinicalTrials.gov through June 2012; and a call for evidence.

Study Selection: Randomized controlled trials of rhBMP-2 versus iliac crest bone graft (ICBG) in spinal fusion surgery for degenerative disc disease and related conditions. Observational studies in similar populations for adverse events.

Data Extraction: Primary outcomes included pain, Oswestry disability index (ODI), SF-36, fusion and adverse events.

Data Synthesis: Of 17 Medtronic trials, 11 were eligible and we obtained IPD from one of two identified non-Medtronic trials (1411 Medtronic and 106 non-Medtronic participants). 35 observational studies provided additional adverse event data. At two years, ODI scores were 3.5% lower with rhBMP-2 than with ICBG (95% CI: 0.49 to 6.47) and radiographic fusion was 14% higher (95% CI: 33% to 25%). At or shortly after surgery, pain was more common with rhBMP-2 (OR 1.78, 95%CI: 1.06, 2.95) and evidence for implant-related (OR 1.87: 95% CI:0.67, 5.21) or neurological events (OR 1.58, 95%CI 0.88,2.83) was inconclusive. Cancer was more common following rhBMP-2 than ICBG (RR 1.84, 95% CI: 0.81 to 4.16), but few events prohibited definite conclusions.

Limitations: Observational studies were diverse and at risk of bias.
Conclusions: At 2 years compared to ICBG, rhBMP-2 increases fusion rates and reduces pain by a clinically insignificant amount. There is evidence that rhBMP-2 increases early post-surgical pain, and inconclusive evidence regarding implant-related or neurological events and cancer incidence.

(Words: 272)

Fifth, please address the requests provided on the attached technical/format review.

Please send your revised manuscript and cover letter to us within four weeks of receiving this letter. Submit your revised manuscript and cover letter at https://www.acponline.org/authors/ by clicking the "Revise Paper" link listed below your "Tasks."

Please keep all editorial correspondence confidential, and refrain from sharing either the correspondence itself or the essence of its content with individuals who are not your collaborators. Doing so helps ensure we can offer you advice that is in the best interests of your paper, without concern for how it might be considered or used by others.

We look forward to receiving your revised manuscript.

Sincerely,

Christine Laine, MD, MPH

Editor
10. Authors: Response Letter 2 (4 March 2013)

To the editor,
On behalf of my co-authors I have re-submitted the manuscript “The safety and effectiveness of recombinant human bone morphogenetic protein-2 (rhBMP-2) for spinal fusion: An individual patient data meta-analysis” with revisions as requested by the reviewers, for consideration by Annals of Internal Medicine.
Below is a full response to all reviewers’ comments.
Yours faithfully,
Dr Mark Simmonds,
on behalf of the authors.

Response to Editors’ Comments

We would like to thank the editors for their further comments on our paper “The safety and efficacy of recombinant bone morphogenetic protein-2 (rhBMP-2) for spinal fusion”, and particularly for taking the time to revise the abstract. Our replies to your comments, detailing the changes made are given below.

First, thank you for explaining in your cover letter how the data you used in this analysis compares to the data Carragee and colleagues reported in reference 5. Please include a similar description in your Discussion... If feasible, you might also expand Appendix Table 1 to indicate whether Carragee and colleagues considered data from each of the listed trials in their analysis.
We have added a short description of how our review differs from Carragee to the discussion (para. 7). We were unable to extend appendix table 1 as we were not able to determine exactly which references were used in the Carragee review.

Second, we continue to be concerned about the description of missing data and the methods used to address this issue... You could develop an appendix table that provides the number of observations per visit in each treatment group in each trial. .... Alternatively (or in addition), you could add the proportion of patients who attended the last scheduled visit in the trial to one of the existing descriptive tables of included studies.
The new appendix table 2 summarises the numbers of missing observations across trials by time point and treatment group for each outcome.
...While we agree with using complete case analysis as the main analytical approach, please provide an alternative method for handling missing data in a sensitivity analysis that addresses this issue more in accordance with current biostatistical methods.
We have replaced the last observation carried forward analyses with multiple imputation analyses. The imputations were based on predictions from regression models of outcome against the most recent non-missing outcome, stratified by trial. We describe this approach briefly in the “Sensitivity analysis” section of the methods appendix, and have altered the main text where necessary. The results were almost identical to the complete case analysis, so no further changes to text or figures have been made.

Third, we ask that you make the following editorial changes to the tables and figures to make the body of the report more manageable while retaining detail for those readers who wish to delve deeper:
• Figure 2: Add the number of patients in each group at each time point
We think that doing this would cause unnecessary clutter to this figure. There are four outcomes and five time points, so this would mean adding 40 numbers to the figure. We note that the relevant numbers are available from appendix table 2.

- **Figure 3: Delete this figure and incorporate results into the text**
  We have done this. See amended text in the fusion results section.

- **Figure 5: Move this figure to an appendix**
  We have done this. This is now appendix figure 3.

- **Figure 6: Delete this figure and incorporate results into the text**
  We think that the results of this figure are difficult to describe in text alone. We have instead moved this figure to the appendix. It is now appendix figure 6, and we have modified the text in the paper to accommodate this.

- **Figure 7: Consolidate the information provided in Figure 7 and Table 2 into a single figure or table**
  We have removed table 2 and added the relevant numbers to the new figure 4. In order to retain the data on adverse events over all time points that were in table 2 we have added a new forest plot of adverse events over all time points as appendix figure 7, with some additional text in the paper (Results: Safety).

- **Figure 8: Move this figure to an appendix**
  We have done this, now appendix figure 8.

- **Figure 9: Add number of patients and number of events per group at each time point.**
  (Note: this is now figure 5) This is an analysis of cancer, so is not done by time point but uses the total number of cancers during the whole two-year follow-up. The relevant numbers are already in the figure. We have added a footnote (see end of the figure legends in main text) to clarify this.

- **Figures 10, 11, and 12: Consolidate into a single figure**
  We have done this, now figure 6

**Fourth, the Abstract would benefit from some reorganization and inclusion of greater detail.**
We suggest the following revised Abstract for your approval...
We thank the editors for amending the abstract. We have generally replaced our abstract with your suggestion. Some edits were required to incorporate funding source and PROSPERO registration details within the word limit.

**Fifth, please address the requests provided on the attached technical/format review.**
We have done this, with one exception:
*Please use generic names in place of brand names where applicable.*
We have kept our use of “INFUSE” and “AMPLIFY” as these brand names are the simplest way of distinguishing between different rhBMP-2 preparations.
11. **Authors: Revised Manuscript 2 (4 March 2013)**

The safety and effectiveness of recombinant human bone morphogenetic protein-2 (rhBMP-2) for spinal fusion

An individual participant data meta-analysis

*Running title:*

*Bone morphogenetic protein-2 in spinal fusion*

Mark C Simmonds MA PhD
Jennifer VE Brown BA MSc
Morag K Heirs MSc MA
Julian PT Higgins BA PhD
Richard J Mannion PhD FRCS
Mark A Rodgers BSc MSc
Lesley A Stewart BSc MSc PhD*

*Centre for Reviews and Dissemination, University of York, UK*

*Corresponding author*

Centre for Reviews and Dissemination
University of York, UK
YO10 5DD
Tel: 01904 321040
Email: lesley.stewart@york.ac.uk

Word count: 4,616 words
Abstract

Background: Recombinant human bone morphogenetic protein-2 (rhBMP-2) is used widely to promote fusion in spinal surgery, but its safety has been questioned.

Purpose: To evaluate the effectiveness and safety of rhBMP-2.

Data Sources: Individual participant data (IPD) obtained from sponsor or investigator. Data extracted from study publications identified by systematic bibliographic searches.

Study Selection: Randomized controlled trials of rhBMP-2 versus iliac crest bone graft (ICBG) in spinal fusion surgery for degenerative disc disease and related conditions. Observational studies in similar populations for adverse events.

Data Extraction: IPD for 11 eligible of 17 provided Medtronic-sponsored trials and one of two further eligible trials were included (1411 and 106 participants, respectively). Additional aggregate adverse event data extracted from 35 published observational studies.

Data Synthesis: Primary outcomes were pain (Oswestry disability index (ODI), SF-36), fusion and adverse events. At two years, ODI scores were 3.5% lower with rhBMP-2 than with ICBG (95% CI: 0.49 to 6.47) and radiographic fusion was 14% higher (95% CI: 33% to 25%). At or shortly after surgery, pain was more common with rhBMP-2 (OR 1.78, 95%CI: 1.06, 2.95). Cancer was more common following rhBMP-2 than ICBG (RR 1.84, 95% CI: 0.81 to 4.16), but few events prohibited definite conclusions.

Limitations: Observational studies were diverse and at risk of bias.

Conclusions: At 2 years compared to ICBG, rhBMP-2 increases fusion rates, reduces pain by an amount that is not clinically significant, and increases early post-surgical pain. There is inconclusive evidence of increased cancer incidence.
Primary funding source: This review was funded by a competitively awarded grant from the Yale University Open Data Access project. A protocol was lodged on the PROSEPRO database (CRD42012001907).

(Words: 274)

**Introduction**

Recombinant human bone morphogenetic protein-2 (rhBMP-2) is widely used as an alternative to iliac crest bone graft (ICBG) to promote fusion in spinal surgery (1) (2). Following Food and Drug Administration approval of rhBMP-2 in anterior lumbar interbody fusion surgery (3), numbers of spinal procedures using rhBMP-2 have grown rapidly, including off-label indications (2) (4). Recently, concerns have grown about adverse events associated with rhBMP-2, with a review of publicly available data suggesting that the risk of adverse events is 10 to 50 times higher than reported in trial publications (5).

The Yale University Open Data Access (YODA) project invited the manufacturer, Medtronic, to provide full data from all their trials of rhBMP-2, to allow independent re-analysis. YODA subsequently invited proposals to undertake independent evaluation, and funded the Centre for Reviews and Dissemination and one other group to do so. This provided an opportunity to undertake an individual participant data (IPD) meta-analysis, which is regarded as a ‘gold standard’ approach to synthesis, and allows more in-depth analysis than is possible in a systematic review of the published literature.

We embedded our meta-analysis of IPD within a wider systematic review and sought to examine all available research evidence. This included the IPD made available by Medtronic from all their trials of rhBMP-2 (randomised and non-randomised), as well as from one additional RCT. To investigate the safety of rhBMP-2 more widely, we sought all observational studies of rhBMP-2 in spinal surgery that reported adverse events.

**Methods**

**Eligibility, search, data collection and critical assessment**
Methods were, in advance of detailed knowledge of the IPD to be provided, pre-specified in a protocol, registered in PROSPERO in February 2012 (registration number CRD42012001907) which was also lodged with the Yale University Open Data Access team at the outset of the project. Eligible for inclusion in our principal analysis were all randomised controlled trials comparing rhBMP-2 with ICBG in spinal fusion surgery, irrespective of spinal level or surgical approach. We included trials of the licensed INFUSE® rhBMP-2 formulation (rhBMP-2 concentration: 1.5mg/cc) and unlicensed AMPLIFY ™ (rhBMP-2 concentration: 2mg/cc) and Biphasic Calcium Phosphate (BCP) formulations. For supporting analyses of safety, all studies comparing rhBMP-2 with any other spinal fusion technique and reporting adverse events were eligible for inclusion.

As our re-analysis of the Medtronic data was done in the context of a full systematic review, we performed a systematic literature search to identify all eligible studies, particularly those not provided by Medtronic. This involved searching multiple bibliographic databases including CENTRAL, MEDLINE, EMBASE and Science Citation Index in January 2012, and automated ‘current awareness’ searches up to June 2012 (see web appendix for detailed search strategies). We also searched clinicaltrials.gov to identify ongoing or unpublished randomised trials and published a call for evidence (see web appendix for details).

The Yale University Open Data Access project team provided us with IPD from 17 trials undertaken by Medtronic. Eleven of these were RCTs comparing rhBMP-2 to ICBG surgery and were eligible for inclusion in our principal evaluation of effectiveness. Of the others, four were single-arm trials of rhBMP-2 and one (19) used a different comparator. These were included in our supporting consideration of adverse events. One trial was stopped early after recruiting only three patients and not considered further.

One researcher collated IPD from across the multiple SAS data files provided by Medtronic, and a second researcher checked the process. The IPD were checked for completeness, internal consistency, improbable values and balance of patient characteristics across treatment arms (see methods appendix for details). The study selection and data extraction from published reports was undertaken by three researchers independently working in pairs, with any differences resolved by discussion or by referring to the third researcher. We assessed risk of bias in the RCTs using the
Cochrane Collaboration’s ‘risk of bias’ tool, and using a modified form of the Newcastle-Ottawa scale for non-randomised studies (7, 8) (see methods appendix for details). Risk of bias was assessed by at least two researchers independently, with any differences resolved by discussion.

**Effectiveness analysis**

We pre-specified our primary outcomes as those that we considered likely to be important to patients. The Oswestry Disability Index (ODI) and Neck Disability Index (NDI) for cervical spinal surgery measure, respectively, lower back and neck pain on a scale from 0 (no pain) to 100% (extreme pain). The SF-36 Physical Component Score (SF-36 PCS) assesses both pain and physical function on a scale from 0 (worst) to 100% (best). Medtronic also provided data on back and leg pain, which were both measured and supplied on a scale from 0 (no pain) to 20 (extreme pain). Spinal fusion was supplied as success or failure as defined radiographically within the trials according to Medtronic criteria that required all of the following: evidence of bridging trabeculae, no evidence of motion (<3mm difference in translation, less than 5 difference in angular motion) and no evidence of radiolucency. We analysed these outcomes at 6 weeks, 3, 6, 12 and 24 months after surgery. We did not analyse the limited data that were available for longer follow-up times because these were available for only a minority of the participants.

We also considered four secondary outcomes: duration of hospital stay, operating time, successful return to work / usual activity and use of pain-relief medication.

**Statistical methods**

Full details of statistical methods and analyses are provided in the methods appendix. In the analyses of effectiveness for continuously distributed outcomes (such as ODI) we calculated mean differences between treatment arms in the change in score from pre-operative values. For dichotomous outcomes (such as successful fusion) we calculated relative risks. These were calculated separately for every trial at every time point. We then used standard random-effects meta-analytic techniques (9) (10) to combine effect estimates across trials. Separate meta-analyses were performed for each of the specified time points. Linear and logistic random-effects regression models were used to combine all data from all trials in “one-stage” meta-analyses as sensitivity analyses. (11, 12) To explore the
influence of missing observations we used multiple imputation to impute data which were missing in sensitivity analyses.

Heterogeneity was assessed in all meta-analyses using the Higgins’ $I^2$ statistic (13) and Cochran’s Q test. We examined whether effects varied according to the type of spinal surgery, or by rhBMP-2 formulation (INFUSE® or AMPLIFY™) using subgroup analysis (stratified by trial). We investigated whether patient-level factors (age, sex, smoking, alcohol consumption, body mass index, diabetic status, and history of spinal surgery for back pain) were associated with the effectiveness of rhBMP-2 surgery using a one-stage random-effects regression model (11) including interaction terms between patient-level factors and treatment.

**Safety analysis**

For evaluation of safety, we examined numbers of adverse events, including cancer, provided in the IPD according to the classifications of adverse events produced by Medtronic. Case report forms did not form part of the Medtronic submission, but summaries of cases were provided for the majority of patients in the clinical study reports. A full clinical assessment of these narratives was beyond the scope of this time-limited project, but checks in one trial (Interfix PLIF) showed that the Medtronic classifications of events seemed appropriate. To our knowledge these data represent all adverse events that occurred during the follow-up periods of these trials. Because the numbers of specific adverse events in most trials were generally small, we used one-stage random-effects logistic regression meta-analysis models (12) to analyse these data (see methods appendix for model details). Results of these analyses are presented as odds ratios.

From other published studies that compared spinal fusion surgery using rhBMP-2 with ICBG and other comparators, we extracted data on numbers of adverse events as specified in our protocol. This was restricted to comparative studies including more than ten adult participants.

**Funding source**

This review was funded by the Yale University Open Data Access project who provided the individual patient data for, and other materials relating to, the Medtronic trials. They were not involved in the analyses of these data or in the production of this paper. There was no direct contact with Medtronic or with the other evaluation team. Our full report to the Yale University Open Data
Access project is publicly available and examines a wide range of issues around the effectiveness and safety of rhBMP-2 in spinal surgery. It also includes an examination of the reliability of the literature published in this field (which will be submitted for publication separately) (14) (15).

Results

Figure 1 shows the result of our systematic search. Other than the trials already supplied by Medtronic, we identified two eligible randomised trials not conducted by Medtronic which compared rhBMP-2 with ICBG surgery. We requested IPD from the authors, and obtained these from one (Glassman)(16). IPD from the other trial in 40 patients undergoing single-level bilateral posterior lateral interbody fusion were not made available by the trial author.(17) This unavailable trial reported 100% fusion in both groups and no difference in back pain or ODI at 12 months, and therefore could not contribute further to our analyses. One ongoing trial was identified, for which recruitment had been suspended, but as it had not been closed it could not be included.(18)

Our IPD meta-analysis was therefore based on data from 1,411 patients included in 11 Medtronic RCTs that compared rhBMP-2 with ICBG surgery plus data from 106 participants in one RCT not sponsored by Medtronic (Glassman) (16). Data from one Medtronic RCT (19) were not included because the comparator arm was not ICBG surgery. Table 1 summarises the 12 included trials. Details of publications associated with each the Medtronic trials are provided in Appendix table 1. Note however that our analyses used the supplied IPD and not data reported in these publications. For all the included trials, data on all pain outcomes were available at all time points from six weeks to 24 months. Similarly data on spinal fusion were available for all trials (except the LT-Cage pilot trial and the Glassman trial) at all times from six months onwards. Data were not available from all trial participants at all time points. Appendix table 2 summarises the levels of missing data for the main pain and fusion outcomes. At 24 months after surgery outcome data were not available for around 15% of participants.

As the trial protocols for all the Medtronic trials were very similar, our assessment of risk of bias was the same for all trials. We found that the randomisation and allocation concealment procedures were adequate for all trials. Neither patients nor physicians were blinded to the treatment received, and all
pain and function outcomes were patient assessed, so there was a potential for bias in these outcomes.

Assessment of successful fusion was performed blind to treatment received, minimising bias.

**Effectiveness**

**Pain**

Figure 2 illustrates meta-analyses across the 12 RCTs for four pain outcomes (SF-36 PCS also incorporates physical function assessment). Points on the plot represent mean differences in changes in scores (from pre-operative values) line between rhBMP-2 and ICBG at each time point, with vertical lines showing the 95% confidence intervals for these. For the first three outcomes, points below the horizontal zero line indicate a benefit of rhBMP-2; for SF-36 PCS, points above the line indicate a benefit of rhBMP-2.

The use of rhBMP-2 generally achieved greater pain reduction (from pre-operative values) than did ICBG from six months after surgery onwards. At 24 months after surgery the ODI was approximately 3.5 percentage points better (Mean difference: -3.48, 95%CI: -6.47 to -0.49, $I^2 = 38\%$) and back pain was similarly better among rhBMP-2 patients, by more than one point on the 20-point scale used (Mean difference: -1.58, 95%CI: -2.65 to -0.51, $I^2 = 44\%$). SF-36 PCS was approximately two percentage points higher for patients receiving rhBMP-2 (Mean difference: 1.93, 95%CI: 0.63 to 3.22, $I^2 = 0$) at 24 months. There was no evidence of a difference in leg pain reduction between treatment groups (Mean difference: -0.59, 95%CI: -1.27 to 0.09, $I^2 = 0$). In general, improvements in pain and function using rhBMP-2 when compared with ICBG were small. Patients in both groups improved considerably over time such that the extra benefit of rhBMP-2 over ICBG surgery was small in comparison (appendix figure 1). ODI improved by approximately 26 percentage points at 24 months for ICBG patients and by 30 percentage points for rhBMP-2 patients.

Results of the one-stage linear regression meta-analysis models of pain outcomes were almost identical to the results presented above (see appendix figure 2). Similarly, using a multiple imputation approach where pain data were missing gave almost identical results.

**Fusion**

Data on fusion were not available for the LT-Cage pilot and Glassman trials (total 120 patients).

Figure 3 shows a forest plot for successful fusion 24 months after surgery, where rhBMP-2 increased
fusion rates by 14% (Relative risk: 1.14, 95%CI: 1.03 to 1.25). Increased fusion rates were also identified at six months (Relative risk: 1.23, 95%CI: 1.01 to 1.51) and 12 months after surgery (Relative risk: 1.13, 95%CI: 1.01 to 1.25). There was however substantial heterogeneity in the relative risk of successful fusion across trials at earlier time points, with I² at six, 12 and 24 months being 97%, 80% and 76% respectively. This heterogeneity is apparent in Figure 3.

Results from the one-stage logistic regression meta-analysis model for successful fusion were almost identical to the results above. Similarly, using a multiple imputation approach where fusion data were missing gave almost identical results.

Investigation by surgical approach

We investigated whether the effectiveness of rhBMP-2 varied across anterior lumbar fusion, posterior lumbar fusion, and anterior cervical fusion (see Table 1) using subgroup analysis. Appendix figure 3 shows the results of these subgroup analyses for ODI and successful fusion 24 months after surgery. A test for heterogeneity showed that there was moderate evidence of a difference between surgery types for ODI (p-value 0.065), but this was primarily due to the very large benefit of rhBMP-2 on the neck disability index observed in the single, small cervical surgery trial (23 patients). Excluding this trial resulted in no clear difference in the effectiveness of rhBMP-2 between anterior or posterior approaches (p = 0.17). There was no evidence of a difference in the relative risks of successful fusion at two years across surgery types (p = 0.88). Nor was there evidence of difference between surgery types for any other outcome.

We also considered whether a range of patient-level factors were associated with the effectiveness of rhBMP-2. There was generally no evidence of interactions between rhBMP-2 and the patient-level factors (age, sex, smoking, alcohol consumption, body mass index, diabetic status, and history of spinal surgery). One possible exception was that, for people with a history of previous spinal surgery, there was no difference in the effectiveness of rhBMP-2 and ICBG, either at reducing ODI or improving fusion rates (whereas there was an overall benefit of rhBMP-2 across all patients). Given the number of analyses performed, this result may be a chance finding. We do not present these results here. They are available in our full report.

Further outcomes
Among secondary outcomes there was no overall evidence of any difference in duration of hospital stay (Mean difference: -0.15 days, 95%CI: -1.35 to 1.06 days) or that rhBMP-2 surgery increased the probability of returning to work or usual activity earlier when compared with receiving ICBG (for example, at two years: Relative risk: 1.01, 95%CI: 0.88 to 1.17). Using rhBMP2 shortened operating times, by 21 minutes (95% CI: 15 to 27, see Appendix figure 4), from an average of 135 minutes. There was no evidence that use of pain-relief medication differed between rhBMP-2 and ICBG arms at any time (see Appendix figure 5).

We investigated the association between successful fusion and change in pain scores by comparing, for each of the four pain outcomes (ODI, SF-36 PCS, back and leg pain), the mean difference in pain 24 months after surgery with the relative risk of successful fusion 24 months after surgery. Appendix figure 6 shows this relationship, where, each point in the figure represents the results for one Medtronic trial. There is no evidence of a consistent relationship between improvements in fusion due to rhBMP-2 and improvements in pain or function. If successful fusion resulted in reduced pain we might expect trials with higher fusion rates on rhBMP-2 to show greater improvement in pain scores, but this does not appear to be the case. In particular those trials where fusion was less common in the rhBMP-2 patients (Inter Fix ALIF pilot and BCP US), still showed a benefit of rhBMP-2 in terms of improved SF-36 PCS. The BCP US trial also showed a benefit of rhBMP-2 on ODI and back pain. The apparent benefits of rhBMP-2 in pain reduction do not, therefore, appear to be due to increased fusion rates.

**Safety**

All Medtronic trials provided data on adverse events at all the specified time points and also at or shortly after (up to four weeks) surgery. Reporting of adverse events in the Glassman trial was not consistent with the Medtronic trials so this trial was not included in these analyses. Our full report to the Yale University Open Data Access project describes these data.

We note that pain was reported as an adverse event in the Medtronic IPD as well as being assessed as an effectiveness outcome using the pain scales discussed earlier. The reasons for this were not clear, but for completeness we analyse pain reported as an adverse event here, particularly because pain...
immediately after surgery was not recorded on the pain scales, only as an adverse event (recorded as present or absent).

Figure 4 illustrates the results of the one-stage meta-analyses for adverse events across these 11 Medtronic RCTs at, or shortly after (up to four weeks), surgery. Some types of adverse event were more common among rhBMP-2 patients: arthritis/bursitis, implant-related events, neurological events, other pain, retrograde ejaculation, wound complications and vascular events all had at least a 50% increase in risk of occurring. Because there were few events, confidence intervals were wide and findings inconclusive. For back and leg pain there was clear evidence of a higher incidence among rhBMP-2 patients (OR 1.92, 95% CI: 1.14 to 3.25, p = 0.004). Appendix figure 7 shows the results of the one-stage meta-analyses for adverse events over all times up to 24 months after surgery. As in the analysis in Figure 4, results were generally inconclusive.

Appendix figure 8 shows the results of meta-analyses for four key categories of adverse event (implant related, infections, neurological, any pain) across all time periods. Again, as events were uncommon the evidence is generally inconclusive. There was no evidence of a difference in the risk of adverse events from three months after surgery onwards. At, or shortly after, surgery there was inconclusive evidence that implant related events were more common in rhBMP-2 patients (Odds ratio: 1.87, 95% CI: 0.67 to 5.21), as were neurological events (Odds ratio: 1.58, 95% CI: 0.88 to 2.83). Only for pain was there clear evidence that this was more common in the rhBMP-2 patients at or shortly after surgery (Odds ratio: 1.78, 95% CI: 1.06 to 2.95, p = 0.007). This contrasts with the results seen in the analyses of ODI, SF-36 PCS and back pain in the effectiveness analyses where pain reduction was greater in the rhBMP-2 patients from three months after surgery onwards.

Cancer

Table 2 summarises the cancers observed in the 11 Medtronic RCTs, of which five observed at least one cancer case. It excludes pre-existing cancers. It includes three cases of cancer in the LT Cage open pivotal trial that were identified through extended follow-up of only the rhBMP-2 patients. However, these three cancers were not included in the quantitative analyses, as this would have been unfair against rhBMP-2 since any equivalent cancers occurring in the ICBG had not been sought. A one-stage random-effects meta-analysis model found that cancer was nearly twice as common among
rhBMP-2 patients (Relative risk: 1.98, 95% CI 0.86 to 4.54), but the 95% confidence intervals were consistent with risk in rhBMP-2 patients being anywhere from 14% lower to 454% higher. The absolute risk of cancer was low (3% in rhBMP-2 patients). A forest plot for the equivalent two-stage analysis is given in Figure 5 (Relative risk: 1.84, 95% CI: 0.81 to 4.16). The relative risk of cancer was similar across trials. In particular, the relative risk of cancer in the AMPLIFY™ trial, which used a different preparation of rhBMP-2 at a higher dose, was no greater than in those trials which used INFUSE ® (p = 0.82). We note that as well as the three additional cancers identified during additional follow up, there were also three cancers among rhBMP-2 patients in the MAVERICK™ trial and two in a single-arm Medtronic trial; these were not included in the analyses here as there was no ICBG comparator.

**Adverse events reported in the wider literature**

We also investigated adverse events data reported in publications of randomised trials with comparators other than ICBG and in non-randomised studies. We identified 35 observational studies (in 43 publications) that reported adverse effects of rhBMP-2 for at least ten adult patients. There were 14 studies of posterior lumbar fusion, five in anterior lumbar fusion, ten in cervical fusion, and eight using multiple spinal fusion procedures. These studies used a variety of spinal fusion techniques as a control, including ICBG, local bone graft, allograft and bone marrow aspirates. Other than for the Medtronic Maverick trial, we did not have IPD for any of these studies. These studies are summarised in Appendix table 3. Quality assessment found that all included a cohort of patients representative of those likely to receive treatment in practice, and exposure to treatment was clearly established. However, most made no attempt to match or control for potential confounding factors and data on the comparability of the treatment groups was generally limited or not reported. Given the diversity in the nature of these studies and their potential for bias (as a body of evidence) we did not combine studies in meta-analyses.

Despite the methodological issues there is some suggestive evidence of higher rates of particular adverse events among rhBMP-2 patients, illustrated in Figure 6. In particular, heterotopic bone formation (reported in nine studies) was more common among rhBMP-2 patients, though whether this led to any clinical consequences for those patients was unclear. Leg pain and radiculitis (five studies)
appeared to be more common among rhBMP-2 patients, as had been observed in the Medtronic trials. Osteolysis was more common among rhBMP-2 patients, but only two studies reported on this event. Dysphagia (eight studies) appeared to be more common for rhBMP-2 patients in cervical spinal surgery, although there were some inconsistencies in the results of these studies. Comer et al (20) compared four consecutive-patient cohorts undergoing ALIF with or without rhBMP-2, reporting a higher rate of retrograde ejaculation among rhBMP-2 treated patients (6.3% vs 0.9%, p=0.001).

Discussion

The analyses in this systematic review are based on data from 1411 individual participants in 12 eligible RCTs, including all trials sponsored by Medtronic (both published and unpublished) and one further trial. Data on all measured outcomes were provided. We found the randomisation procedures to be adequate in all trials, but participants were not blind to the treatment received. While assessment of some outcomes such as radiologic assessment of fusion was blinded, patient reported outcomes relating to pain were not. Follow up was reasonably complete up to our final analysis time point of 24 months. Although there are some caveats, in general we consider the body of evidence for comparative effectiveness to be strong.

We found clear evidence that rhBMP-2 when compared with ICBG, improves rates of fusion, according to Medtronic definitions which may have been stringent given that only 69% of ICBG patients achieved fusion within two years, which is lower than would be expected generally. There were high levels of inconsistency across trials, with large I² values at all time points. We also found that rhBMP-2 improves back pain and quality of life when compared with ICBG at between six months and two years after surgery. These improvements in pain are modest, and fall below the thresholds of what are regarded to be clinically meaningful improvements in pain (estimated as being between 4 and 17 percentage points for ODI, and at least 5.4 points for SF-36 PCS (21) (22)). Thus it does not seem that improved rates of fusion lead to clinically meaningful improvements in the level of long-term pain.

In general there does not appear to be a strong correlation between successful fusion and pain reduction. Those trials with higher fusion rates on rhBMP-2 did not also achieve greater pain
reduction. In the trials where fusion rates were, in fact, lower among rhBMP-2 patients there was still a greater pain reduction in rhBMP-2 patients. It therefore appears that, either rhBMP-2 surgery has an effect on pain over and above that brought about by fusion, which seems medically unlikely, or there was some bias in the interpretation of pain. As participants were neither blinded to treatment received nor to their fusion status, they may have reported exaggerated benefits on the 'new' treatment and biased pain assessment in favour of rhBMP-2.

In contrast, the analysis of adverse events reported in the IPD showed an increased risk of pain associated with rhBMP-2 in the immediate post-surgical period. Although this might appear to contradict the finding that rhBMP-2 reduces pain from six months onward, it may be that rhBMP-2 surgery leads to increased pain shortly after surgery but reduces pain in the longer term.

The IPD also indicate a possible increase in the risk of cancer associated with rhBMP-2 with nearly double the number of new cancers occurrences when compared to ICBG patients. The overall absolute risk of cancer is low in both groups, however, and so whether this increased risk is genuine is uncertain, but it is consistent with the literature suggesting a possible link between BMP and cancer. (23)

This review differs from the existing review by Carragee et al in that we had access to more extensive and more detailed data. (5). The Caragee review used aggregate data extracted from publications of industry-sponsored trials and from publically available Food and Drug Administration summaries and public meeting documents. The Food and Drug Administration materials appear to provide incomplete outcome data from a subset of trials evaluating rhBMP-2. Our review, analysed individual participant data from all Metronic sponsored trials, irrespective of whether they had been published or submitted to the Food and Drug Administration. This review included both the licensed INFUSE® preparation of rhBMP-2 and unlicensed AMPLIFY™ and Biphasic Calcium Phosphate (BCP) preparations, which use a higher dose of rhBMP-2. We found no evidence of any difference in effectiveness, safety or cancer risk between licensed and unlicensed preparations.

Data on adverse events in the wider literature raise concerns that rhBMP-2 may increase the risk of heterotopic bone formation, osteolysis, radiculitis and retrograde ejaculation. However, these are based on only the published literature, and observations are from non-randomised studies most of
which provided little information about the comparability of groups. These findings should therefore be interpreted cautiously.

Conclusions

The use of rhBMP-2 in spinal fusion surgery increases the likelihood of successful fusion at up to two years, but this does not seem to translate into clinically significant reduction in pain. The small improvements in fusion and in the level of pain reduction, which manifest after six months, also seem to come at the expense of more frequent presence of pain in the immediate post-operative period, and a possibly increased number of cancer cases. We suggest that it is important that these findings are explained clearly to patients so that they are able to make informed choices about the type of surgery they would prefer.

References

Table legends

Table 1: Summary of 12 RCTs providing IPD used in these analyses

Table 2: Incidence of cancer in the Medtronic RCTs.

Appendix table 1: Details of publications of Medtronic trials

Appendix table 2: Missing outcome data in the Medtronic trials

Appendix table 3: Details of published studies reporting on the safety of rhBMP-2 surgery

Figure legends

Figure 1: Flowchart displaying the flow of materials through the review process

Figure 2: Meta-analyses of pain outcomes at 6 weeks, 3, 6, 12 and 24 months after surgery.

Figure 3: Forest plot of relative risk of successful fusion 24 months after surgery.

Figure 4: Meta-analysis of adverse events by category at or shortly after surgery in 11 Medtronic RCTs.

Figure 5: Forest plot of cancer incidence in the Medtronic RCTs.

Figure 6: Heterotopic bone formation, radiculitis and dysphagia in non-randomised studies of rhBMP-2

Appendix figure 1: Reduction in pain scores from pre-operative results by treatment received.

Appendix figure 2: Results of the one-stage meta-analyses of pain and function outcomes at 6 weeks, 3, 6, 12 and 24 months after surgery.

Appendix figure 3: Subgroup analyses by surgical approach for Oswestry Disability Index and successful fusion.

Appendix figure 4: Mean difference in operating time between rhBMP-2 and ICBG in the Medtronic trials

Appendix figure 5: Meta-analysis of use of four types of pain-relief medication in the Medtronic trials

Appendix figure 6: Relationship between relative risk of fusion and mean difference in pain outcomes across trials 24 months after surgery.

Appendix figure 7: Meta-analysis of all adverse events by category in 11 Medtronic RCTs

Appendix figure 8: Meta-analyses of adverse events according to adverse event category.

Footnote to Figure 5

* The numbers of cancers in the figure are the total number occurring during the two years of follow-up
Acknowledgements
We thank Leah Carreon for providing the IPD for the Glassman trial, Kath Wright for performing the literature search and Charlotte Seneschall for assistance in extracting data for the safety evaluation.

Funding
Project funded by a competitive grant from the Yale University Open Data Access (YODA) project.

Contact details
Address for authors Simmonds, Brown, Heirs, Higgins, Rodgers, Stewart and for reprint requests
Centre for Reviews and Dissemination
University of York
York, UK
YO10 5DD

Address for R Mannion
Cambridge University Hospitals NHS Foundation Trust
Hills Road
Cambridge, UK
CB2 0QQ
<table>
<thead>
<tr>
<th>Trial name</th>
<th>Date of 1st patient enrollment</th>
<th>Surgical Approach</th>
<th>Intervention/Control</th>
<th>Patients randomised/enrolled</th>
<th>Patients excluded</th>
<th>Patients evaluated</th>
<th>Mean age</th>
<th>Percentage male</th>
<th>Previous spinal surgery (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT Cage pilot</td>
<td>January 1997</td>
<td>Anterior lumbar interbody fusion (ALIF)</td>
<td>rhBMP-2/ACS + LT-CAGE (1.5 mg/ml)</td>
<td>11</td>
<td>0</td>
<td>11</td>
<td>42</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>LT Cage open</td>
<td>August 1998</td>
<td>ALIF Single level open</td>
<td>rhBMP-2/ACS + LT-CAGE (1.5 mg/ml)</td>
<td>145</td>
<td>2</td>
<td>143</td>
<td>43</td>
<td>55</td>
<td>38</td>
</tr>
<tr>
<td>Bone dowel pilot</td>
<td>April 1998</td>
<td>ALIF Single level open</td>
<td>rhBMP-2/ACS + cortical bone dowels (1.5 mg/ml)</td>
<td>24</td>
<td>0</td>
<td>24</td>
<td>41</td>
<td>33</td>
<td>46</td>
</tr>
<tr>
<td>Bone dowel pivotal</td>
<td>September 2000</td>
<td>ALIF Single level open</td>
<td>rhBMP-2/ACS + cortical bone dowels (1.5 mg/ml)</td>
<td>64</td>
<td>9</td>
<td>55</td>
<td>39</td>
<td>44</td>
<td>33</td>
</tr>
<tr>
<td>Inter Fix PLIF</td>
<td>March 1999</td>
<td>Posterior lumbar interbody fusion (PLIF) Single level</td>
<td>rhBMP-2/ACS + INTER FIX cages (1.5 mg/ml)</td>
<td>35</td>
<td>1</td>
<td>34</td>
<td>46</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>Cornerstone pilot</td>
<td>September 1999</td>
<td>One or two level anterior cervical interbody fusion</td>
<td>rhBMP-2/ACS + Cornerstone implant + Atlantis plate (1.5 mg/ml)</td>
<td>18</td>
<td>0</td>
<td>18/18 had single level surgery</td>
<td>51</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td>Trial name</td>
<td>Date of 1st patient enrollment</td>
<td>Surgical Approach</td>
<td>Intervention/Control</td>
<td>Patients randomised/enrolled</td>
<td>Patients excluded</td>
<td>Patients evaluated</td>
<td>Mean age</td>
<td>Percentage male</td>
<td>Previous spinal surgery (%)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------</td>
<td>-------------------</td>
<td>----------------------------------------------------------</td>
<td>-----------------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>----------</td>
<td>-----------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Mastergraft pilot</td>
<td>April 2003</td>
<td>Posterior lumbar fusion (PLF)</td>
<td>Single level</td>
<td>ICBG/Cornerstone implant + Atlantis plate</td>
<td>15</td>
<td>0</td>
<td>15</td>
<td>47</td>
<td>47</td>
</tr>
<tr>
<td>Inter Fix ALIF pilot</td>
<td>February 1999</td>
<td>ALIF</td>
<td>Single-level open</td>
<td>rhBMP-2/ACS/MASTERGRAFT/CD HORIZON (12mg)</td>
<td>27</td>
<td>2</td>
<td>25</td>
<td>56</td>
<td>0</td>
</tr>
<tr>
<td>Inter Fix ALIF pilot</td>
<td>February 1999</td>
<td>ALIF</td>
<td>Single-level open</td>
<td>rhBMP-2/ACS+NOVUS LC Device</td>
<td>23</td>
<td>2</td>
<td>21</td>
<td>57</td>
<td>0</td>
</tr>
<tr>
<td>BCP US</td>
<td>May 1999</td>
<td>PLF</td>
<td>Single level</td>
<td>rhBMP-2/BCP (3mg/ml)</td>
<td>22</td>
<td>0</td>
<td>11</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>BCP Canada</td>
<td>September 1999</td>
<td>PLF</td>
<td>One or two level</td>
<td>rhBMP-2/BCP + TSRH (2mg/ml)</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>52</td>
<td>40</td>
</tr>
<tr>
<td>Amplify</td>
<td>March 2002</td>
<td>PLF</td>
<td>Open bilateral</td>
<td>rhBMP-2/CRM/CD Horizon (2mg/ml)</td>
<td>262</td>
<td>23</td>
<td>239</td>
<td>53</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICBG + CD Horizon</td>
<td>256</td>
<td>32</td>
<td>224</td>
<td>52</td>
<td>42</td>
<td>28</td>
</tr>
<tr>
<td>Trial name</td>
<td>Date of 1st patient enrollment</td>
<td>Surgical Approach</td>
<td>Intervention/Control</td>
<td>Patients randomised/enrolled</td>
<td>Patients excluded</td>
<td>Patients evaluated</td>
<td>Mean age</td>
<td>Percentage male</td>
<td>Previous spinal surgery (%)</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
<td>------------------------------</td>
<td>------------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>----------</td>
<td>-----------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Glassman (not Medtronic)</td>
<td>May 2004</td>
<td>PLF Single or multilevel</td>
<td>rhBMP-2/ACS + fixation</td>
<td>n/a</td>
<td>n/a</td>
<td>52</td>
<td>69</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ICBG + fixation</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>54</td>
<td>70</td>
<td>33</td>
<td>37</td>
</tr>
</tbody>
</table>
Table 2: Incidence of cancer in the Medtronic RCTs.

<table>
<thead>
<tr>
<th>Cancer</th>
<th>rhBMP-2</th>
<th>ICBG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skin cancers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanoma</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Basal cell carcinoma</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Breast</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Colon</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Larynx</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Leukaemia</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lung</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Non-Hodgkin’s lymphoma</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ovarian</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Pancreatic</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Prostate</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Stomach</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Testicular</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Thyroid</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total cancers</strong></td>
<td><strong>20</strong></td>
<td><strong>8</strong></td>
</tr>
<tr>
<td><strong>Total patients</strong></td>
<td><strong>672</strong></td>
<td><strong>609</strong></td>
</tr>
</tbody>
</table>

* Includes 3 cancers in rhBMP-2 patients identified through extended follow-up of the LT CAGE open trial (one thyroid, one testicular, one melanoma)
### Appendix table 1: Details of publications of Medtronic trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Publication</th>
<th>Full paper</th>
<th>Abstract only</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RCTs (rhBMP-2 vs ICBG)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LT Cage pilot</td>
<td>Boden et al 2000</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LT Cage open</td>
<td>Dickman 2001</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Gornet et al 2001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2002</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2003a</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2003b</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chhabra et al 2006</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Chhabra et al 2007</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Van Genugten et al 2008</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Alt et al 2009</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2009</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2011</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Bone dowel pilot</td>
<td>Burkus et al 2001</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2002</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2005</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Bone dowel pivotal</td>
<td>Burkus et al 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus 2005</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Inter Fix PLIF</td>
<td>Alexander et al 2002</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Burkus et al 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Haid et al 2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Cornerstone pilot</td>
<td>Baskin et al 2003</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Mastergraft pilot</td>
<td>Bae et al 2007</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>BCP US</td>
<td>Boden et al 2002</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>BCP Canada</td>
<td>Assiri 2004</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Alexander et al 2007</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abraham et al 2008</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abraham et al 2010</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Amplify</td>
<td>Glassman et al 2005</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dimar et al 2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dimar et al 2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glassman et al 2007</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dimar et al 2009</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>McInnis 2010</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Inter Fix ALIF pilot</td>
<td>Not published</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cornerstone ACDF pivotal</td>
<td>Not published</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>(terminated after recruitment of 3 patients)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT (no ICBG arm)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study Details</td>
<td>References</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Maverick Disc pivotal</strong></td>
<td></td>
<td>Gornet et al 2007</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gornet et al 2011</td>
<td></td>
</tr>
<tr>
<td><strong>Single-arm rhBMP-2 trials</strong></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>LT Cage laparoscopic</strong></td>
<td></td>
<td>Kleeman et al 2001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zdeblick et al 2001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Burkus et al 2003a</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chhabra et al 2006</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chhabra et al 2007</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Van Genugten et al 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alt et al 2009</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Burkus et al 2009</td>
<td></td>
</tr>
<tr>
<td><strong>Telamon</strong></td>
<td></td>
<td>Not published</td>
<td></td>
</tr>
<tr>
<td><strong>CRM Two-level pilot</strong></td>
<td></td>
<td>Not published</td>
<td></td>
</tr>
<tr>
<td><strong>BCP Mexico pilot</strong></td>
<td></td>
<td>Not published</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix table 2: Missing outcome data in the Medtronic trials

<table>
<thead>
<tr>
<th></th>
<th>Number (and percentage) of missing observations</th>
<th>6 weeks</th>
<th>3 months</th>
<th>6 months</th>
<th>12 months</th>
<th>24 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oswestry ICBG</td>
<td></td>
<td>18</td>
<td>14</td>
<td>29</td>
<td>42</td>
<td>79</td>
</tr>
<tr>
<td>Disability</td>
<td></td>
<td>3.0%</td>
<td>2.3%</td>
<td>4.8%</td>
<td>6.9%</td>
<td>13.0%</td>
</tr>
<tr>
<td>Index rhBMP-2</td>
<td></td>
<td>18</td>
<td>10</td>
<td>20</td>
<td>33</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.6%</td>
<td>1.4%</td>
<td>2.9%</td>
<td>4.8%</td>
<td>8.6%</td>
</tr>
<tr>
<td>SF-36 PCS ICBG</td>
<td></td>
<td>53</td>
<td>53</td>
<td>39</td>
<td>59</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.7%</td>
<td>8.7%</td>
<td>6.4%</td>
<td>9.7%</td>
<td>17.1%</td>
</tr>
<tr>
<td>rhBMP-2</td>
<td></td>
<td>91</td>
<td>88</td>
<td>54</td>
<td>64</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.1%</td>
<td>12.7%</td>
<td>7.8%</td>
<td>9.2%</td>
<td>14.4%</td>
</tr>
<tr>
<td>Back pain ICBG</td>
<td></td>
<td>36</td>
<td>33</td>
<td>45</td>
<td>59</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.9%</td>
<td>5.4%</td>
<td>7.4%</td>
<td>9.7%</td>
<td>15.6%</td>
</tr>
<tr>
<td>rhBMP-2</td>
<td></td>
<td>40</td>
<td>40</td>
<td>49</td>
<td>61</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.8%</td>
<td>5.8%</td>
<td>7.1%</td>
<td>8.8%</td>
<td>12.4%</td>
</tr>
<tr>
<td>Leg pain ICBG</td>
<td></td>
<td>36</td>
<td>33</td>
<td>45</td>
<td>59</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.9%</td>
<td>5.4%</td>
<td>7.4%</td>
<td>9.7%</td>
<td>15.6%</td>
</tr>
<tr>
<td>rhBMP-2</td>
<td></td>
<td>40</td>
<td>40</td>
<td>49</td>
<td>61</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.8%</td>
<td>5.8%</td>
<td>7.1%</td>
<td>8.8%</td>
<td>12.4%</td>
</tr>
<tr>
<td>Spinal fusion ICBG</td>
<td></td>
<td>96</td>
<td>87</td>
<td>106</td>
<td>15.8%</td>
<td>14.3%</td>
</tr>
<tr>
<td>rhBMP-2</td>
<td></td>
<td>98</td>
<td>89</td>
<td>112</td>
<td>14.1%</td>
<td>12.8%</td>
</tr>
</tbody>
</table>

* 694 patients received rhBMP-2 surgery; 608 ICBG surgery
**Appendix table 3: Details of published studies reporting on the safety of rhBMP-2 surgery**

<table>
<thead>
<tr>
<th>Comparator</th>
<th>Surgical approach</th>
<th>Author</th>
<th>rhBMP-2 No. Patients</th>
<th>Comparator</th>
<th>rhBMP-2 No. Patients</th>
<th>Adverse events reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICBG</td>
<td>Anterior lumbar</td>
<td>Carragee (2012)(20)</td>
<td>239</td>
<td>ICBG</td>
<td>233</td>
<td>Wound complications</td>
</tr>
<tr>
<td>ICBG</td>
<td>Anterior lumbar</td>
<td>Pradhan (2006) (26)</td>
<td>9</td>
<td>ICBG</td>
<td>27</td>
<td>Neurologic events, hardware failure, leg pain/radiculitis</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Glassman (2007) (28)</td>
<td>91</td>
<td>ICBG</td>
<td>35</td>
<td>Hardware failure</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Lee (2010) (29)</td>
<td>86</td>
<td>ICBG</td>
<td>41</td>
<td>Infection, wound complications, Heterotopic bone formation</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Mannion (2010) (30)</td>
<td>17</td>
<td>ICBG</td>
<td>19</td>
<td>Neurologic events</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Mummanen (2004) (31)</td>
<td>25</td>
<td>ICBG</td>
<td>19</td>
<td>Heterotopic bone formation, osteolysis, infection, hardware failure, leg pain/radiculitis, inflammatory cyst formation</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Rihn (2009) (32)</td>
<td>86</td>
<td>ICBG</td>
<td>33</td>
<td>Heterotopic bone formation, osteolysis, infection, hardware failure, leg pain/radiculitis, inflammatory cyst formation</td>
</tr>
<tr>
<td>ICBG</td>
<td>Posterior lumbar</td>
<td>Singh (2006) (33)</td>
<td>41</td>
<td>ICBG</td>
<td>11</td>
<td>Heterotopic bone formation, dysphagia</td>
</tr>
<tr>
<td>Allograft</td>
<td>Cervical</td>
<td>Lu (2010) (34)</td>
<td>100</td>
<td>ICBG</td>
<td>50</td>
<td>Heterotopic bone formation, wound complications</td>
</tr>
<tr>
<td>Allograft</td>
<td>Anterior lumbar</td>
<td>Slosar (2007) (35)</td>
<td>45</td>
<td>ICBG</td>
<td>30</td>
<td>Hardware failure</td>
</tr>
<tr>
<td>Allograft</td>
<td>Various</td>
<td>Vaidya (2007b) (36)</td>
<td>36</td>
<td>ICBG</td>
<td>41</td>
<td>Dysphagia</td>
</tr>
<tr>
<td>Allograft</td>
<td>Posterior lumbar</td>
<td>Mindea (2009) (37)</td>
<td>35</td>
<td>ICBG</td>
<td>8</td>
<td>Leg pain/radiculitis</td>
</tr>
<tr>
<td>Autologous bone</td>
<td>Posterior lumbar</td>
<td>Gray (2010) (38)</td>
<td>82</td>
<td>ICBG</td>
<td>39</td>
<td>Heterotopic bone formation, osteolysis, hardware</td>
</tr>
<tr>
<td>Bone Source</td>
<td>Levels</td>
<td>Study (Year)</td>
<td>n1</td>
<td>n2</td>
<td>Additional Complications</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>----</td>
<td>----</td>
<td>------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Autologous bone</td>
<td>Posterior lumbar</td>
<td>Joseph (2007) (39)</td>
<td></td>
<td></td>
<td>Failure, leg pain/radiculitis, Heterotopic bone formation</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Cervical</td>
<td>Longley (2009) (40)</td>
<td>84</td>
<td>84</td>
<td>Dysphagia</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Cervical</td>
<td>Williams (2011) (43)</td>
<td>4532</td>
<td>652</td>
<td>Dysphagia, wound complications, Heterotopic bone formation, infection, hardware failure, dysphagia, recurrent laryngeal neck palsy, wound complications</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Cervical</td>
<td>Xu (2011) (44)</td>
<td>48</td>
<td>156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Posterior lumbar</td>
<td>Rowan (2011) (46)</td>
<td>64</td>
<td>40</td>
<td>Leg pain/radiculitis</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Posterior lumbar</td>
<td>Williams (2011) (43)</td>
<td>6049</td>
<td>18267</td>
<td>Wound complications, inflammatory cyst formation, Dysphagia, wound complications</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Various</td>
<td>Cahill (2009) (2)</td>
<td>17626</td>
<td>53026</td>
<td>Wound complications, Heterotopic bone formation, neurologic events, hardware failure, wound complications</td>
<td></td>
</tr>
<tr>
<td>No rhBMP-2</td>
<td>Various</td>
<td>Williams (2011) (43)</td>
<td>11933</td>
<td>43929</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone marrow aspirate</td>
<td>Anterior lumbar</td>
<td>Gerszten (2011) (47)</td>
<td>45</td>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone marrow aspirate</td>
<td>Posterior lumbar</td>
<td>Taghvai (2010) (48)</td>
<td>24</td>
<td>38</td>
<td>Hardware failure</td>
<td></td>
</tr>
<tr>
<td>Silicated Calcium Phosphate</td>
<td>Posterior lumbar</td>
<td>Pimenta (2011) (49)</td>
<td>15</td>
<td>15</td>
<td>Heterotopic bone formation</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>Cervical</td>
<td>Hiremath (2009) (50)</td>
<td>16</td>
<td>67</td>
<td>Neurologic events, hardware failure, dysphagia, recurrent neck palsy, wound complications</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1.
Figure 2.
### Estimates with 95% confidence intervals

<table>
<thead>
<tr>
<th>Trial</th>
<th>rhBMP Fused/Total</th>
<th>ICBG Fused/Total</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT Cage Open</td>
<td>123/130</td>
<td>106/119</td>
<td>1.06 (0.99, 1.15)</td>
</tr>
<tr>
<td>Bone Dowel Pilot</td>
<td>24/24</td>
<td>13/19</td>
<td>1.46 (1.08, 1.98)</td>
</tr>
<tr>
<td>Bone Dowel Pivotal</td>
<td>43/44</td>
<td>22/27</td>
<td>1.20 (1.00, 1.44)</td>
</tr>
<tr>
<td>Interfix ALIF Pilot</td>
<td>16/20</td>
<td>13/14</td>
<td>0.86 (0.66, 1.12)</td>
</tr>
<tr>
<td>Cornerstone Pilot</td>
<td>11/11</td>
<td>11/11</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Interfix PLIF</td>
<td>24/26</td>
<td>21/27</td>
<td>1.19 (0.94, 1.49)</td>
</tr>
<tr>
<td>Mastergraft Pilot</td>
<td>18/19</td>
<td>14/20</td>
<td>1.35 (1.00, 1.84)</td>
</tr>
<tr>
<td>BCP US</td>
<td>7/10</td>
<td>3/3</td>
<td>0.70 (0.47, 1.05)</td>
</tr>
<tr>
<td>BCP Canada</td>
<td>86/89</td>
<td>64/91</td>
<td>1.37 (1.20, 1.58)</td>
</tr>
<tr>
<td>Amplify</td>
<td>186/194</td>
<td>151/169</td>
<td>1.07 (1.01, 1.14)</td>
</tr>
</tbody>
</table>

*Pooled*

1.14 (1.03, 1.25)

---

**Relative risk of fusion (rhBMP-2 vs ICBG)**

ICBG better → rhBMP-2 better

Figure 3.
Figure 4.
Figure 5.
### Estimates with 95% confidence intervals

<table>
<thead>
<tr>
<th>Study</th>
<th>Comparator</th>
<th>Surgery</th>
<th>rhBMP-2 events/total</th>
<th>Comparator events/total</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterotopic bone formation</td>
<td>Rinh</td>
<td>KBG</td>
<td>TLIF</td>
<td>2/96</td>
<td>0/33</td>
</tr>
<tr>
<td></td>
<td>Mannion</td>
<td>KBG</td>
<td>TLIF</td>
<td>2/17</td>
<td>0/19</td>
</tr>
<tr>
<td></td>
<td>Gray</td>
<td>Autologous bone</td>
<td>PLIF</td>
<td>43/82</td>
<td>19/29</td>
</tr>
<tr>
<td></td>
<td>Joseph</td>
<td>Autologous bone</td>
<td>TLIF/PLIF</td>
<td>5/23</td>
<td>1/10</td>
</tr>
<tr>
<td></td>
<td>Pimenta</td>
<td>SCP</td>
<td>UIF</td>
<td>1/15</td>
<td>0/15</td>
</tr>
<tr>
<td>Radiculitis</td>
<td>Rinh</td>
<td>KBG</td>
<td>TLIF</td>
<td>2/96</td>
<td>0/33</td>
</tr>
<tr>
<td></td>
<td>Mannion</td>
<td>KBG</td>
<td>TLIF</td>
<td>2/17</td>
<td>0/19</td>
</tr>
<tr>
<td></td>
<td>Gray</td>
<td>Autologous bone</td>
<td>PLIF</td>
<td>43/82</td>
<td>19/29</td>
</tr>
<tr>
<td></td>
<td>Joseph</td>
<td>Autologous bone</td>
<td>TLIF/PLIF</td>
<td>5/23</td>
<td>1/10</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>Pimenta</td>
<td>SCP</td>
<td>UIF</td>
<td>1/15</td>
<td>0/15</td>
</tr>
<tr>
<td></td>
<td>Rinh</td>
<td>KBG</td>
<td>TLIF</td>
<td>2/86</td>
<td>0/33</td>
</tr>
<tr>
<td></td>
<td>Mannion</td>
<td>KBG</td>
<td>TLIF</td>
<td>2/17</td>
<td>0/19</td>
</tr>
<tr>
<td></td>
<td>Gray</td>
<td>Autologous bone</td>
<td>PLIF</td>
<td>43/82</td>
<td>19/29</td>
</tr>
<tr>
<td></td>
<td>Joseph</td>
<td>Autologous bone</td>
<td>TLIF/PLIF</td>
<td>5/23</td>
<td>1/10</td>
</tr>
<tr>
<td></td>
<td>Pimenta</td>
<td>SCP</td>
<td>UIF</td>
<td>1/15</td>
<td>0/15</td>
</tr>
</tbody>
</table>

**Relative risk of adverse event**

![Figure 6.](image-url)
Appendix Figure 1.
Appendix Figure 2.
Appendix Figure 3.
Appendix Figure 4.
Appendix Figure 5.
Appendix Figure 6.
### Estimates with 95% confidence intervals

<table>
<thead>
<tr>
<th>Type of adverse event</th>
<th>Total ICBG</th>
<th>Total rhBMP-2</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arm and neck pain</td>
<td>12</td>
<td>13</td>
<td>1.02 (0.48, 2.27)</td>
</tr>
<tr>
<td>Arthritis/bursitis</td>
<td>16</td>
<td>24</td>
<td>1.42 (0.75, 2.69)</td>
</tr>
<tr>
<td>Back and leg pain</td>
<td>207</td>
<td>268</td>
<td>1.20 (0.99, 1.48)</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>99</td>
<td>93</td>
<td>0.66 (0.65, 1.16)</td>
</tr>
<tr>
<td>Death</td>
<td>10</td>
<td>6</td>
<td>0.58 (0.29, 1.15)</td>
</tr>
<tr>
<td>Dural injury</td>
<td>12</td>
<td>13</td>
<td>0.84 (0.42, 2.10)</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>2</td>
<td>3</td>
<td>0.62 (0.19, 2.06)</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>115</td>
<td>117</td>
<td>0.90 (0.69, 1.17)</td>
</tr>
<tr>
<td>Implant</td>
<td>10</td>
<td>22</td>
<td>1.93 (0.91, 4.11)</td>
</tr>
<tr>
<td>Infection</td>
<td>90</td>
<td>104</td>
<td>1.06 (0.79, 1.42)</td>
</tr>
<tr>
<td>Neurological</td>
<td>119</td>
<td>152</td>
<td>1.15 (0.90, 1.48)</td>
</tr>
<tr>
<td>Other pain</td>
<td>83</td>
<td>97</td>
<td>1.38 (1.00, 1.92)</td>
</tr>
<tr>
<td>Respiratory</td>
<td>27</td>
<td>34</td>
<td>1.11 (0.66, 1.85)</td>
</tr>
<tr>
<td>Retrograde ejaculation</td>
<td>1</td>
<td>5</td>
<td>4.76 (0.55, 40.95)</td>
</tr>
<tr>
<td>Spinal</td>
<td>63</td>
<td>70</td>
<td>0.99 (0.79, 1.21)</td>
</tr>
<tr>
<td>Trauma</td>
<td>137</td>
<td>153</td>
<td>1.01 (0.60, 2.28)</td>
</tr>
<tr>
<td>Urogenital</td>
<td>63</td>
<td>81</td>
<td>1.19 (0.85, 1.66)</td>
</tr>
<tr>
<td>Vascular</td>
<td>4</td>
<td>6</td>
<td>1.55 (0.37, 6.88)</td>
</tr>
<tr>
<td>Vertebral fracture</td>
<td>1</td>
<td>3</td>
<td>2.79 (0.29, 26.89)</td>
</tr>
<tr>
<td>Wound complication</td>
<td>4</td>
<td>6</td>
<td>1.50 (0.42, 5.36)</td>
</tr>
</tbody>
</table>

**Odds ratio of adverse events (rhBMP-2 vs ICBG)**

More common with ICBG  →  More common with rhBMP-2

---

Appendix Figure 7.
Appendix Figure 8.
Search strategies

**OVID Medline (1948 to present)**

1. bone morphogenetic proteins/ or bone morphogenetic protein 2/ (10872)
2. ((bone morphogen$ or osteogen$ or osteoinduct$) adj (protein$ or factor$ or polypeptide$ or poly-peptide$)).af. (16131)
3. (bmp or bmp2 or bmp-2).af. (11384)
4. (rhbmp or rhbmp2 or rhbmp-2).af. (1262)
5. (rh-bmp or rh-bmp2 or rh-bmp-2).af. (72)
6. (infuse or amplify).af. (15590)
7. or/1-6 (33622)
8. Spinal Fusion/ (14504)
9. (spine or spinal).af. (322484)
10. spondylosyndes$.af. (7)
11. spondylodes$.af. (631)
12. lumbar interbody arthrodesis.af. (22)
13. ((lumbar or cervical or posterior or anterior or lumbosacral or transforminal or posterolateral) adj3 fusion$).af. (7931)
14. fusion cage.af. (125)
15. or/8-14 (323015)
16. 7 and 15 (1312)
17. exp animals/ not humans.sh. (371340)
18. 16 not 17 (827)

**EMBASE (1974 to present)**

1. bone morphogenetic protein/ or bone morphogenetic protein 2/ (12676)
2. ((bone morphogen$ or osteogen$ or osteoinduct$) adj (protein$ or factor$ or polypeptide$ or poly-peptide$)).af. (19870)
3. (bmp or bmp2 or bmp-2).af. (11817)
4. (rhbmp or rhbmp2 or rhbmp-2).af. (1435)
5. (rh-bmp or rh-bmp2 or rh-bmp-2).af. (90)
6. (infuse or amplify).af. (17204)
7. or/1-6 (38898)
8. Spine Fusion/ (13270)
9 (spine or spinal).af. (369003)
10 spondylosyndes$.af. (10)
11 spondylodes$.af. (1709)
12 lumbar interbody arthrodesis.af. (23)
13 ((lumbar or cervical or posterior or anterior or lumbosacral or transforminal or posterolateral) adj3 fusion$).af. (9892)
14 fusion cage.af. (196)
15 or/8-14 (369756)
16 7 and 15 (1898)
17 animal experiment/ (1579120)
18 16 not 17 (1542)

Cochrane Central Register of Controlled Trials (CENTRAL)
#1 MeSH descriptor Bone Morphogenetic Proteins, this term only
#2 MeSH descriptor Bone Morphogenetic Protein 2, this term only
   (morphogen* NEXT (protein* or factor* or polypeptide* or poly-peptide*)) or
#3 (osteogen* NEXT (protein* or factor* or polypeptide* or poly-peptide*)) or
   (osteoinduct* NEXT (protein* or factor* or polypeptide* or poly-peptide*))
#4 (bmp or bmp2 or bmp-2)
#5 (rhbmp or rhbmp2 or rhbmp-2)
#6 (rh-bmp or rh-bmp2 or rh-bmp-2)
#7 (infuse or amplify)
#8 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7)
#9 MeSH descriptor Spinal Fusion, this term only
#10 (spine or spinal)
#11 (spondylosyndes*)
#12 (spondylodes*)
#13 ("lumbar interbody arthrodesis")
#14 ((lumbar or cervical or posterior or anterior or lumbosacral or transforminal or posterolateral) NEAR/3 fusion*)
#15 ("fusion cage")
#16 (#9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15)
#17 (#8 AND #16)

**Science Citation Index Expanded (SCI-EXPANDED)**

**1899-present**

# 24  #14 not #23
# 23 #22 OR #21 OR #20 OR #19 OR #18 OR #17 OR #16 OR #15
# 22 Title=(genera or taxonomy or species or fauna or habitat or marine or ecology)
# 21 Title=(cow or cattle or bovine or livestock or swine or poultry)
# 20 Title=(rabbit or rabbits or moss or mosses or fungus or fungi)
# 19 Title=(fossil or fossils or lichen or lichens or mushroom or mushrooms)
# 18 Title=(bat or bats or bee or bees or grass or grasses or bird or birds or avian)
# 17 Title=(bovine or sheep or fly or flies or fish or fishes or fisheries or horse or horses or equine)
# 16 350,781  Title=(animal or animals or dog or dogs or canine or cat or cats or feline)

# 15 Title=(rat or rats or mouse or mice or hamster or hamsters)
# 14 #13 AND #6
# 13 #12 OR #11 OR #10 OR #9 OR #8 OR #7
# 12 Topic="fusion cage"
# 11 Topic=((lumbar or cervical or posterior or anterior or lumbosacral or transforminal or posterolateral) NEAR/3 fusion*)
# 10 Topic="lumbar interbody arthrodesis"
# 9  Topic=(spondylodes*)
# 8  Topic=(spondylosyndes*)
# 7 Topic=(spine or spinal)
# 6 #5 OR #4 OR #3 OR #2 OR #1
# 5 Topic=(infuse or amplify)
# 4 Topic=(rh-bmp or rh-bmp2 or rh-bmp-2)
# 3 Topic=(rhbmp or rhbmp2 or rhbmp-2)
# 2 Topic=(bmp or bmp2 or bmp-2)
Call for evidence

Systematic review of bone morphogenic protein-2 (rhBMP-2) for spinal fusion - call for evidence

The Centre for Reviews and Dissemination (CRD) is undertaking a systematic review and individual participant data (IPD) meta-analysis of the comparative effectiveness of rhBMP-2 (marketed as INFUSE) for spinal fusion. The review has been commissioned by the Yale University Open Data Access (YODA) initiative as part of an overarching project to systematically review the safety and effectiveness of rhBMP-2, including re-analysis of IPD that have been made available to Yale on an unrestricted basis by the manufacturer (Medtronic Inc). YODA aims to improve access to patient-level data from clinical trials, and provide independent, scientifically rigorous, objective and fair analyses of such data.

CRD will undertake a comprehensive and rigorous systematic review and meta-analysis of individual participant data (IPD) of all relevant randomised controlled trials that have compared rhBMP-2 with standard bone graft therapy.

We will include all relevant randomised controlled trials, irrespective of whether conducted by the manufacturer or not, and irrespective of whether published or not.

We are therefore interested in hearing from anyone who has conducted, or is aware of, unpublished or partially published research in this area. For example, trials which have been presented at conferences but not fully reported elsewhere.

We are currently aware of 17 trials funded by the manufacturer and have searched the published literature but welcome any information regarding further unpublished research. If you know of any such trials please contact [CRD details deleted].

Link to CRD project page
http://www.york.ac.uk/inst/crd/projects_in_progress.cfm
Link to YODA page
http://medicine.yale.edu/core/projects/yodap/index.aspx
Methods Appendix

Data sources used

Analyses of efficacy were restricted to randomised controlled trials in spinal fusion surgery in which rhBMP-2 had been compared with conventional iliac crest bone graft surgery (ICBG). Single-arm trials of rhBMP-2 or trials with comparators other than IBCG were excluded. One trial (rhBMP-2/BCP US pilot RCT) had two rhBMP-2 arms using different fixation procedures. Only the primary arm was used in these analyses. The second rhBMP-2 arm (of 11 patients) was combined with the first in sensitivity analyses (not presented in this paper). We performed all analyses using the patient-level data supplied by Medtronic and from the Glassman trial. Although intention-to-treat analyses were intended this was not possible as a number of randomised patients withdrew before surgery and no outcome data were available for them.

Outcomes of interest

Primary outcomes

- Disease-specific pain and functionality
  - Oswestry Disability Index (ODI; or Neck Disability Index (NDI) for cervical spinal surgery); this measures lower back (or neck) pain on a scale from 0 (no pain) to 100% (extreme pain).
  - SF-36 Physical Component Score (SF-36 PCS), which assesses both pain and physical function on a scale from 0 (worst) to 100% (best).
  - Back and leg pain; both measured on a scale from 0 (no pain) to 20 (extreme pain).
- Successful spinal fusion
  - Defined radiographically by Medtronic as requiring all of the following: evidence of bridging trabeculae, no evidence of motion (<3mm difference in translation, less than 5° difference in angular motion) and no evidence of radiolucency.

Secondary outcomes

- Duration of hospital stay
- Operating time
- Successful return to work or usual activity
- Use of pain medication

Safety outcomes

We analysed adverse events supplied by Medtronic that we considered to be potentially related to spinal surgery according to the categorisations provided in the Medtronic IPD. We also considered the following broad categories of adverse events:

- Pain: back, leg, lower extremity, arm, neck and upper extremity pain
- Implant related (hardware failure): displacement, breakage, loosening, malpositioning and subsidence
- Infection
- Neurological events: including numbness, tingling, “pins and needles”
- Cancer

For adverse events not generally reported by Medtronic we searched the wider literature and analysed the following:
• Leg pain (including radiculitis)
• Heterotopic bone formation
• Dysphagia (in cervical spinal surgery)
• Retrograde ejaculation

Patient-level and trial-level factors affecting effectiveness

We investigated how the effectiveness of rhBMP-2 might be influenced by trial and patient-level characteristics. We investigated the following trial-level factors:

- spinal location of surgery (e.g., cervical or lumbosacral)
- surgical approach (e.g., anterior lumbar fusion, posterior lumbar interbody fusion)

We also investigated the following patient-level factors:

- previous spinal surgical interventions
- age
- gender
- smoking status
- diabetic status
- body mass index

Time points

These outcomes were analysed at a range of different time points after surgery, namely: six weeks, three, six, twelve and 24 months. Data on successful fusion were available only from six months after surgery onwards. Data on adverse events were provided at these times and were also available and analysed at or immediately after surgery (up to four weeks). For all trials data were provided at all the described time points.

Data management and checking of individual patient data

Data were provided by Medtronic for each trial in a range of separate SAS-format data files according to the types of outcomes reported. From these, we collated individual-level data on all the available effectiveness outcomes at all the time points listed above. Data from the single non-Medtronic (Glassman) trial were provided as a single Excel spreadsheet. For each trial for which IPD were provided we checked the consistency of the data, both to ensure that the data as provided were valid, and to check for errors in our data collation process. Data checking procedures included:

- Checking uniqueness and consistency of patient identification numbers
- Checking consistency of treatment allocation records
- Ensuring all demographic data (e.g., ages) were within plausible ranges
- Ensuring all pain and function measurements (e.g., ODI score, SF36 Physical Component Score) were within range, with no outliers
- Checking for balanced randomisation in terms of age, sex and other patient-level factors
- Checking that summary scores (e.g., ODI score) agreed with the raw scores from each question from the questionnaire

It was not possible to check judgements of fusion status without access to the raw radiological data and a radiologist. However, if fusion status was assessed by blinded experts any possible reporting bias or inconsistencies in judgment should be shared across both rhBMP-2 and the control arms. We have assumed that the assessments of spinal fusion provided by Medtronic are valid.
Quality assessment and Risk of Bias

In addition to the data checking procedures we also assessed each trial using the Cochrane Risk of Bias Tool [1]. This was developed by the Cochrane Collaboration to assess those aspects of trial design and conduct that have been demonstrated empirically to affect treatment effect estimates. The tool does not address aspects of trial design that relate to applicability or generalisability. The Risk of Bias tool covers six key areas of potential bias: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessors, incomplete outcome data and selective reporting of outcomes. Each area was given a judgement of high risk of bias, unclear risk of bias or low risk of bias, and a reason for each judgement was recorded in the main data extraction spreadsheet. We used the guidelines from The Cochrane Collaboration on what constituted high, low and unclear risk of bias, with additional details based on discussions with the clinical member of the team (RM) Error! Reference source not found.

Two reviewers independently completed Risk of Bias assessments for all trials included in the efficacy analyses. Judgements were made for each type of outcome reported in the trials: fusion, patient-reported and adverse events.

These decisions were based on the full trial protocols provided by Medtronic and the brief protocol provided by the authors of the Glassman trial.

To address incomplete outcome data we used standard data checking procedures described in the previous section to compare loss to follow-up in each arm. We also checked that we had been provided with IPD for all outcomes that were listed in the trial protocols. We requested from Medtronic any available data for patients who were recruited but not reported in the trials. Some tabulated information was provided on why these patients did not receive surgery, and so were excluded from the IPD, but further data were unavailable. No post randomisation data had been collected for these individuals. We also checked loss to follow-up and missing outcome data (by treatment arm) for main outcomes at each analysis time point.

Quality of the non-randomised studies were assessed using a domain-based approach, based on a modified version of the Newcastle-Ottawa scale [2]. However, we avoided scoring or assigning star values to the studies as this runs the risk of producing an uninformative summary score. Instead, we tabulated the relevant information for each of the following domains:

- Representativeness of the exposed cohort (did the patients require spinal fusion surgery, did they form a particular subgroup e.g. all smokers)
- Selection of the control group (were they drawn from the same source e.g. same hospital or database)
- Ascertainment of exposure (given the nature of the topic, this was usually via secure record in which the exposure to treatment was clearly known)
- Outcome of interest not present at start of the study (was this explicitly checked for e.g. pre-existing cancer)
- Confounders or other factors used to match or controlled for in analysis
- Outcome assessment (independent/blind assessment, via secure medical record, self-report, or no details)
- Follow-up (we considered adequacy of duration in relation to the specific adverse events reported and whether this was similar across the two groups)

Assessments were checked, with disagreements resolved by discussion and/or consultation with a third researcher.
**Statistical analysis**

Our primary statistical method for estimating the efficacy of rhBMP-2 surgery for all the specified outcomes was to use standard two-stage meta-analytic techniques.[3-4] IPD from each trial were analysed separately, using the same methods across trials, for all the efficacy outcomes. Separate meta-analyses were performed at each of the specified time points. We also used a “one-stage” meta-analysis approach in some analyses, primarily as a sensitivity analysis to confirm the results of two-stage analyses. These approaches are described in more detail below.

All main analyses used a complete-case approach, where participants with missing data were excluded from the analysis.

**Estimates of effect**

*Continuously distributed outcomes*

For the continuously distributed outcomes (ODI, SF-36, back pain, leg pain) we assessed efficacy in terms of the mean difference (MD) in outcome between the rhBMP-2 and ICBG arms. For each patient, the change in the score from baseline to the time point of interest was calculated, and hence the mean change in each arm within each trial, and thus the mean difference between arms at that time. This mean difference along with its associated standard error was calculated for each trial.

*Dichotomous outcomes*

For dichotomous outcomes (successful fusion, successful return to work, use of pain-relief medication, cancer) we assessed efficacy in terms of the relative risk (RR) for the outcome between the rhBMP-2 and ICBG arms. In the one-stage random-effects meta-analyses relative risks could not be calculated because algorithms did not converge successfully (i.e. they crashed). For these models, results were therefore calculated in terms of the odds ratio (OR), with its corresponding 95% confidence interval.

**Two-stage meta-analyses**

We combined the effect estimates from each trial (mean difference or relative risk) across trials using a standard DerSimonian-Laird random-effects meta-analysis to account for potential heterogeneity in effects across trials.[5] Separate analyses were performed at each time point. We present the results as summary plots across all times. This is called a “two-stage” approach because it is performed in two stages: first we estimate effects within trials, and then combine results across trials in a meta-analysis.

**One-stage meta-analyses**

We performed one-stage meta-analyses of the pain and function outcomes as a comparison in order to confirm the validity of the two-stage analyses. In a one-stage analysis all patient data from all trials are combined simultaneously in a single regression model that is stratified by trial (hence in “one stage”). For ODI, for example, we used a random-effects linear regression model of change in ODI from baseline against treatment received. This model included data at all time points simultaneously, but with a separate treatment effect estimated at each time point. The model therefore does not assume any particular model for changes in effects over time. However, it does assume the same amount of between-patient variation at every time point, and the same amount of between-study variation in treatment effects (heterogeneity) at every time point. The model was also stratified according to the trial to which each patient belonged.[6] This model had the form:
$y_{ijk} = \alpha_{ik} + \beta_{ik}x_{ij} + \epsilon_{ij}$

$\epsilon_{ij} \sim N(0, \sigma^2_i)$

$\beta_{ik} \sim N(\beta_k, \tau^2)$

Where $y_{ijk}$ is the change from baseline in ODI at time $k$ for patient $j$ in trial $i$. And $x_{ij}$ is a coding for treatment received ($1 = $ rhBMP-2, $0 = $ ICBG). $\alpha_{ik}$ is the baseline change in score in trial $i$ at time $k$, $\beta_k$ is the mean difference between rhBMP-2 and ICBG surgery at time $k$ (i.e. the treatment effect), and $\tau^2$ is the heterogeneity in the treatment effect across trials.

For dichotomous outcomes, (successful fusion and adverse events), a similar random-effects logistic regression model, also stratified by trial, was used, of the form: [7]

$log \left( \frac{p_{ijk}}{1 - p_{ijk}} \right) = \alpha_{ik} + \beta_{ik}x_{ij}$

$\beta_{ik} \sim N(\beta_k, \tau^2)$

Where $p_{ijk}$ is the probability of successful fusion at time $k$ for patient $j$ in trial $i$, and so $\beta_{ik}$ is the log odds ratio of event (eg. successful fusion).

The numbers of events in any particularly adverse event category in any trial at any time points were often small, and some trials had no adverse events in a particular category. In such a situation, where events are rare, two-stage meta-analyses of relative risk may be inaccurate because of the corrections required to adjust for trials with no events. For all analyses of adverse events we therefore used only one-stage meta-analysis.

Assessment of heterogeneity

We assessed heterogeneity in all two-stage meta-analyses by calculating the $I^2$ statistic for heterogeneity.[8]

Exploring clinical heterogeneity

We investigated how trial-level and participant-level factors influenced the effectiveness of rhBMP-2 surgery to try to explain any heterogeneity in the meta-analyses. For trial-level factors, this was achieved using subgroup analysis. For example, for type of surgery, trials were divided into subgroups according to the surgery type (ALIF, PLIF or PLF, ACDF). Random effects meta-analyses were conducted within each surgery subgroup to estimate the effect of rhBMP-2 in each subgroup and results presented for each. The results across subgroups were compared using tests of heterogeneity, for example to identify any notable differences between surgery types.

For participant-level factors we used random-effects one-stage meta-analysis methods.[6] For example, to investigate the effect of age we took the linear regression model used for the one-stage analysis described for ODI above and incorporated an age parameter and an interaction term between age and treatment. This extended model had the form:

$y_{ijk} = \alpha_{ik} + \beta_{ik}x_{ij} + \mu z_{ij} + \gamma x_{ij}z_{ij} + \epsilon_{ij}$

$\epsilon_{ij} \sim N(0, \sigma^2_i)$

$\beta_{ik} \sim N(\beta_k, \tau^2)$

Where $z_{ij}$ is the initial age of patient $j$ in trial $i$, and $\gamma$ is the interaction between age and treatment, which measures the extent to which age affects the efficacy of rhBMP-2 surgery. This model is stratified by trial.

Sensitivity analyses

Some data were missing at some times after surgery because some patients were lost to follow-up or were not available at certain times. To investigate the potential for bias due to such missing data we performed a multiple imputation analysis for the main pain outcomes and for spinal fusion. This was achieved by performing a regression of the outcome at each time point against
the outcome at the previous time point, stratified by trial, using the complete-case data. Where a participant had no recorded outcome at some time point the predicted outcome from the model and its standard error were used to impute ten possible outcomes, assuming that outcomes were normally distributed. The resulting ten compete data sets were analysed and the results averaged to obtain a summary result after imputation.

**Multiple testing**

We note that we have performed a large number of analyses, for many outcomes at multiple time points. This increases the probability of identifying a significant difference between treatment arms when there is in fact no difference in any specific analysis to above the nominal 5% significance level (i.e. that associated with 95% confidence intervals).

We have not performed any corrections for multiple testing because the variation in numbers of outcomes considered and the different numbers of time periods involved in some analyses would mean that any correction would be arbitrary, and would make comparisons between outcomes and between main analyses and sensitivity and subgroup analyses difficult.

**Statistical software**

The IPD were supplied by Medtronic in SAS format. All data management, data checking and data extraction for the IPD was performed using SAS statistical software. Statistical analyses were performed using the R statistical software package. In particular, two-stage meta-analyses were conducted using the meta library in R, and one-stage analyses were performed using the lme4 library for multilevel modelling. Forest plots were produced in R using an in-house package. All other figures were produced in R using the ggplot library.

**References**

Systematic Review and Meta-Analysis of the Safety and Efficacy of Recombinant Human Bone Morphogenetic Protein-2 (rhBMP-2)

PROTOCOL

Background

Human bone morphogenetic protein-2 (rhBMP-2) is used in orthopaedic and spinal surgery to promote fusion. Following 2002 FDA approval for use in anterior lumbar fusion (ALIF) surgery, numbers of spinal procedures using rhBMP-2 grew rapidly, including many in off-label indications. This is despite the fact that the majority of procedures will fuse without use of rhBMP-2 and that in many cases fusion per se is not a pre-requisite for successful surgery. More recently, a number of small studies have raised concern over high rates of adverse events (AE) some of which are potentially life-threatening and which had not been reported in licensing studies. The need for rhBMP-2, its efficacy and its AE profile is therefore under considerable scrutiny and a robust re-evaluation of the research evidence is vital.

In recognition of the importance of this issue, the manufacturers of rhBMP-2 (Medtronic Inc) released all of its clinical research data that are relevant to the use of rhBMP-2 to Yale University for independent scrutiny and review. Yale has contracted two academic groups to carry out independent and unrestricted systematic reviews of the safety and efficacy of rhBMP-2 in spinal fusion, including re-analysis of the individual participant data from Medtronic studies.

This protocol describes the systematic review and associated methodological comparisons to be carried out by the Centre for Reviews and Dissemination (CRD), University of York, UK. The systematic review will consider data from Medtronic studies alongside any non-industry funded clinical research data on rhBMP-2.

In addition to the reviews of benefits and harms described below, the planned research will inform discussions around the comparability of IPD and aggregate data syntheses, and provide objective evidence around data disclosure, selective reporting, industry sponsorship, the availability of individual participant data, and how these might affect the ability to learn about efficacy and adverse events. In addition, this work will contribute to a wider debate on standards for data disclosure and dealing with potential conflict of interest.

Objectives

To evaluate whether rhBMP-2 is more or less effective than standard bone graft therapy (SBGT) in spinal fusion by rigorous systematic review and meta-analysis of relevant studies, including an analysis of individual patient data (IPD), considering

(i) Potential benefits of rhBMP-2, focusing on evidence from randomised controlled trials and on outcomes that are meaningful to patients.

(ii) Potential harms of rhBMP-2 by identifying serious adverse events that have been reported from its use in clinical trials and in general medical practice.

Methods for synthesis of evidence of clinical effectiveness and safety
Inclusion and exclusion criteria

Two reviewers will independently screen all titles and abstracts retrieved from electronic database and other searches. Full paper manuscripts of any publications that may be relevant will be obtained (where possible) and the relevance of each study assessed by two reviewers according to the criteria below. Studies that do not meet all of the criteria will be excluded and their bibliographic details listed with reasons for exclusion. Any discrepancies will be resolved by consensus and, if necessary, a third reviewer will be consulted.

Participants

Studies including patients undergoing spinal fusion surgery for treatment of degenerative disc disease, spondylolisthesis or any other relevant spinal condition will be included. Although the licensed indication is for use with anterior lumbar interbody fusion, inclusion will not be restricted by operative approach as this will allow evaluation of evidence pertaining to off-label use, particularly with respect to adverse events in cervical and lumbosacral spinal surgery. Anterior lumbar interbody fusion (ALIF), posterolateral lumbar fusion (PLF), posterolateral lumbar interbody fusion (PLIF), transforaminal lumbar interbody fusion (TLIF), and anterior cervical disectomy and fusion (ACDF) will all be eligible. Inclusion will not be restricted by type of surgery (open, minimally invasive or laparoscopic).

Studies of rhBMP-2 use outside spinal fusion surgery (e.g. in long bone fractures) will be excluded, as will all animal and in vitro studies. This study will focus only on rhBMP2 and not on other recombinant forms of BMP, such as rhBMP7.

Interventions

Studies evaluating rhBMP-2 in spinal fusion will be included in the review, including both Medtronic’s INFUSE/Inductos and AMPLIFY rhBMP-2 carrier/preparations.

Comparators

For the evaluation of benefits, studies comparing rhBMP-2 against any standard bone graft techniques (SBGT) will be eligible for inclusion. Studies without a comparator will also be eligible for inclusion in the evaluation of AEs potentially related to the use of rhBMP-2.

Outcomes

Inclusion will not be limited by which outcomes are reported. Any studies that are eligible but do not report outcomes of interest will be included and reported as such.

Study designs

Effectiveness
Only RCTs meeting the above criteria will be included in the review of comparative effectiveness.

**Adverse events**

In accordance with best practice,¹ the review of AEs will not be restricted to RCTs. A decision about which study designs will be included will be taken once we have completed a mapping exercise to determine the volume and scope of studies reporting adverse events, and made a preliminary assessment of the completeness and quality of the available registry data. Attention will be paid to the potential for overlap among identified studies (i.e. the same patients being included in case reports, cohort studies and registries).

**Search strategy**

A systematic literature search will be performed of the following databases: BIOSIS Previews, Cochrane Central Register of Controlled Trials (Central), the Database of Abstracts of Reviews of Effects (DARE), EMBASE, MEDLINE, MEDLINE in Process and Other Non-Indexed Citations, PubMed, Science Citation Index, TOXLINE, and the FDA website.

The search strategy will be designed to retrieve any studies relevant to the effectiveness or adverse events (AEs) of rhBMP-2 in spinal fusion. Synonyms will be searched for in title and abstract and appropriate indexing/keywords selected. No search filters for specific study designs will be used owing to limitations of searching beyond randomised controlled trials (RCTs) for AEs.² Searches will not be restricted by publication status or date. References of relevant papers will be checked for further relevant studies. Authors of included trials will be asked to notify us of any unpublished studies of which they are aware.

A summary of the search strategies used are presented in Appendix 1.

**Provision of IPD**

Yale University has secured the release of all clinical trial data (published and unpublished), post-marketing surveillance data and spontaneous adverse event data from Medtronic (the manufacturer of rhBMP-2). Yale will make these data available for inclusion in this systematic review.

We will also seek IPD for any additional RCTs identified by the literature searches.

**Obtaining Data**

**Data extraction strategy (for published aggregate data)**

A data extraction form will be developed, piloted and adjusted as necessary. Data extracted will include details of study design, setting, and sponsor as well as outcome, trial and patient characteristics.

Data will be extracted into EPPI-Reviewer/Excel. Data extraction will be undertaken independently by two researchers with discrepancies resolved by consensus or recourse to a third researcher if necessary.
Provision of individual patient data

Investigators of trials for which IPD is not supplied by Medtronic will be contacted and asked to participate in the review by providing individual participant data for inclusion and re-analysis. If they agree to participate, fully anonymised data on all randomised patients relating to the outcomes and trial and patient characteristics described above will be requested.

Data will be accepted in any suitable electronic format, but an example format detailing the recommended coding will be created and offered to all collaborators. Simple checks on the data will be made to ensure data are correctly coded, that missing data are correctly identified and to ensure that the data are consistent with published results. Data from all trials (including Medtronic-sponsored trials) will be incorporated into a single, database with fields that are consistent (as far as possible) across both Medtronic and non-Medtronic trials.

Data storage and confidentiality

All IPD and adverse event data from Medtronic will be transferred to CRD via password-protected memory stick. Data from other investigators will be in a de-identified format and received via secure FTP transfer or encrypted email. All data will be held in a password protected area of the CRD server. Access will be limited to staff working directly on the project. Copying data to laptop computers or memory sticks will be prohibited.

Outcomes

We will consider a range of outcomes and place emphasis on clinical or functional over radiological outcomes, particularly those that are directly meaningful to patients.

Effectiveness outcomes

The outcomes of interest will be:
- disease-specific questionnaires (e.g. Oswestry Disability Index, Neck Disability Index)
- patient QL / functional status questionnaires (e.g. SF-36)
- post-operative pain – surgical site and bone graft donor site
- duration of hospital stay
- operating time
- successful return to work/usual activity
- fusion status
- time to discharge*

Analysis of outcomes marked * are likely to be possible only in the IPD review.

Adverse event outcomes

General
- heterotopic bone formation
• osteolysis
• infection
• neurological events (new / worse leg pain, sensory disturbance, reflex changes, bladder disturbance)
• cancer
• hardware failure (e.g. cage subsidence, implant breakages)

Surgery-specific

• ALIF – major vascular injury, retrograde ejaculation, urinary retention
• ACDF - dysphagia, airway obstruction, neck pain, recurrent laryngeal nerve palsy
• PLIF/TLIF/PLF – leg pain/radiculitis, leg weakness, inflammatory cyst formation

Data on serious adverse events (AEs) will be extracted from relevant observational sources as well as from RCTs (where reported)

Other data

In order to investigate how the efficacy and safety of rhBMP-2 might be influenced by trial and patient-level characteristics we will also extract/obtain data on the following, if available:

• spinal location of surgery (eg. cervical or lumbosacral)
• type of surgery (eg. ALIF, PLIF etc.)
• rhBMP-2 dose/volume
• cage type
• nature of spinal condition (eg. degenerative disc disease, spondylolisthesis)
• comparator treatment
• previous surgical interventions *
• age *
• gender *
• smoking status *

Data marked * may only be available from IPD

Risk of bias (quality) assessment

Critical appraisal of RCTs will be based on trial publications, protocols, and where available on IPD. Risk of bias in RCTS will be assessed using the Cochrane Risk of Bias tool. Other study designs will be assessed based on CRD guidance. Assessment will be undertaken independently by two researchers with any discrepancies resolved by consensus or recourse to a third researcher if necessary.

All IPD will be subject to detailed checking including examination of patterns of missing data, integrity of randomisation via pattern of randomisation and balance across baseline characteristics, and internal consistency.

Methods of analysis/synthesis
A narrative and tabular summary of key study characteristics will be undertaken. Published main results and quality assessment of individual RCTs will also be tabulated.

**Aggregate data from RCTs**

Where appropriate (based on clinical similarity of trials and the necessary data being available) aggregate study results (from publications/FDA reports) will be combined in a series of random-effects meta-analyses. Separate analyses will be conducted for each outcome listed above.

It is anticipated that the measures used to assess continuous outcomes will vary between studies, and in these cases standardized mean differences will be calculated where appropriate, and combined in random-effects meta-analyses.

Heterogeneity and inconsistency across trials will be assessed using $\chi^2$ tests and quantified using the $I^2$ statistic.\(^5\)

**Exploring clinical heterogeneity (subgroup analyses and meta-regression)**

We will investigate how trial-level and (where feasible) patient-level covariates influence the effectiveness of rhBMP-2 therapy. Trials will be grouped by type of surgery as follows: ALIF; PLIF; TLIF; PLF; ACDF. Meta-analyses will be conducted within each surgery subgroup and the results across subgroups compared using tests of heterogeneity to identify differences between surgery types. Similar subgroup analyses and tests for differences across them will be undertaken to identify differences in effectiveness of treatment at differing spinal locations. If feasible, meta-regression will be used to investigate dose-response relationships.

**Individual patient data**

**Two stage modelling**

IPD will be analysed using standard two stage meta-analytic techniques. IPD from each trial will be analysed separately, using the same methods across trials, for each outcome listed above.\(^6,7\) The resultant summary statistics for each trial will then be combined using random-effects models to give overall estimates of the effect for each outcome explored. For time to event outcomes, we will apply Cox proportional hazards models to each trial, providing the proportional hazards assumption is not clearly violated.\(^8\)

**Exploring clinical heterogeneity (subgroup analyses)**

Characteristics that vary at the trial level (e.g. comparator type) will be investigated by analysing grouped trials or by meta-regression as described above for aggregate data.

Patient-level characteristics, including underlying condition (degenerative disc disease, spondylolisthesis) age, smoking history, and previous surgery at the same site will be explored (data permitting) in univariate two-stage subgroup analyses.\(^9,10\)
Characteristics such as type of surgery that may vary between trials (i.e. trials include only one type of surgery) will be analysed using whichever of the above approaches best suits the data.

**One stage modelling**

Time and data permitting, we will also analyse the IPD in a one-stage modelling framework. This will enable us to take account of multiple patient characteristics when comparing rhBMP-2 and SBGT (stratified by trial) and also enable simultaneous exploration of multiple potential interactions between treatment and patient or trial-level covariates. For these analyses, we will implement multilevel modelling approaches for binary, continuous, ordinal, and time-to-event outcomes. All IPD analyses will use complete case analysis as imputation of missing data is not feasible within the time scale of this project.

**Combining aggregate and individual participant data**

Where IPD is not available for relevant RCTs we will seek to combine aggregate data and IPD in meta-analyses using two-stage methods as described above and, time-permitting, using more advanced methods which have been described elsewhere. Results from these analyses will be compared to analyses of aggregate data only and of IPD only as a sensitivity analysis.

**Adverse events**

Adverse events will be categorised as described above and results tabulated. Where possible, rates of adverse events in rhBMP-2 surgery and in SBGT will be compared statistically. Where possible, meta-analyses of adverse events will be conducted to compare the incidence and nature of events across trials. Such analyses may be limited given the possible rarity and limited reporting of adverse events in RCTs. The underlying principles of meta-analysis will be applied even where statistical combination is not feasible: specifically, this will include evaluation of consistency of findings, direction and magnitude of effect, and strength of evidence, in addition to considerations of the quality of the evidence.

**Methodological exploration**

**Investigations of bias**

We will investigate possible sources of bias in RCTs and studies of adverse events. In particular we will explore the role of funding source (industry/non industry), publication status (published/unpublished) study design (randomised/ non randomised; controlled/uncontrolled; different observational designs) and components of risk of bias assessments using subgroup meta-analyses and/or meta-regression to compare results from studies that differ by the factor in question.

**Comparing aggregate and IPD approaches**

IPD analyses will be compared with aggregate data analyses. Where results differ we will establish which differences are attributable to data availability (unpublished trials, unreported outcomes, excluded patients) and which are attributable to the methods of analyses.
The summary statistics generated for each trial from the IPD will also be compared with the corresponding published results.

**Data disclosure timeline**

Construction of a timeline of what was reported where and when will provide background information on publication practice and the way that information about efficacy and adverse effects enter the public domain.

**Outputs and dissemination**

A comprehensive report detailing all the analyses described above will be submitted to the coordinating centre by the agreed deadline. This report will meet the requirements of the PRISMA statement for the reporting of systematic reviews and meta-analyses for the appropriate sections of the review.17

We will also disseminate the findings through high-impact journal publications, conference presentations and other relevant channels. We intend that the first publication will coincide with the release of the full report by Yale.

**References**

Dear Dr. Simmonds:

Congratulations! We are pleased to accept for publication your manuscript, M12-2603, "The safety and efficacy of recombinant human bone morphogenetic protein-2 (rhBMP-2) for spinal fusion: an individual patient data meta-analysis." A production editor will inform you when your paper has been scheduled for a specific issue and when you can expect proofs.

Annals now gives authors the opportunity to submit brief videos about their articles. These videos can increase the attention an article receives from readers, reporters for national and international news venues, and researchers. Instructions for video development and submission can be found here. Please notify Rob Blackwell: rblackwell@acponline.org if you plan to submit a video.

Please take care about the people that you talk with about your article and its planned publication in Annals. Doing so benefits you, the author, in several ways. First, it reduces the risk that another journal will learn about your article and publish a competing article before your article, which will make your work appear less novel and newsworthy. Second, some articles are of interest to the financial world. Premature disclosure of the content could appear as an attempt to aid and abet people who want to profit from insider knowledge. For these reasons, please do not talk about your article except to people that you know and trust. If you have any questions, please contact the editors.

Annals sends advance information about each issue to members of the news media and ask that they not release news stories before a specified embargo date of 5:00 p.m. Eastern time of the day before the publication date. If you are contacted prior to publication by members of the media who wish to discuss the content of your paper, we ask that you honor the embargo date. We also ask that you not distribute copies of your paper before publication. If the public relations office of your institution is planning its own publicity for your article, they are under the same constraints as the media regarding release of information and should contact the American College of Physicians' press office at 215-351-2653.

My colleagues and I very much appreciate the opportunity to publish your fine work.

Sincerely,

Christine Laine, MD, MPH

Editor