Effectiveness and harms of recombinant human bone morphogenetic protein-2 in spine fusion: a systematic review and meta-analysis

Fu R, Selph S, McDonagh M, Peterson K, Tiwari A, Chou R, Helfand M

CRD summary
This well-conducted review concluded that there was no evidence of clinical benefit from recombinant human bone morphogenetic protein-2, over bone graft, in spinal fusion, and it might be associated with important harms, making it difficult to recommend its use. These conclusions reflected the evidence presented and appear to be reliable.

Authors' objectives
To assess the effectiveness and harms of recombinant human bone morphogenetic protein-2 (rhBMP-2) in spine fusion.

Searching
Individual participant data (IPD) were obtained from 17 industry-sponsored eligible studies, and related internal documents were checked. MEDLINE (from 1996), EMBASE, The Cochrane Library, Scopus, ClinicalTrials.gov and the US Food and Drug Administration website, were searched up to August 2012. Search terms were reported. Reference lists of relevant publications were screened.

Study selection
Randomised controlled trials (RCTs) and cohort studies comparing rhBMP-2 with any control, in spinal fusion, were eligible for inclusion. Uncontrolled case series were eligible for the evaluation of harm. Studies that combined the results of rhBMP-2 with those of other bone morphogenetic proteins, where rhBMP-2 was not predominantly used, were excluded.

Most of the included trials compared rhBMP-2 with iliac crest bone graft; one compared it with artificial disc replacement. All the trials used similar eligibility criteria and recruited similar samples for each surgical approach. Most of the included patients underwent posterolateral lumbar fusion, and the others underwent anterior lumbar interbody fusion. The primary effectiveness outcomes were overall success (at 24 months) and fusion. Other outcomes were pain, disability, neurologic status, function and return to work. The adverse event data were overall adverse events (one of any type) and specific events, such as urine retention, wound infection, wound dehiscence, or possible lumbar radiculitis. Four definitions for lumbar radiculitis were used. Most of the trials were sponsored by Medtronic; one was sponsored by Norton Healthcare.

Two reviewers independently assessed studies for inclusion.

Assessment of study quality
The quality of the studies was assessed using the criteria adapted from the Cochrane Back Review Group and the US Preventive Services Task Force. The quality was rated as good, fair or poor. The strength of evidence by outcome was rated on the basis of the aggregate risk of bias, consistency, directness and precision of the evidence. The quality of Medtronic-funded studies was also assessed based on information from trial protocols and internal reports.

Two reviewers independently assessed quality, with any discrepancies resolved through consensus.

Data extraction
Data were collected for each trial to calculate mean differences and 95% confidence intervals, for continuous variables, and relative risks, with 95% confidence intervals, for binary variables. Effectiveness outcomes from IPD were recalculated using consistent definitions. Some adverse event data were extracted on event rates to enable the calculation of relative risks, with 95% confidence intervals.

One reviewer extracted the data and a second reviewer checked it for accuracy.

Methods of synthesis
For continuous outcomes, the data for all the trial participants were pooled, using a linear mixed-effects model. For binary outcomes, the data were pooled, using a generalised linear mixed-effects model for common outcomes, or a generalised linear fixed-effect model for rare outcomes. The combined mean differences or relative risks, with 95% confidence intervals, were calculated. A separate model was fitted for each time point.

The analyses were stratified by spinal area and surgical approach, for all outcomes except cancer and death. Statistical heterogeneity was assessed using the estimated between-study variance from the mixed-effects model. Imputation was used for missing data for overall success and fusion, which were defined using multiple criteria. Sensitivity analyses were conducted to assess the impact of excluding patients with any missing data, as well as including patients who were classified as failures (those who met some criteria, but had missing data for others).

Sensitivity analyses were conducted by excluding poor-quality studies and studies that used a lower rhBMP-2 concentration, as well as excluding graft site-related adverse events for the analysis of harms. For cancer, sensitivity was analysed by excluding events that were not reportable to the National Cancer Institute's Surveillance, Epidemiology and End Results Programme.

**Results of the review**

Thirteen RCTs (1,981 patients), 31 cohort studies, 47 intervention series, and 34 case series were included in the review. The total number of participants, in studies other than RCTs, was not reported. Nine trials reported follow-up rates at 24 months that were greater than 90% in both groups. For RCTs, the main source of bias was lack of blinding of surgeons, patients and outcome assessors (except for radiologic outcomes). Despite some baseline differences between patients receiving iliac crest bone graft and those receiving rhBMP-2, a pattern favouring the latter was not observed. For cohort studies, most reported baseline differences between groups or lacked the information on baseline characteristics, and they neither reported blinding of outcome assessors, nor adjusted for potential confounders.

Individual patient data were available for 465 participants for anterior lumbar interbody fusion (five RCTs) and 722 participants for posterolateral lumbar fusion (four RCTs). These trials were of fair quality, overall.

For anterior lumbar interbody fusion, there were no significant differences in the rate of overall success (RR 1.19, 95% CI 0.99 to 1.42; four RCTs) and fusion (RR 1.05, 95% CI 0.88 to 1.24; five RCTs), at 24 months after surgery, between the rhBMP-2 and iliac crest bone graft groups. There were no significant differences in back pain and leg pain scores.

For posterolateral lumbar fusion, there were no significant differences in the rate of overall success (RR 1.05, 95% CI 0.91 to 1.21; four RCTs) and fusion (RR 1.16, 95% CI 0.96 to 1.41; four RCTs), at 24 months after surgery, between the two groups. There were no significant differences in back pain and leg pain scores.

There were no significant differences in the rate of at least one adverse event (any type), between groups, at 24 months, for anterior lumbar interbody fusion (RR 0.96, 95% CI 0.85 to 1.09; five RCTs) and for posterolateral lumbar fusion (RR 1.02, 95% CI 0.95 to 1.10; four RCTs).

For anterior cervical spine fusion, rhBMP-2 was associated with an increased risk of wound complications and dysphagia at 24 months after surgery (four cohort studies), compared with controls. Overall, it was associated with an increased cancer risk (RR 3.45, 95% CI 1.98 to 6.00; five studies), but there was no significant difference at 48 months after surgery (four studies).

The results of the sensitivity analyses were generally similar to the overall results. The results for other outcomes were reported, as were the results of the assessment of reporting bias in published articles of industry-sponsored studies.

**Authors’ conclusions**

For spinal fusion, there was no evidence of clinical benefits with rhBMP-2, over bone graft, and it might be associated with important harms, making it difficult to recommend rhBMP-2.

**CRD commentary**

The review question was clear and it was supported by appropriate inclusion criteria. Several relevant databases were searched. Data from unpublished studies were included, which reduced the potential for publication bias. Individual
patient data were available for most of the eligible trials, reducing the potential for availability bias. Sufficient attempts were made to minimise bias and error in the review.

Study quality was assessed, using appropriate criteria, and sensitivity analysis was conducted to assess the impact of poor-quality studies on the results. The authors rated the evidence from included trials to be of fair quality overall. The trial characteristics were provided. Heterogeneity was assessed and appropriate methods were used to pool the results.

The authors' conclusions reflected the evidence presented, and they acknowledged that there was little evidence for each surgical approach. Overall, this review was well conducted, and the authors' conclusions appear to be reliable.

**Implications of the review for practice and research**

**Practice:** The authors stated that it was difficult to identify clear indications to recommend rhBMP-2 in spinal fusion.

**Research:** The authors stated that research was required to provide more reliable estimates of the cancer risk and other adverse events, and to identify patient populations where rhBMP-2 might be beneficial.

**Funding**

Funded by Yale University, USA, and Medtronic.

**Bibliographic details**


**PubMedID**

23778906

**Linked records**

- Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion: a meta-analysis of individual-participant data

**DOI**

10.7326/0003-4819-158-12-201306180-00006

**Original Paper URL**


**Other publications of related interest**


**Indexing Status**

Subject indexing assigned by NLM

**MeSH**

Bone Morphogenetic Protein 2 /adverse effects /therapeutic use; Conflict of Interest; Drug Industry; Humans; Ilium /transplantation; Incidence; Intervertebral Disc Degeneration /surgery; Neoplasms /epidemiology; Off-Label Use;
Publication Bias; Publishing /standards; Recombinant Proteins /adverse effects /therapeutic use; Spinal Fusion /methods; Transforming Growth Factor beta /adverse effects /therapeutic use; Treatment Outcome

**AccessionNumber**
12013033500

**Date bibliographic record published**
18/06/2013

**Date abstract record published**
03/07/2013

**Record Status**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.