WASHINGTON STATE HEALTH CARE AUTHORITY

On- and off-label uses of rhBMP-2 or rhBMP-7 for spinal fusion

Health Technology Assessment

Tuesday, February 14, 2012

Health Technology Assessment Program
676 Woodland Square Loop SE
P.O. Box 42712
Olympia, WA 98504-2712
http://www.hta.hca.wa.gov

On- and off-label uses of rhBMP-2 or rhBMP-7 for spinal fusion

Provided by:

Spectrum Research, Inc.

Final Report Prepared by:

Robin Hashimoto, PhD
Annie Raich, MPH
Emily Yoder, BA
Kara McMullen, MPH
Joseph R. Dettori, PhD MPH

With assistance from
Nora B. Henrikson, PhD MPH
Erika D. Brodt, BS
Andrea Skelly, PhD MPH

February 13, 2012

This technology assessment report is based on research conducted by a contracted technology assessment center, with updates as contracted by the Washington State Health Care Authority. This report is an independent assessment of the technology question(s) described based on accepted methodological principles. The findings and conclusions

contained herein are those of the investigators and authors who are responsible for the content. These findings and conclusions may not necessarily represent the views of the HCA/Agency and thus, no statement in this report shall be construed as an official position or policy of the HCA/Agency.

The information in this assessment is intended to assist health care decision makers, clinicians, patients and policy makers in making sound evidence-based decisions that may improve the quality and cost-effectiveness of health care services. Information in this report is not a substitute for sound clinical judgment. Those making decisions regarding the provision of health care services should consider this report in a manner similar to any other medical reference, integrating the information with all other pertinent information to make decisions within the context of individual patient circumstances and resource availability.

TABLE OF CONTENTS

TABLES

FIGURES

APPENDICES

Appendices are published in a separate document that is publicly available on the WA HTA website.

EXECUTIVE SUMMARY

Introduction

Autogenous iliac crest bone graft (ICBG) harvested from the posterior iliac crest (hip) has long been the "gold standard" for patients receiving spinal fusion. In addition to being osteogenic, osteoconductive, and osteoinductive, ICBG has the added benefit of being histocompatible and non-immunogenic. However, autogenous bone graft harvesting can result in pain, infection, nerve and artery damage, as well as an increased risk of stress factor at the harvest site. Harvest site pain is the primary motivation to pursue technologies other than autograft. The last decade has seen the development of bone morphogenetic protein (BMP) products to serve as substitutes to autograft in spinal fusion procedures that would promote the same high rate of fusion observed in patients who undergo spinal arthrodesis with ICBG but that would allow patients to avoid harvest site pain and morbidity.

To date, two rhBMPs (rhBMP-2 and rhBMP-7) and associated delivery vehicles have received approval from the FDA. InFUSE Bone Graft/LT-CAGE Lumbar Tapered Fusion Device (Medtronic), which is made of rhBMP2 on an absorbable collagen sponge carrier, has received FDA premarket approval for treatment of single level anterior open or anterior laparoscopic spinal fusion procedures in adults with degenerative disc disease (DDD) in the lumbar spine (between L4 and S1) and who have failed conservative care. OP-1 Putty (Stryker), comprised of rhBMP7 and bovine collagen, has received a humanitarian device exemption from the FDA for use in compromised patients undergoing revision posterolateral lumbar spinal fusion for whom autologous bone graft or bone marrow harvest are contraindicated or not expected to result in fusion. In addition, rhBMP2 and rhBMP7 are being used for other off-label uses in the lumbar and cervical spine.

Significant questions remain about the safety, efficacy and effectiveness and cost effectiveness of the use of rhBMP-2 or rhBMP-7 for spinal fusion procedures. The primary efficacy and effectiveness outcomes include function and pain; radiographic fusion is also often considered to be a primary outcome but serves as a surrogate outcome for patient-reported function and pain. Secondary outcomes of interest may include (but are not limited to) perioperative outcomes, patient satisfaction, return to work, medication usage. Safety questions in particular have recently arisen regarding whether the potential beneficial outcomes of BMP use outweigh their risks: the risks of the intervention, and how often complications arise. Therefore, this health technology assessment set out to answer the following key questions regarding patients undergoing spinal fusion:

Key question 1

What are the expected treatment outcomes of primary single or multilevel lumbar or cervical spinal fusion for degenerative disc disease (DDD), and of revision posterolateral lumbar spinal fusion in compromised patients (i.e., osteoporosis, smoking, diabetes)? Are there validated instruments related to outcomes in patients undergoing these

procedures? Has clinically meaningful improvement in outcomes been defined in these patient populations?

Key question 2

What is the evidence of efficacy and effectiveness of:

- rhBMP-2 (InFUSE) for on-label lumbosacral spine fusion in patients with DDD?
- rhBMP-7 (OP-1) for on-label revision posterolateral lumbar spine fusion in compromised (e.g., osteoporosis, smoking, diabetes) patients?
- rhBMP-2 (InFUSE) for off-label lumbosacral spine fusion?
- rhBMP-7 (OP-1) for off-label lumbosacral spine fusion?
- rhBMP-2 (InFUSE) for off-label cervical spine fusion?
- rhBMP-7 (OP-1) for off-label cervical spine fusion?

Including consideration of perioperative outcomes (including length of surgery) as well as short term and long term impact on function, pain, radiographic fusion, patient satisfaction, quality of life, activities of daily living and return to work, as well as other reported measures.

Key question 3

What is the evidence of the safety of on- or off-label use of rhBMP-2 or rhBMP-7 for spinal fusion compared with spinal fusion using ICBG or alternative bone graft substitutes? Including consideration of short- and long term adverse events and complications type and frequency (pain, donor site morbidity, resorption/osteolysis, heterotopic bone formation, graft subsidence, graft migration, dysphagia or respiratory difficulties, elevated antibody responses to BMPs or collagen, wound complications (infection, hematoma, seroma, or dehiscence), local or systemic toxicity, mispositioned graft, neurological complications, retrograde ejaculation, urogenital complications, allergic reactions, mortality, other major morbidity), as well as revision/re-operation risks.

Key question 4

What is the evidence that on- or off-label use of rhBMP-2 or rhBMP-7 for spinal fusion compared with spinal fusion using ICBG or alternative bone graft substitutes has differential efficacy or safety issues in sub-populations? Including consideration of gender, age, baseline functional or pain status, comorbidities (including but not limited to tobacco use, alcohol use, or psychological), other patient characteristics or evidence-based patient selection criteria, provider type, setting or other provider characteristics, or payor/beneficiary type (including worker's compensation, Medicaid, state employees).

Key question 5

What evidence of cost implications and cost-effectiveness of on- or off-label use of use of rhBMP-2 or rhBMP-7 exists? Including consideration of costs (direct and indirect) and cost effectiveness in the short term and long term.

Methods

We conducted a formal, structured systematic search of the peer-reviewed literature across a number of databases in addition to searches of pertinent databases related to clinical guidelines and previously performed assessments. Pertinent studies were critically appraised using our Level of Evidence (LoE) system which evaluates the methodological quality based on study design as well as factors which may bias studies. An overall Strength of Evidence (SoE) combines the LoE with consideration of the number of studies and consistency of the findings to describe an overall confidence regarding the stability of estimates as further research is available. Included economic studies were also formally appraised based on criteria for quality of economic studies and pertinent epidemiological precepts.

We selected articles to summarize based on the inclusion and exclusion criteria in the following table:

Study Component	Inclusion	Exclusion
Participants	Patients with back and/or leg or neck pain	◆ Skeletally immature patients (< 18 years of age) ◆ Pregnancy ◆ History of tumor in the implantation site ◆ Infection at the implantation site
Intervention	◆ FDA-approved ("on-label") and -unapproved ("off-label") implantation of rhBMP-2 (InFUSE) or rhBMP-7 (OP-1) in the lumbar or cervical spine	◆ Implantation of rhBMP-2 or rhBMP-7 into sites other than the spine ◆ Spine fusion not using rhBMP-2 or rhBMP-7
Comparators	◆ Placebo ◆ Standard care ◆ Physical therapy ◆ Autograft bone, allograft bone, bone marrow, demineralized bone matrix, stem cells, and/or other bone substitutes used to enhance bone remodeling)	
Outcomes	Perioperative outcomes Short- and long- term: ◆ Functional outcomes ◆ Pain ◆ Radiographic fusion ◆ Patient satisfaction ◆ Quality of life ◆ Activities of daily living ◆ Return to work ◆ Complications/Adverse events (safety) ◆ Reoperation (safety) ◆ Prognostic factors	◆ Non-clinical outcomes
Study Design	◆ Reliability/validity studies for question 1. ◆ Comparative studies for questions 2-4. ◆ Case series and case reports designed to evaluate adverse events for question 3. ◆ Formal economic studies will be sought for question 5	◆ Non-clinical studies
Publication	◆ Studies published in English in peer reviewed journals, published HTAs or publically available FDA reports ◆ Full formal economic analyses (e.g. cost-utility studies) published in English in HTAs or in a peer-reviewed journal published after those represented in previous HTAs.	◆ Abstracts, editorials, letters ◆ Duplicate publications of the same study which do not report on different outcomes ◆ Single reports from multicenter trials ◆ Studies reporting on the technical aspects of BMP use in fusion surgery ◆ White papers ◆ Narrative reviews ◆ Articles identified as preliminary reports when results are published in later versions ◆ Incomplete economic evaluations such as costing studies

Quality of literature available
For Key Question 1 we identified a total of 36 studies.

For Key Question 2 on efficacy/effectiveness we found a total of 14 RCTs (LoE I: 0 studies; LoE IIa: 1 study; LoE IIb: 13 studies) and 15 cohort studies (LoE II: 2 studies; LoE III: 13 studies).

For Key Question 3 on safety we found 2 additional subsets of RCTs included in Key Question 2 that presented more safety data [1, 2]), 12 cohort studies (LoE II: 1 study; LoE III: 11 studies), 33 case series, and 16 case reports. All studies included in Key Question 2 were evaluated for safety in Key Question 3. Thus Key Question 3 evaluates 14 RCTs (+ 2 additional subset analyses from these studies), 27 cohort studies, 33 case series, and 16 case reports.

Note that for Key Question 3, we identified three publically available FDA reports, all three of which had peer-reviewed publications associated. In general, we used the safety data from the peer-reviewed publication when possible. If the peer-reviewed publication did not have the outcome of interest, we looked to the FDA report. If the peer-reviewed publication and the FDA report both provide data on the outcome of interest, but the data are reported differently, we used the most conservative data; that is, we used the report in which the comparison most favored the control group. For the InFUSE FDA SSED, we noted that most of the data presented were a compilation of three different datasets: a small pilot study, a large pivotal RCT and a single arm case series. When using the FDA InFUSE SSED, if the data were segregated by dataset, we used the pivotal RCT. If not, we then looked to the pooled data that included all three datasets.

To address outcomes following fusion with versus without BMP in special populations (Key Question 4), we included 8 cohort studies (LoE II: 1 study; LoE III: 7 studies).

For Key Question 5, we identified 3 studies that met our inclusion criteria.

Results/Summary of evidence

Key Question 1: Expected treatment outcomes, validated instruments, and clinically meaningful improvement

What are the expected treatment outcomes of lumbar or cervical spinal fusion? Are there validated instruments related to outcomes in patients undergoing these

procedures? Has clinically meaningful improvement in outcomes been defined in these patient populations?

Summary: We identified four outcome measures commonly used in the comparative studies in this HTA. Only one, the SF-36, was evaluated for validity in spinal fusion patients. One study demonstrated criterion validity and internal consistency of the SF-36 in patients undergoing lumbar fusion. Responsiveness and other aspects of validity were not examined in this study. The other three outcome measures (the ODI, the NDI, and pain assessed by a VAS) have been shown to have a degree of validity, reliability, and responsiveness in various spine populations, some of which might be eligible for fusion. The minimal clinically important difference (MCID) was variously defined in fusion patients for the ODI in several studies; in characterizing the ODI results here we will use the FDA accepted MCID for ODI of 15 points. The MCID for pain as measured by the 10 cm VAS has been defined as 0.2 to 2.9 cm. For patients with chronic LBP, a 2.0 cm improvement for pain has been recommended by some as the minimal clinically important difference. For this report, we will use the latter value for defining the MCID for VAS scores, which corresponds to a difference of 4.0 points on the more commonly reported 20-point scale. There is some cause for concern surrounding the definition of MCID for ODI, including the wide range and variability of reported values, the feasibility of the MCID being able to detect an improvement considered important to patients, and the various calculation methods of the MCID.

Key Question 2: Efficacy and effectiveness

What is the evidence of efficacy and effectiveness of:

a) rhBMP-2 (InFUSE) for on-label lumbosacral spine fusion in patients with DDD?

b) rhBMP-7 (OP-1) for on-label revision posterolateral lumbar spine fusion in compromised (e.g., osteoporosis, smoking, diabetes) patients?

c) rhBMP-2 (InFUSE) for off-label lumbosacral spine fusion?

d) rhBMP-7 (OP-1) for off-label lumbosacral spine fusion?

e) rhBMP-2 (InFUSE) for off-label cervical spine fusion?

f) rhBMP-7 (OP-1) for off-label cervical spine fusion?

Including consideration of perioperative outcomes (including length of surgery) as well as short term and long term:

- Impact on function, pain, radiographic fusion, patient satisfaction, quality of life, activities of daily living and return to work
- Other reported measures

rhBMP-2 on-label use: lumbar spine

EFFICACY

Studies:

Two LoE IIb RCTs were identified that met our inclusion criteria. Study size ranged from 14 to 279 patients. These two studies served as the pilot and pivotal trials in the 2002 FDA Summary of Strength and Effectiveness Data (SSED) for InFUSE (P000058), and both studies were sponsored by Medtronic. Patients were followed for 24 months. Because the studies were similar in design, we were able to pool outcomes data from both studies. Patients with DDD, radiculitis, and/or up to 25% spondylolisthesis and who were refractory to conservative care underwent primary single-level open anterior lumbar fusion with either rhBMP-2/ACS (InFUSE) (n = 154) or iliac crest bone autograft (ICBG) (n = 139). RhBMP-2 was used at a dose ranging from 4.2 to 8.4 mg per patient. Additional details are available in Table 8 and the surrounding text.

Summary:

Perioperative outcomes (Table 9):

Operative time: The mean length of operative time was similar in both groups (1.6 versus 2.0 hours for rhBMP-2 versus ICBG) (2 RCTs). The strength of this evidence is *low*.

Blood loss: The mean perioperative blood loss was lower in the rhBMP-2 group compared with the ICBG group (108.9 versus 153.3 mL) (2 RCTs). The strength of this evidence is *low*.

Length of hospital stay: The mean length of hospital stay was similar in both groups (3.0 vs. 3.3 days for rhBMP-2 versus ICBG) (2 RCTs). The strength of this evidence is *low*.

Fusion (Table 10): The percentages of patients with successful fusion were similar in both treatment groups at all follow-ups. The strength of this evidence is *low*. By 24 months, 94.2% of rhBMP-2 and 88.5% of ICBG patients had successful fusion.

ODI (Table 11): ODI outcomes were similar between groups at all reported follow-ups (2 RCTs). The strength of this evidence is *low*. At 24 months, 84.4% of rhBMP-2 and 82.0% of ICBG patients had ODI "success", which was defined as improvement from baseline by at least 15%. Mean score improvements at 24 months were 29.6 and 23.7 points, the difference between which is not considered clinically meaningful.

Pain (Table 11): There were not clinically meaningful differences between groups in back and leg pain VAS scores as reported by one RCT (N = 279). The strength of this evidence is *low*. The percentage of patients with back pain "success", which was defined as improvement by > 3 points from baseline, was similar between groups at all follow-ups between 1.5 and 24 months. At 24 months, 75% and 79% of patients

in the rhBMP-2 and ICBG groups had achieved back pain "success", while the mean VAS score improvement was 8.5 and 8.2 points, respectively. Similarly, the proportion of patients with leg pain "success" was similar between groups at 24 months (80% versus 74%, respectively. The 24-month mean improvement in leg pain VAS scores was similar at 1.5 months and identical in both groups by 24 months (6.2 points).

SF-36: function (Table 11): There was no difference in function as measured by the SF-36 between groups as reported by one RCT (N = 14). The strength of this evidence is *low*. The mean improvement in SF-36 physical function subscale scores was similar in both groups at all reported follow-ups between 3 and 24 months (38 vs. 37 for rhBMP-2 vs. ICBG, respectively at 24 months).

Patient satisfaction (Table 11): Patient satisfaction rates were similar in both groups at 24 months (2 RCTs). The strength of this evidence is *low*.

Work status (Table 11): There was no difference in work status between groups as reported by two RCTs. The strength of this evidence is *low*. At 24 months, the percentage of patients who were working was similar in both groups: 67.5% (104/154) compared with 56.1% (78/139) of patients in the pooled rhBMP-2 and ICBG groups, respectively.

Neurological status (Table 12): As reported by one RCT (N = 279), the percentage of patients with neurological success was similar between groups at all reported follow-ups (3 to 24 months). The strength of this evidence is *low*. At 24 months, 83% and 84% of those in the BMP and control groups had neurological success.

EFFECTIVENESS

Studies:

One integrated analysis (LoE II) met our inclusion criteria. The analysis retrospectively pooled data from 679 patients that was reported in three studies, including one RCT (reported in efficacy) (n = 279), one case series (n = 22), and one unpublished study (n = 378). *Note that patients from the Burkus RCT make up 41% of the population reported in the integrated analysis (279/679 patients); thus results in this section partially overlap with those reported above*. A weakness of this integrated analysis is that more than half of the data (56%) were taken from an unpublished cohort study. Patients with DDD and radiculitis who were refractory to conservative care underwent primary single-level open (41%) or laparoscopic (59%) anterior lumbar fusion with either rhBMP-2/ACS (InFUSE) (n = 277) or iliac crest bone autograft (ICBG) (n = 402). Patients were followed for 24 months. Additional details are available in Table 13 and the surrounding text.

Summary:

Perioperative outcomes (Table 14): The following three outcomes were better in the rhBMP-2 group than in the ICBG group: operating time (1.8 versus 2.7 hours), blood

loss (127 versus 193 mL), and length of hospital stay (2.2 versus 3.1 days). The strength of this evidence is *low*.

Fusion (Table 14): The percentage of patients with successful fusion was similar between groups at 6, 12, and 24 months. The strength of this evidence is *low*. By 24 months, 94% of the BMP and 89% of the ICBG patients had successful fusion.

ODI (Table 14): At six months and later, there were no clinically meaningful differences between groups in ODI outcomes. The strength of this evidence is *low*. At three months, the BMP group did have clinically meaningful improvement in ODI scores while the ICBG group did not (mean scores improvement of 31 versus 5 points, respectively). At six, 12, and 24 months, however, both groups had clinically meaningful improvement, and the differences between groups were not clinically meaningful (mean score improvement at 24 months: 31 versus 26 points, respectively).

SF-36: pain and function (Table 14): SF-36 pain and function outcomes were similar between groups. The strength of this evidence is *low*. While the mean score improvements in SF-36 pain index and physical component subscale scores were higher in the rhBMP-2 versus the ICBG group at all follow-ups, it is unlikely that these differences are clinically meaningful. The mean score improvement at 24 months for the pain index subscale was 39 (BMP) versus 33 (ICBG) points, and for the physical component subscale it was 16 (BMP) versus 12 (ICBG) points.

Work status (Table 14): A similar percentage of patients had returned to work by 24 months (75% versus 65% for rhBMP-2 versus ICBG, respectively). However, those in the rhBMP-2 group returned to work a median of 55 days sooner than those in the control group. The strength of this evidence is *low*.

rhBMP-7 on-label use: lumbar spine

No studies were identified that evaluated on-label use of rhBMP-7 (OP-1).

Although OP-1 has received a humanitarian device exemption (HDE) from the FDA (H020008) for "use as an alternative to autograft in compromised patients [i.e., osteoporotic, smokers, diabetics] requiring revision posterolateral (intertransverse) lumbar spinal fusion, for whom autologous bone and bone marrow harvest are not feasible or are not expected to promote fusion,"[3] the pilot and pivotal trials evaluated primary (not revision) posterolateral fusion patients and these trials are therefore not in accordance with on-label use of OP-1.

rhBMP-2 off-label use: lumbar spine

EFFICACY

Studies:

Six RCTs met our inclusion criteria. One study received an LoE grade of IIa, and the remaining five studies were graded LoE IIb. Study size ranged from 27 to 463 patients. Patients with DDD, radiculitis, and/or up to 25% spondylolisthesis who were refractory to conservative care underwent primary single- (or in one study, multi-) level posterior (four studies), anterior (one study), or posterolateral (one study) lumbar fusion with either rhBMP-2/ACS (InFUSE) or iliac crest bone autograft (ICBG). Due to heterogeneity in surgical procedures (i.e., approach, use of ceramic granules, use of cage versus allograft dowel versus no device, single- versus multilevel design), we were not able to pool outcomes data from the six studies. Patients received BMP in a variety of forms: rhBMP-2/CRM; InFUSE; and AMPLIFY. Doses of rhBMP-2 varied and ranged from 4.2 to 40 mg per patient (when reported). Patients were followed for 17 (mean) to 24 months. Five RCTs were sponsored by Medtronic and one by a Norton Healthcare grant. Additional details are available in Table 15 and the surrounding text.

Summary:

Perioperative outcomes (Table 15):

Operative time: Operative times were similar or improved with rhBMP-2 compared with control treatment (6 RCTs). The strength of this evidence is *moderate*. Three of the studies reported similar mean operative times for both groups, with the difference between groups ranging from 0.1 to 0.4 hours). The other three reported statistically shorter operative time for patients receiving rhBMP2 compared with ICBG (difference between groups ranged from 0.4 to 0.5 hours).

Blood loss: Mean perioperative blood loss volumes were similar or improved (ie., lower) with rhBMP-2 compared with control treatment (6 RCTs). The strength of this evidence is *moderate*. Four studies reported similar mean blood loss between groups, with differences ranging from 5 to 123 mL. Two studies reported statistically lower mean blood loss in the rhBMP-2 group, with differences ranging from 98 to 106 mL, although it is unclear whether these differences are clinically meaningful.

Length of hospital stay: There was no difference between groups in the mean length of hospital stay (5 RCTs). The strength of this evidence is *high*. Mean length of stay ranged from 2.9 to 4.1 days in the rhBMP-2 groups and from 3.3 to 5.2 days in the control group.

Fusion (Table 17): Fusion rates were similar or improved with rhBMP-2 versus ICBG treatment (6 RCTs). While there may be a slight improvement in fusion

success in those treated with rhBMP-2 compared with ICBG, it is not clear that the differences are clinically meaningful. The strength of this evidence is *moderate*. With six studies reporting, the proportions of patients with successful fusion at 24 months ranged from 86-100% in the rhBMP-2 group and from 67-89% in the ICBG group (the latter excludes one study with only 5 patients in the control group; otherwise the range is 40-89% for the ICBG group). More specifically, outcomes were reported as similar or improved: three studies reported that the percentage of patients with successful fusion was similar between treatment groups at all follow-ups, while another three studies reported statistically meaningful improvements in the rhBMP-2 versus the ICBG groups.

ODI (Table 18): In general, ODI outcomes were similar between groups at all reported follow-ups in terms of both the percentage of patients with ODI "success" (defined as improvement from baseline scores by at least 15% to 20%, depending on the study) (reported by three RCTs) and the mean ODI score improvement from baseline (reported by six RCTs). The strength of this evidence is *high*. Regarding ODI "success", all three studies reporting statistically similar incidences of "success" at last follow-up (17-24 months), with the proportion of patients with ODI "success" ranging from 2 to 21% higher in the BMP. Regarding mean score improvement, in five RCTs there was clinically meaningful improvement (ie., improvement by at least 15 points) in both treatment groups at 24 months, but the difference between the groups was not clinically meaningful. In one RCT (N = 46), the BMP group had clinically meaningful improvement in their mean ODI scores while the ICBG group did not; however, the difference in mean score improvement was only 2 points (15 versus 13 points, respectively).

Pain (Table 18): Back pain outcomes were clinically similar in five RCTs and clinically improved in one RCT with rhBMP-2 treatment at final follow-up of 17 to 24 months. The strength of this evidence is *moderate*. Mean improvements in back pain scores at 17-24 months ranged from 3.1 to 9.6 points in the rhBMP-2 groups and from 3 to 8 points in the ICBG groups; differences between groups ranged from 0.1 to 4.5 points (6 RCTs). One RCT (N = 67) reported clinically and statistically improved 24 month scores in rhBMP-2 versus the ICBG groups (9 versus 4.5 (of 20 possible) points, respectively). Clinically meaningful improvement is considered to be 20mm on a 100 cm scale[4], which translates to 4 points on the 20 point scales used here. Regarding leg pain, all six RCTs reported similar clinical improvements in leg pain VAS scores between groups at final follow-up of 17 to 24 months. The strength of this evidence is *high*. Mean VAS score improvements in leg pain ranged from 3.6 to 9.3 points in the rhBMP-2 groups and from 3.1 to 7.2 points in the ICBG groups; differences between groups ranged from 0.5 to 2.1 points.

SF-36: function (Table 18): Outcomes were similar between groups; the strength of this evidence is *high*. All six RCTs reported improvements in the SF-36 physical component subscale for both the rhBMP-2 and ICBG treatment groups, which ranged from 7 to 15 points in the rhBMP-2 and 7 to 17 points in the ICBG groups at final

follow-up (17 to 24 months). Regarding the differences between groups in mean score improvement, five RCTs reported no differences between treatment groups at 17 to 24 months, while one RCT reported that the rhBMP-2 group had statistically better improvements compared with the ICBG group at 6, 12, and 24 months ($P \leq .02$). In this case, however, it is unlikely that the small score differences reported between treatment groups (of 6, 8, and 3 points at each follow-up, respectively) are clinically meaningful.

Patient satisfaction: Patient satisfaction was similar between treatment groups at 17 and 24 months as reported by two RCTs (N = 27 – 67). The strength of this evidence is *low*.

Work status: There were no differences in work status between treatment groups as reported by four RCTs. The strength of this evidence is *high*. Two RCTs reported that similar percentages of patients between groups were working or had returned to work in both treatment groups at 24 months follow-up, and two other RCTs reported no difference in mean time to return to work between treatment groups.

Neurological status: There were no differences in neurological success between groups at 24 months as reported by one RCT (N = 67). The strength of this evidence is *low*.

Overall success: The composite measure of "overall success" was reported by one RCT (N = 41) and defined as a combination of successful fusion, ODI success, an absence of severe adverse events, an absence of secondary surgical procedures at the index level, and maintenance or improvement in neurological status. The incidence of "overall success" was similar between groups at 24 months. The strength of this evidence is *low*.

EFFECTIVENESS

Studies:

Eight cohort studies met our inclusion criteria. Of these eight studies, there were two prospective cohort studies, one prospective case control study, three retrospective cohort studies, and two retrospective cohort studies with historical controls. One prospective cohort study received an LoE grade of II, while the remaining seven studies were graded LoE III. Study size ranged from 36 to 126 patients. Patients were followed for a mean of 9 to 39 months. Six of the cohort studies included patients with DDD, radiculitis; some of these studies also include those with up to grade 1 or 2 spondylolisthesis, scoliosis, instability, nonunion, or adjacent segment degeneration. One study treated patients with symptomatic pseudarthrosis following previous PLIF for DDD; one study evaluated patients with scoliosis with degeneration distal to a prior long idiopathic scoliosis fusion site. Patients underwent primary or revision single- or multi-level anterior (two studies), posterior (two studies), transforaminal (one study), or posterolateral (three studies) lumbar fusion

with rhBMP-2 or iliac crest bone autograft (ICBG), allograft chips, or local or rib autograft. Doses of rhBMP-2 varied and ranged from 3 to 36 mg per patient (when reported). Due to heterogeneity in control treatment, patient diagnosis, and surgical procedures (i.e., approach, use of local autograft or ICBG or bone graft extenders, use of cage versus allograft dowel versus no device, single- versus multilevel design, primary versus revision surgery), we were not able to pool outcomes data. Study sponsorship was reported as follows: Medtronic (1 study), Medtronic and Norton Healthcare grants (1 study), no funding (2 studies), no direct funding but benefits may have been received (1 study), or funding not reported (3 studies). Additional details may be found in Table 19 and the surrounding text.

Summary:

Perioperative outcomes (Table 20):

Operative time: There were no differences in mean operative time between groups as reported by one study (N = 64). The strength of this evidence is *insufficient*.

Blood loss: Blood loss was lower in the rhBMP-2 group compared with the autograft group as reported by one study (N = 64) (mean blood loss of 1221 versus 1938 mL, respectively). The strength of this evidence is *insufficient*.

Length of hospital stay: There was no evidence on length of hospital stay.

Fusion (Table 21): Fusion outcomes were similar between rhBMP-2 and control groups at final follow-up (7 studies), while one study reported that outcomes were improved with rhBMP-2. The strength of this evidence is *low*. Specifically, the proportion of patients with successful fusion was similar between rhBMP-2 and autograft groups by final follow-up in five of six studies reporting. One study (N = 50) reported that more patients had successful fusion following PLF with rhBMP-2 versus ICBG (94% versus 77%), however the control group only had 11 patients. One study (N = 75) reported that fusion rates were higher following ALIF with rhBMP-2 compared with allograft chips (99% versus 82% at 24 months).

ODI (Table 22): There were no clinically meaningful differences in mean ODI score improvement at final follow-up (24 to 61 months) as reported by two studies (N = 64 – 75). The strength of this evidence is *insufficient*. The differences in mean score improvement ranged from 3 points (favoring rhBMP-2 over allograft chips) to 3 points (favoring autograft).

Pain (Table 22): There were no differences between groups in various reported pain outcomes (back pain VAS, leg pain VAS, unspecified pain VAS, Prolo Scale Pain Subscale, and the SRS-30 Pain Subscale) at final follow-up as reported by five studies. The strength of this evidence is *low*.

Function (Table 22): There were no differences between treatment groups in functional improvement as reported by two cohort studies. One study (N = 44) reported function using the Prolo Scale Functional Subscale, while the other (N = 64) utilized the SRS-30 Functional Subscale. The strength of this evidence is *insufficient*.

Patient satisfaction (Table 22): Patient satisfaction was similar between treatment groups as reported by two studies (N = 64 – 75). The strength of this evidence is *insufficient*.

Medication use (Table 22): Medication usage was similar in both treatment groups at a mean of 8 to 11 months as reported by one study (N = 44). The strength of this evidence is *insufficient*.

Mental health/self image (Table 22): Both mental health and self-image were similar between treatment groups at a mean of 40 to 61 months follow-up as reported by one study (N = 64). The strength of this evidence is *insufficient*.

Overall outcome (Table 22): There were no differences in the percent of patients between treatment groups who considered themselves to have "good" or "excellent" outcomes as reported by one study (N = 75). The strength of this evidence is *insufficient*.

rhBMP-7 off-label use: lumbar spine

EFFICACY

Studies: Five RCTs met our inclusion criteria, all of which were graded LoE IIb. Study size ranged from 20 to 293 patients. Patients with degenerative (or in one study, isthmic) spondylolisthesis up to grade 1 (or 2) who had not responded to six months of nonsurgical treatment underwent primary single- level posterior (four studies) or posterolateral (one study) lumbar fusion with either OP-1 (rhBMP-7) or iliac crest bone autograft (ICBG) (four studies) or autograft (1 study). RhBMP-7 was used at a dose of 7 mg per patient. The two studies by Vaccaro et al. were similar in design and length of follow-up; data were pooled from these studies when helpful. The remaining three studies were heterogeneous in design and patient characteristics and thus data from these RCTs were not pooled. The mean length of follow-up ranged from 12 to 54 months. Studies were sponsored as follows: Stryker Biotech (1 RCT); funding received but source not stated (2 RCTs); no direct funding but benefits may have been received (2 RCTs). Additional details are available in Table 23 and in the surrounding text.

Summary:

Perioperative outcomes (Table 24):

Operative time: Operative time was shorter or similar for patients treated with OP-1 compared with ICBG (3 RCTs). The strength of this evidence is *low*. More specifically, one large RCT (N = 293) reported statistically lower operative time in the OP-1 group than in the ICBG group (2.4 versus 2.7 hours, respectively; *P* = .006). The two smaller RCTs (both with N = 36) reported no difference in mean operative times between groups.

Blood loss: Blood loss was lower or statistically similar for patients who received OP-1 versus ICBG as reported by two RCTs. The strength of this evidence was *low*. One study reported statistically lower blood loss in OP-1 versus ICBG patients (difference of 162 mL) (N = 293), while the other study reported statistically similar volumes of blood loss (difference of 49 mL) (N = 36).

Length of hospital stay: There was no difference in the mean length of hospital stay between treatment groups as reported by three RCTs (N = 36 – 293); though the large RCT did not report data, only that there was no difference between treatment groups (*P* = .529). The strength of this evidence is *high*. The mean length of hospital stay (in the two studies reporting the data) ranged from 3.9 to 10.5 days in the OP-1 group compared with 4.3 to 10.9 days in the ICBG group.

Fusion (Table 25): Overall, there were no differences in fusion success between treatment groups as reported by all five RCTs at 12-48 months follow-up (N = 20 – 293). The strength of this evidence is *high*.

ODI (Table 26): Both the percentage of patients with ODI "success" (defined as improvement from baseline scores by at least 20%) (as reported by two RCTs (N = 36 – 293)) and the mean ODI score improvement from baseline (as reported by three RCTs (N = 20 – 293)) were similar between groups at all reported follow-ups. The strength of this evidence is *high*. At 36 to 48 months, 69 to 74% of patients had ODI "success", compared with 57 to 77% of control patients. Regarding mean ODI score improvement, OP-1 and control treatment groups appeared to have clinically meaningful improvement at 12 to 36 month follow-ups, but the differences of 3 to 7 points between the treatment groups were not clinically meaningful.

Pain (Table 26): There were no differences in back or leg pain outcomes as reported by one study each. The strength of this evidence is *low*. Specifically, similar percentages of patients in both treatment groups had no back pain at 12 months follow-up as reported by one small RCT (N = 20). Another RCT reported no difference in the mean VAS score improvement for leg pain at 36 months follow-up between the OP-1 and ICBG treatment groups (N = 293) (3.2 versus 2.8 (on a 10-point scale), respectively).

SF-36: function (Table 26): The mean improvement in SF-36 physical component subscale scores was reported to be similar in both treatment groups at 36 months by one RCT (N = 293), however, no data were reported. The strength of this evidence is *low*.

Neurological success (Table 27): Neurological success was similar in both OP-1 and ICBG treatment groups at 36 months follow-up (or longer) as reported by one RCT (N = 293). The strength of this evidence is *low*.

Overall success (Table 27): The percentage of patients who achieved the composite measure of overall success was similar in both treatment groups as reported by two RCTs at 36 to 48 months follow-up (N = 36 – 293). The strength of this evidence is

low. This composite measure required ODI success (improvement by 20%), lack of device-related serious adverse events, and radiographic fusion; the smaller RCT additionally required maintenance or improvement in neurologic fusion. Percentages of patients with overall success ranged from 47-62% (mean 48.4% (88/182) in the OP-1 group and from 33-47% (mean 45.9% (34/74)) in the ICBG group.

EFFECTIVENESS

No studies were identified that evaluated the effectiveness off-label use of rhBMP-7 (OP-1).

rhBMP-2 off-label use: cervical spine

EFFICACY

Studies:

One RCT[5] was identified that met our inclusion criteria (LoE IIb). Thirty-three patients with cervical disc disease with radiculopathy and/or myelopathy underwent one- or two-level plated anterior cervical discectomy with fusion (ACDF) with either rhBMP-2 (n = 18) or ICBG (n = 15). BMP was used at a dose of 0.6 to 1.2 mg per patient. Patients were followed for 24 months. The study reported no direct funding, but benefits may have been received. Additional details may be found in Table 30 and the surrounding text.

Summary:

Perioperative outcomes (Table 31): All perioperative outcomes (operative time, blood loss, and length of stay) were similar in both treatment groups. The strength of this evidence is *low*.

Fusion (Table 32): Fusion was identical in both treatment groups at all time points, with 100% of patients achieving fusion success at 6, 12, and 24 months post-ACDF. The strength of this evidence is *low*.

NDI (Table 33): The mean score improvement in the NDI was higher in patients who received rhBMP-2 compared with ICBG at 24 months post-ACDF (53 versus 37 points, respectively). The strength of this evidence is *low*. This score difference of 16 points is likely clinically meaningful in neck pain patients, although we did not find an accepted definition of MCID for NDI in spine fusion patients. There were no differences in NDI score improvements between groups at 6 or 12 months.

Pain (Table 33): Neck pain outcomes were clinically similar in both treatment groups at all reported follow-ups (1.5 to 24 months). At 24 months, 100% of patients in both treatment groups were considered to have neck pain "success" (defined as an improvement in VAS score by at least 3 points from baseline). The strength of this

evidence was *low*. Arm pain was clinically improved in rhBMP-2 patients compared with those who received ICBG at 24 months (VAS scores of 14 versus 8 (on a 20-point scale). The strength of this evidence is *low*.

Function: SF-36 (Table 33): There were no differences in the SF-36 physical component subscale scores between treatment groups at 6, 12, or 24 months following ACDF with rhBMP-2 or ICBG. The strength of this evidence is *low*.

Patient satisfaction (Table 33): Patient satisfaction was similar in both groups at 24 months (92-93%). The strength of this evidence is *low*.

Mental health: SF-36 (Table 33): Mean score improvement in the SF-36 physical component subscale scores were similar between treatment groups at 6, 12, or 24 months following ACDF with rhBMP-2 or ICBG. The strength of this evidence is *low*.

Neurological status (Table 34): The percentage of patients who achieved neurological success was similar for both treatment groups at 6, 12, and 24 months. The strength of this evidence is *low*. By 24 months, 100% of patients in both groups were considered to have neurological success, which was defined as maintenance or improvement in both motor and sensory function.

EFFECTIVENESS

Studies:

Five cohort studies met our inclusion criteria, including one prospective cohort study, three retrospective cohort studies, and one retrospective case-control database study. All were graded LoE III. Study size ranged from 58-775 patients. Two of the cohort studies included patients with DDD; another included patients with DDD, herniated nucleus pulposus, or stenosis. A fourth study treated patients for stenosis, spondylosis, or nonunion from a previous fusion. The fifth study did not report patient diagnoses. Patients underwent primary or revision single- or multi-level anterior (two studies), posterior (two studies) cervical fusion with rhBMP-2 or iliac crest bone autograft (ICBG) (two studies), allograft and demineralized bone matrix (one study), a combination of autograft and/or allograft materials (one study). One study did not report surgical approach or the details of the control treatment (referred to as "non-BMP"). BMP was used a dose that ranged from 0.9 to 12 mg per patient (when reported). Due to heterogeneity in control treatments and surgical procedures (i.e., approach, use of local autograft or ICBG or allograft, single- versus multilevel design, primary versus revision surgery), we were not able to pool outcomes data. Patients were followed for 1 to 36 months. Studies were sponsored as follows: no funding received (1 study); funding received but source not stated (1 study); no direct funding but benefits may have been received (2 studies), and funding not reported (1 study). Additional details are available in Table 35 and in the surrounding text.

Summary:

Perioperative outcomes (Table 36):

Operative time: There were no differences in the mean operative time between groups as reported by two studies (N = 66 – 77). The strength of this evidence is *insufficient*.

Blood loss: Mean blood loss was similar in both treatment groups as reported by three studies. The strength of this evidence is *low*.

Length of hospital stay: Length of hospital stay was similar between treatment groups in four (N = 58 – 204) of five studies reporting, while one large study (N = 775) reported a longer postoperative stay for those patients treated with rhBMP-2 (7.2 days) than those who did not receive rhBMP-2 (4.3 days, n = 156). The strength of this evidence is *insufficient.*

Fusion (Table 37): Use of rhBMP-2 was associated with similar or higher rates of fusion as reported by two cohort studies. The strength of this evidence is *insufficient*. One study (N = 58) reported no difference in fusion rates between the rhBMP-2 and allograft groups at up to 24 months follow-up (100% versus 96%, respectively), while the other study (N = 204) reported higher rates of fusion in patients who received rhBMP-2 (100%) versus autograft and/or allograft (87.6%).

ODI (Table 38): ODI outcomes were clinically similar in both treatment groups at all follow-ups as reported by two studies (N = 58 – 66). The strength of this evidence is *insufficient* low. At final follow-up (24 to 36 months), the difference in mean score improvement between treatment groups ranged from 1 to 9 points.

Pain (Table 38): Neck pain outcomes were similar in two studies and worse in one study following fusion with rhBMP-2 compared with control. The strength of this evidence is *insufficient*. Specifically, two studies reported differences in neck pain VAS scores between treatment groups that ranged from 0 to 2 points at 24 to 36 months follow-up (N = 58 – 66); these differences are not clinically meaningful. One study reported that more patients in the rhBMP-2 group had persistent neck pain at a mean of 24 months compared to those in the control group (48% versus 23%) (N = 204). Arm pain improvement was similar in both treatment groups at all follow-ups out to 24 to 36 months as reported by two studies (N = 58 – 66). The strength of this evidence is *insufficient*.

There were no differences in Nurick or ASIA scores (1 study), improvement in neurological deficits (1 study), use of narcotic medications (1 study), or patient-reported success (1 study).

rhBMP-7 off-label use: cervical spine

No studies were identified that evaluated efficacy or effectiveness off-label use of rhBMP-7 in the lumbar spine.

Key Question 3: Safety

What is the evidence of the safety of on- or off-label use of rhBMP-2 or rhBMP-7 for spinal fusion compared with spinal fusion using ICBG or alternative bone graft substitutes? Including consideration of:

- Short- and long term adverse events and complications type and frequency (pain, donor site morbidity, resorption/osteolysis, heterotopic bone formation, graft subsidence, graft migration, dysphagia or respiratory difficulties, elevated antibody responses to BMPs or collagen, wound complications (infection, hematoma, seroma, or dehiscence), local or systemic toxicity, mispositioned graft, neurological complications, retrograde ejaculation, urogenital complications, allergic reactions, mortality, other major morbidity).
- Revision/re-operation rates

Overgrowth and uncontrolled bone formation

Summary:

On-label use: There were no on-label comparative studies reporting on this outcome.

Off-label use: The risk of uncontrolled bone formation varied widely among three RCTs and four cohort studies assessing off-label use of rhBMP. While the majority of studies reported no cases of uncontrolled bone formation in either the rhBMP or control groups, one RCT identified an incidence of 75% in the rhBMP group compared with 13% in the control group two years after treatment. Some of the differences among studies may be a result of whether the studies assessed uncontrolled bone formation using standard radiography or computerized tomography (CT). Due to the high variability in the results of this outcome, the strength of evidence for these estimates is *low*.

Data are summarized in Table 40.

Osteoclast activity

Summary:

On-label use: The occurrence of resorption, osteolysis, or graft subsidence/migration/loosening occurred infrequently in both treatment groups in the FDA pilot and pivotal RCTs for InFUSE, 1.3% in the rhBMP group and 0.0% in the control groups. The strength of evidence for these estimates is *low*.

Off-label use: Three RCTs consistently reported similar risks of subsidence or migration between the rhBMP-2 and control groups with risks ≤ 6% in each group. One cohort study in patients with a variety of indications used spinal levels as the unit of measure. That study reported a high subsidence risk of 62% of the levels in the rhBMP-2 and 10% of the levels in the control group. The strength of evidence that the risk of resorption, osteolysis or graft subsidence is similar between groups in off-label use is *moderate*.

Data are summarized in Table 41.

Wound infections

Summary:

On-label use: The strength of evidence for risk of superficial wound infection is *insufficient* with respect to on-label use of rhBMP-2 compared with autologous bone graft. There was only one very small pilot study (n=14) reporting on this outcome.

Off-label use: The risk of superficial wound infections (including superficial infection, dehiscence, edema, and superficial hematoma or seroma) was low (<10%) and similar between treatment groups as reported by two RCTs and five cohort studies. The strength of evidence for these results is *moderate*.

All data are summarized in Table 42.

Infection, seroma, or hematoma (type unspecified)

Summary:

Many studies reported infection, seroma, or hematoma but did not specify whether these were superficial or deep infections.

On-label use: The risk of infection, seroma or hematoma not specified as superficial or deep in the FDA trial for InFUSE was similar for the rhBMP and control groups, 12.2% and 11.5%, respectively. The strength of evidence for these estimates is *low*.

Off-label use: The risk of infection, seroma or hematoma not specified as superficial or deep was reported in four RCTs and two cohort studies of off-label use of rhBMP. While the risks varied from study to study (from 0% to 20% depending on the study), they were similar between the rhBMP and control groups. The strength of evidence for these results is *moderate*.

Data are reported in Table 43.

Deep infection or epidural hematoma and/or surgical evacuation

Summary:

On-label use: There were no on-label comparative studies reporting on this outcome.

Off-label use: There were no differences in the risk of this outcome between rhBMP-2 and control groups with respect to deep infections; the risks across one RCT and four cohort studies were ≤10% in each group. There were no reports of long-term sequelae resulting from deep infection. We did not identify any studies evaluating rhBMP-7 use that reported on deep infection or surgical evacuation. The strength of evidence for these results is *low*.

All data are summarized in Tables 44 and 45.

Dysphagia

Summary:

On-label use: The strength of evidence for dysphagia is *insufficient* with respect to on-label use of rhBMP-2 compared with autologous bone graft. One cohort study (FDA summary on InFUSE) reported "respiratory" complications with 1.7% having this complication among those in the rhBMP-2 compared with 8.6% in the control group.

Off-label use: One RCT reported "respiratory" complications in approximately 5 to 6% of patients in both treatment groups (rhBMP-2 and controls). When rhBMP-7 was used for off-label indications in the lumbar spine, two RCTs and the FDA SSPB for rhBMP-7 reported low risks of "respiratory, thoracic, and mediastinal" complications in both treatment groups. One large retrospective database study reported that a similarly low percentage of rhBMP patients experienced dysphagia or hoarseness as the control group following primary or revision fusion in the lumbar spine. The strength of evidence that the risk of dysphagia and respiratory difficulties in off-label use in the lumbar spine is similar between rhBMP and control groups is *moderate.*

In studies that evaluated rhBMP-2 use in cervical spine fusion, risks of dysphagia were consistently higher in the rhBMP group in four cohort studies, pooled risk of 34.9% in rhBMP-2 patients compared with 9.2% in the control patients. Two database studies reported statistically higher risks of dysphagia in patients who underwent cervical fusion with versus without BMP; one of these studies found that the difference in dysphagia risks was statistically meaningful in those who underwent anterior but not posterior cervical fusion. The strength of evidence that the risk of dysphagia is higher with the use of rhBMP versus control in the cervical spine is *moderate*.

All data are summarized in Table 46.

Retrograde ejaculation

Summary:

On-label use: The strength of evidence for retrograde ejaculation is *low* regarding on-label use of rhBMP-2 compared with autologous bone graft. The evidence base for on-label use consists of one cohort study (FDA summary on InFUSE) that reported a higher risk of retrograde ejaculation in the rhBMP-2 group, 7.9% versus 1.4%.

Off-label use: One retrospective cohort study gathered data from patients who underwent 1- or 2-level ALIF spanning L5/S1 and identified a 12-fold increase of retrograde ejaculation in those receiving rhBMP-2, 7.2% compared with 0.6%. The strength of evidence in off-label use regarding retrograde ejaculation is *low.*

All data are summarized in Table 47.

Ileus/ bowel obstruction

Summary:

On-label use: The strength of evidence for ileus/bowel obstruction is *insufficient* with respect to on-label use of rhBMP-2 compared with autologous bone graft. There was only one very small pilot study (n=11) reporting on this outcome.

Off-label use: The strength of evidence for ileus/bowel obstruction is *insufficient* with respect to off-label use of rhBMP-2 compared with autologous bone graft. There was only one retrospective cohort study that reported low risks of ileus in both the rhBMP-2 (1%) and ICBG (3%) treatment groups following primary or revision single-level TLIF.

Data are summarized in Table 48.

Urinary retention

Summary:

On-label use: The strength of evidence for urinary retention is *insufficient* with respect to on-label use of rhBMP-2 compared with autologous bone graft. There was only one very small pilot study (n=11) reporting on this outcome.

Off-label use: There were no off-label comparative studies reporting on this outcome.

All data are summarized in Table 49.

Episodes of radiculitis

Summary:

On-label

Only one nonrandomized comparative study was found, the FDA SSED for InFUSE, which reported similar risks of radiculitis following ALIF with rhBMP-2 compared with ICBG (23% vs. 22%, respectively). The strength of evidence is *insufficient* regarding these estimates.

Off-label

Two comparative studies evaluating rhBMP-2 in the lumbar spine, one RCT and one cohort study, reported similar low risks of radiculitis in the BMP compared with the control groups (0-2%). For rhBMP-7 use in the lumbar spine, only one RCT was found which reported a lower risk in the BMP compared with the control group (6% vs. 13%). The strength of evidence is *low* regarding these estimates.

All data are summarized in Table 50.

Dural injury or CSF leak

Summary:

On-label

Only one nonrandomized comparative study was found, the FDA SSED for InFUSE, which reported similar risks of dural injury or durotomy following ALIF with rhBMP-2 compared with ICBG (0% vs. 0.7%, respectively). The strength of evidence is *insufficient* regarding these estimates.

Off-label

Evidence from three RCTs and seven cohort studies evaluating the use of rhBMP-2 or rhBMP-7 in the lumbar or cervical spine shows similar risks of dural injury or durotomy in the BMP groups compared with the control groups; risks ranges from 2.4%–11% irrespective of treatment group. The strength of the evidence is *high* regarding these between-group comparisons.

All data are summarized in Table 51.

Neurological (unspecified or other) adverse events

Summary:

On-label

Only one nonrandomized comparative study was found, the FDA SSED for InFUSE, which reported similar risks of neurological adverse events following ALIF with rhBMP-2 compared with ICBG (12.5% vs. 15.1%, respectively). The strength of evidence is *insufficient* regarding these estimates.

Off-label
Evidence from four RCTs and four cohort studies (to include the FDA SSPB for OP-1) evaluating the use of rhBMP-2 or rhBMP-7 in the lumbar spine shows similar risks of neurological adverse events in the BMP groups compared with the control groups; risks ranged from 4.0%–26.0% irrespective of treatment group. The strength of the evidence is *high* regarding these between-group comparisons.

All data are summarized in Table 52.

Antibody responses to BMP

Summary:

On-label
One RCT was found which reported similar low risks of elevated anti-BMP antibodies following the use of rhBMP-2 in the lumbar spine compared with controls (0.7% vs. 0.8%, respectively). The strength of the evidence is *low* regarding these estimates.

Off-label
Four RCTs reported similar low risks of elevated anti-BMP antibodies following the use of rhBMP-2 in the lumbar (3 RCTs) and cervical spine (1 RCT) compared with controls (0–0.7%).
One RCT which evaluated rhBMP-7 use in the lumbar spine reported a higher proportion of patients with elevated anti-BMP antibodies in the BMP group compared with the control group (93.7% vs. 21%). The strength of evidence is *high* regarding the results for rhBMP-2 and *low* for rhBMP-7.

None of the studies reported any negative consequences to elevated or positive antibody responses.

All data are summarized in Tables 53.

Antibody responses to collagen

Summary:

On-label
One RCT was found which reported similar low risks of anti-bovine collagen or elevated anti-human collagen antibodies following the use of rhBMP-2 in the lumbar spine compared with controls (0.7% vs. 0.8%, respectively). The strength of the evidence is *low* regarding these estimates.

Off-label

Pooled estimates from two RCTs showed similar risks of anti-bovine collagen or elevated anti-human collagen antibodies following the use of rhBMP-2 in the lumbar spine compared with controls (9% vs. 11%, respectively). The strength of the evidence is *low* regarding these estimates.

None of the studies reported any negative consequences to elevated or positive antibody responses.

Data are summarized in Table 54.

Cancer

Summary:

On-label

Only one RCT was found which reported no difference in the risk of cancer at 24 months following ALIF with rhBMP-2 (0.7%) compared with ICBG (0.7%). The strength of evidence is *low* regarding these estimates.

Off-label

Three RCTs and one cohort studies were identified which generally reported higher cancer risks at 1, 2, 4 and 5 years following the use of rhBMP-2 or rhBMP-7 in the lumbar spine. One RCT of a higher dose (40 mg) of rhBMP-2 reported higher cancer risks following PLF with rhBMP-2 compared with controls at 24 months (3.8% vs. 0.9%) and at 60 months (6.3% vs. 2.2%). Similarly, higher incidences of cancer were reported following PLF with rhBMP-7 compared with controls in two RCTs, one with 12 months (5.6% vs. 0%) and one with 48 months (12.5% vs. 8.3%) follow-up. One cohort study, a retrospective chart review, reported higher risks of cancer following various surgical approaches with rhBMP-2 (16.7%) compared with control (7.6%). The strength of evidence is *moderate* regarding these between-group comparisons.

Data are summarized in Table 56.

Cardio/vascular

Summary:

On-label

One RCT and one nonrandomized comparative study, the FDA SEED for InFUSE, reported similar risks of cardio/vascular events following rhBMP-2 use in the lumbar spine compared with controls (4.2%–10.1 vs. 2.2%–12.2%, respectively). The strength of evidence is *low* regarding these between-group comparisons.

Off-label

In general, results from four RCTs and three cohort studies (to include the FDA SSPB for OP-1) show similar risks of cardio/vascular events following rhBMP-2 or rhBMP-

7 use in the lumbar spine compared with controls (3.9%–22.2% vs. 2%–24.1%, respectively). The strength of evidence is *high* regarding these between-group comparisons.

All data are summarized in Table 57.

Deep vein thrombosis

Summary:

On-label

There was no difference in the incidence of DVT in patients treated with rhBMP-2 compared with control as reported by one RCT: 0% versus 1.5%, respectively. The strength of evidence is *low* regarding these estimates.

Off-label

According to three comparative studies, the risks of DVT in patients treated with rhBMP-2 in the lumbar (1 RCT, 1 cohort) and cervical spine (1 cohort) were similar compared with controls: 0%–9% versus 1.9%–12%, respectively. The strength of evidence is *low* regarding these between-group comparisons.

All data are summarized in Table 58.

Death

Summary:

On-label

No difference in the incidence of death between patients treated with rhBMP-2 in the lumbar spine compared with control was reported by one RCT at 24 months (0% vs. 0.7%, respectively). The strength of evidence is *low* regarding these estimates.

Off-label

In the lumbar spine, similar risks of death following the use of rhBMP-2 or rhBMP-7 (1.6%–5.3%) compared with controls (1.7%–6.0%) were reported by four RCTs and two cohort studies at 24 and 36 months, respectively. Following cervical fusion, one retrospective cohort study reported a statistically higher risk of death up to 90 days post-operative in patients treated with (4.2%; 11/260) versus without (1.7%; 9/515) BMP (type unspecified); $P = .047$. The causes of death were not reported, and the significance of this result should be interpreted with some caution as no demographic or surgical details were provided and there is thus an absence of controlling for possible confounding between treatment groups. The strength of evidence is *high* regarding these estimates in the lumbar spine and *insufficient* in the cervical spine.

All data are summarized in Table 59.

Secondary surgical procedures

Revision: *surgery that modified or adjusted the original implant*

Summary:

On-label

One nonrandomized comparative study (an integrated analysis) reported no difference in the risks of revision, defined as surgery that modified or adjusted the original implant, between patients treated with rhBMP-2 in the lumbar spine compared with controls (0.4% vs. 2.0%, respectively). The strength of evidence is *insufficient* regarding these estimates.

Off-label

In general, risks of revision following lumbar spinal fusion were similar between rhBMP and control groups as reported by seven RCTs (five rhBMP-2 and two rhBMP-7) over a range of 17 to 48 months follow-up: 6.0% versus 6.2%, respectively (pooled results). Overall risks were slightly higher with rhBMP-7 use (9.5% vs. 11%) compared with rhBMP2 use (3.8% vs. 4.8%). Results from three cohorts, two evaluating rhBMP-2 in the lumbar spine and one in the cervical spine, indicate lower risks of revision in the BMP groups compared with the control groups: 3% versus 10% (lumbar) and 0% versus 4% (cervical), respectively. The strength of evidence is *high* regarding these between-group comparisons.

All data are summarized in Table 60.

Hardware removal: *removal of one or more components of the original implant (including replacement with a different implant)*

Summary:

On-label

The incidence of hardware removal, defined as removal of one or more components of the original implant (including replacement with a different implant), was similar between rhBMP-2 and control groups as reported by one RCT and one cohort study (integrated analysis with partial overlap of data with the RCT) at 24 months: 1.4% versus 0% and 1.4% vs. 1.7%, respectively. The strength of evidence is *low* regarding these estimates.

Off-label

Risks of hardware removal were slightly less in patients receiving rhBMP-2 in the lumbar spine compared with controls across four RCTs with 24 months follow-up (2.8% vs. 7.2%) and similar as reported by two cohort studies with 3 to 28 months follow-up (8.0% for both groups). The strength of evidence is *moderate* regarding these between-group comparisons.

All data are summarized in Table 61.

Supplemental fixation: *surgery to provide additional stabilization to the index site*

Summary:
On-label
Patients treated with rhBMP-2 in the lumbar spine had lower risks of supplemental fixation, defined as surgery to provide additional stabilization to the index site, compared with patients who received ICBG according to data from one small pilot RCT and one pivotal RCT (5.2% vs. 10.8%, respectively). The strength of evidence is *low* regarding these between-group comparisons.

Off-label
The incidence of supplemental fixation following rhBMP-2 or rhBMP-7 in the lumbar spine compared with controls varied across a total of eight studies, depending on the protein evaluated. Two RCTs with 24 months follow-up and four cohort studies with follow-up periods ranging from 3 to 36 months reported lower mean risks of supplemental fixation in the rhBMP-2 groups compared with the controls: 2.5% versus 6.2% and 6.7% versus 9.5%, respectively. Conversely, two small RCTs found a higher mean risk among those treated with rhBMP-7 (10%) than with control (0%) at 24 months. Only one nonrandomized comparative study was found that reported incidence of supplemental fixation following rhBMP-2 use in the cervical spine and showed a lower risk compared with controls at 30 months (0% versus 3.0%). The strength of evidence is *moderate* regarding the risk of supplemental fixation with off-label use of rhBMP-2 in the lumbar spine, *low* for the estimates of the risk for rhBMP-7 use in the lumbar spine and *insufficient* for the estimates of the risk for rhBMP-2 use in the cervical spine.

All data are summarized in Table 62.

Reoperation: *additional procedure at the index level besides a revision, hardware removal, or supplemental fixation*

Summary:
On-label
One nonrandomized comparative study, an integrated analysis, reported a lower risk of reoperation, defined as an additional procedure performed at the index level besides a revision, hardware removal, or supplemental fixation, following rhBMP-2 use in the lumbar spine compared with control (2.9% vs. 8.0%, respectively). The strength of evidence is *insufficient* regarding these estimates.

Off-label

Three RCTs, two investigating rhBMP-2 and one rhBMP-7, and two cohort studies of rhBMP-2 all reported similar risks of reoperation following lumbar spinal fusion with BMP compared with control over 3 to 27 months of follow-up (1.0%–10% vs. 2.0%–10%, respectively). In contrast, one case-control database study of nearly 5000 patients found lower rates of repeat fusion after BMP use compared with those who did not receive BMP. One nonrandomized comparative study also reported a similar risk of reoperation following rhBMP-2 use in the cervical spine compared with controls at 30 months (0% versus 3.0%). The strength of evidence is *high* regarding these between-group comparisons for the off-label use of rhBMP.

All data are summarized in Table 63.

Fusion at a different spinal level

Summary:

On-label

There were no on-label studies reporting on this outcome.

Off-label

Risks of fusion at a different spinal level were similar between patients treated with rhBMP-2 and control in the lumbar spine as reported by two RCTs at 24 months follow-up (5.0% vs. 4.0%, respectively). In the cervical spine, risks were similar following rhBMP-2 (3.8%–5.6%) compared with controls (0%) at 24 to 30 months as reported by one small RCT and two small cohort studies. The strength of evidence is *moderate* for these estimates.

All data are summarized in Table 64.

Graft site morbidity

Summary: Following ICBG harvesting, hip pain VAS (0-10) ranged from 5.7-8.0 in the perioperative period as reported by four studies[6-9] and from 0.2-2.8 at 12-24 months (last follow-up) as reported by six studies[6-11]. The percentage of patients experiencing pain (definitions varied by study, see table for details) at last follow-up (6-36 months) ranged from 10-66% as reported by nine studies[2, 9, 10, 12-20]. Additional complications included injury to lateral femoral cutaneous nerve, ASIS fractures, superficial infection, deep infection requiring surgery, and hematoma.

Key Question 4: Differential efficacy or safety in subpopulations

What is the evidence that on- or off-label use of rhBMP-2 or rhBMP-7 for spinal fusion compared with spinal fusion using ICBG or alternative bone graft substitutes has differential efficacy or safety issues in sub-populations? Including consideration of:

- Gender
- Age
- Baseline functional or pain status
- Comorbidities (including but not limited to tobacco use, alcohol use, psychological or psychological)
- Other patient characteristics or evidence-based patient selection criteria
- Provider type, setting or other provider characteristics
- Payor/ beneficiary type: including worker's compensation, Medicaid, state employees

Summary: We found *insufficient* evidence of the differential effectiveness of spinal fusion using rhBMP-2 or rhBMP-7 versus spinal fusion using ICBG or alternative bone graft substitutes in any subpopulation. Although eight studies examined outcomes in various subpopulations, none of these studies pre-specified the subgroup analyses, none of the studies performed a test of interaction as the method of subgroup analysis, and some of the studies were inadequately powered to detect differences in treatment effect. In general, fusion without rhBMP tended to have lower risks of complications, while fusion with rhBMP tended to have better radiographic outcomes across most subpopulations examined, although in many cases the differences were small.

Key Question 5: Cost-effectiveness

What evidence of cost implications and cost-effectiveness of on- or off-label use of use of rhBMP-2 or rhBMP-7 exists? Including consideration of:

- Costs (direct and indirect) and cost effectiveness
- Short term and long term

Summary:

There is conflicting evidence about whether the use of rhBMP-2 for on-label lumbar spinal fusion results in better outcomes and/or lower costs than control or standard treatment. The strength of the evidence is *low.*

- One well conducted cost effectiveness analysis performed by the AHRQ suggested that when analyzed as part of the treatment cost, on-label use of rhBMP-2 for lumbar spinal fusion results in lower costs per QALY only when it is assumed to be a part of the Medicare reimbursement and no cost differential is calculated. However the more common payer strategy assumes the cost of rhBMP-2 is added to treatment, in which case the group treated

with rhBMP-2 had higher QALYs and higher cost, a common outcome for new technologies.
- One CUA concluded on-label use of rhBMP-2 for lumbar spinal fusion unlikely to be cost-effective due to higher costs and similar outcomes compared with the control group.

One cost-utility study showed that off-label use of rhBMP-2 was more cost-effective than ICBG for posterolateral spinal fusion in patients at least 60 years of age.
- One moderately well conducted cost utility analysis determined that off-label use of rhBMP-2 in posterolateral spinal fusion was associated with similar efficacy and somewhat lower risks of complications compared with ICBG, resulting in a decreased overall cost of $2319.

No studies were identified that evaluated the cost-effectiveness of rhBMP-2 for use in the cervical spine.

No studies were identified that evaluated the cost-effectiveness of rhBMP-7 use in the spine.

Summary Strength of Evidence

Key Question 1: Validated instruments for measuring treatment outcomes		
	SoE	**Conclusions/Comments**
Measures		The most commonly used outcome measures in comparative studies evaluating BMP use in lumbar or cervical spinal fusion were identified. The following outcome measures have undergone psychometric analysis in spine patients: Measures: • ODI (18 studies) • SF-36 (17 studies) • Pain assessed by VAS (14 studies) • NDI (1 cervical study)
Validity, reliability, and responsiveness		One outcome measure (SF-36) has been shown to have criterion validity and reliability in patients undergoing spinal fusion by one study. All four outcome measures have been shown to have a degree of validity, reliability, and responsiveness in various spine populations, some of which might be eligible for fusion.
MCID		For the ODI (scale 0-100), the MCID has been variously defined in fusion patients as 10-22.9 depending on the study population and calculation method. However, there is some cause for concern regarding the definition. For VAS pain (scale 0-10), the MCID has been defined by one study as 1.8-1.9. However, there is some cause for concern regarding the definition. No studies were found that examined the MCID of the SF-36 or NDI in any spine population.

MCID: minimal clinically important difference; NDI: Neck Disability Index; ODI: Oswestry Disability Index; SF-36: Short-Form 36; SoE: Strength of Evidence; VAS: Visual Analogue Scale

Summary of evidence for Key Question 2

Key Question 2: Efficacy and effectiveness		
	SoE	**Conclusions/Comments**
On-label use of rhBMP-2 in the lumbar spine		
Efficacy		***Study characteristics*** • Evidence base: 2 RCTs[6, 21] (LoE IIb). Study size ranged from 14 to 279 patients. • Interventions: Primary single-level open anterior lumbar fusion with either rhBMP-2/ACS (InFUSE) (n = 154) or iliac crest bone autograft (ICBG) (n = 139). RhBMP-2 was used at a dose ranging from 4.2-8.4 mg/patient. • Population: Patients with DDD, radiculitis, and/or up to 25% spondylolisthesis who were refractory to conservative care. • Length of follow-up: 24 months. • Sponsorship: Both studies were sponsored by Medtronic. • These studies served as the pilot and pivotal trials in the 2002 FDA Summary of Strength and Effectiveness Data (SSED) for InFUSE (P000058) [22]. • The studies were similar in design, thus we were able to pool outcomes data. • Additional details: Table 8 and surrounding text.
	Low	***Conclusions*** The following outcomes were similar in both treatment groups: mean operative time (2 RCTs), length of hospital stay (2 RCTs), fusion (2 RCTs), ODI outcomes (2 RCTs), back and leg pain outcomes (1 RCT), SF-36 physical function scores (1 RCT), patient satisfaction (2 RCTs), return to work (2 RCTs), and neurological success (1 RCT). The following outcomes were improved in patients treated with rhBMP-2 compared with ICBG: perioperative blood loss (2 RCTs).
Effectiveness		***Study characteristics*** • Evidence base: 1 integrated analysis[23] (LoE II) based on the following studies: one RCT[6] (reported in efficacy) (n = 279), one case series[24] (n = 22), and one unpublished study (n = 378). • Interventions: Primary single-level open (41%) or laproscopic (59%) anterior lumbar fusion with either rhBMP-2/ACS (InFUSE) (n = 277) or iliac crest bone autograft (ICBG) (n = 402). • Population: Patients with DDD and radiculitis who were refractory to conservative care. • Length of follow-up: 24 months. • Sponsorship: Both studies have been reported to be sponsored by Medtronic. • Additional details: Table 13 and surrounding text.
	Low	***Conclusions*** The following outcomes were improved in patients treated with rhBMP-2 compared with ICBG: perioperative outcomes (operating time, blood loss, and length of hospital stay). The following outcomes were similar in both treatment groups: fusion, ODI outcomes, SF-36 pain index and physical component subscale scores, and return to work.

Key Question 2: Efficacy and effectiveness		
	SoE	**Conclusions/Comments**
On-label use of rhBMP-7 in the lumbar spine		
Efficacy		No studies were identified that evaluated the efficacy of on-label use of rhBMP-7 in the lumbar spine.
Effectiveness		No studies were identified that evaluated the effectiveness of on-label use of rhBMP-7 in the lumbar spine.
Off-label use of rhBMP-2 in the lumbar spine		
Efficacy		***Study characteristics*** • Evidence base: 6 RCTs[1, 7-9, 12, 25, 26]: LoE IIa (1 study)[1, 12], LoE IIb (5 studies)[7-9, 25, 26]. Study size ranged from 27 to 463 patients. • Interventions: Various. Patients underwent primary single- (or in one study, multi-) level posterior (four studies), anterior (one study), or posterolateral (one study) lumbar fusion with either rhBMP-2/ACS (InFUSE) or iliac crest bone autograft (ICBG). Due to heterogeneity in surgical procedures (i.e., approach, use of ceramic granules, use of cage versus allograft dowel versus no device, single- versus multilevel fusion), we did not pool outcomes data from the six studies. Patients received BMP in a variety of forms: rhBMP-2/CRM; InFUSE; and AMPLIFY. Doses of rhBMP-2 varied and ranged from 4.2-40 mg per patient (when reported). • Population: Patients with DDD, radiculitis, and/or up to 25% spondylolisthesis who were refractory to conservative care. • Length of follow-up: 17 (mean) - 24 months. • Sponsorship: Medtronic (5 RCTs)[1, 7-9, 12, 25]; Norton Healthcare grant (1 RCT)[26]. • Additional details: Table 15 and surrounding text. ***Conclusions***
	High	The following outcomes were similar in both treatment groups: length of hospital stay (5 RCTs), ODI outcomes (3 – 6 RCTs), leg pain (6 RCTs), SF-36 scores (6 RCTs), and work status (4 RCTs).
	Low	The following outcomes were similar in both treatment groups: patient satisfaction (2 RCTs), neurological status (1 RCT), and overall success (1 RCT).
	Moderate	The following outcomes were reported as either similar or improved in patients treated with rhBMP-2 compared with ICBG: • mean operative time (similar in 3 RCTs; statistically improved in 3 RCTs), • perioperative blood loss (similar in 4 RCTs; statistically improved in 2 RCTs), • fusion (similar in 3 RCTs; statistically improved in 3 RCTs), • back pain (similar in 5 RCTs; clinically improved in 1 RCT)
Effectiveness		***Study characteristics*** • Evidence base: 8 cohort studies[13, 15, 27-32], (including 2 prospective cohort studies[30, 32], 1 prospective case control study[31], 3 retrospective cohort studies[13, 15, 29], and 2 retrospective cohort studies with historical controls[27, 28]): LoE II (1 study)[32]; LoE III (7 studies)[13, 15, 27-31]. Study size ranged from

Key Question 2: Efficacy and effectiveness		
	SoE	**Conclusions/Comments**
		36-126 patients. • Interventions: Primary or revision single- or multi-level anterior (two studies), posterior (two studies), transforaminal (one study), or posterolateral (three studies) lumbar fusion with rhBMP-2 or iliac crest bone autograft (ICBG), allograft chips, or local or rib autograft. Due to heterogeneity in control treatment, patient diagnosis, and surgical procedures (i.e., approach, use of local autograft or ICBG or bone graft extenders, use of cage versus allograft dowel versus no device, single- versus multilevel design, primary versus revision surgery), we did not pool outcomes data. Doses of rhBMP-2 varied and ranged from 3-36 mg per patient (when reported). • Population: Six of the cohort studies included patients with DDD, radiculitis; some of these studies also include those with up to grade 1 or 2 spondylolisthesis, scoliosis, instability, nonunion, or adjacent segment degeneration. One study treated patients with symptomatic pseudarthrosis following previous PLIF for DDD; one study evaluated patients with scoliosis with degeneration distal to a prior long idiopathic scoliosis fusion site. • Length of follow-up: Mean of 9-39 months. • Sponsorship: Medtronic (1 study)[32], Medtronic and Norton Healthcare grants (1 study)[28], no funding (2 studies)[15, 30], no direct funding but benefits may have been received (1 study)[27], or funding not reported (3 studies)[13, 29, 31]. • Additional details: Table 19 and surrounding text. ***Conclusions***
	Low	The following outcomes were similar in both treatment groups: fusion (similar in 7 studies, improved in 1 study) and pain (5 studies)
	Insufficient	The following outcomes were similar in both treatment groups: operative time (1 study), ODI scores (2 studies), function (2 studies), patient satisfaction (2 studies), overall patient-reported clinical outcome (1 study), medication use (1 study), and mental health/self image (1 study).
	Insufficient	The following outcomes were statistically improved in patients treated with rhBMP-2 compared with control: perioperative blood loss (1 study)
Off-label use of rhBMP-7 in the lumbar spine		
Efficacy		***Study characteristics*** • Evidence base: 5 RCTs{Delawi, 2010 #121;Johnsson, 2002 #119;Kanayama, 2006 #82;Vaccaro, 2005 #118;Vaccaro, 2008 #94;Vaccaro, 2004 #117;Vaccaro, 2008 #96; all LoE IIb. Study size ranged from 20-293 patients. • Interventions: Primary single- level posterior (four studies) or posterolateral (one study) lumbar fusion with either OP-1 (rhBMP-7) or iliac crest bone autograft (ICBG) (four studies) or autograft (1 study). RhBMP-7 was used at a dose of 7 mg per patient. • Population: Patients with degenerative (or in one study, isthmic) spondylolisthesis up to grade 1 (or 2) who had not responded to six months of nonsurgical treatment.

Key Question 2: Efficacy and effectiveness		
	SoE	**Conclusions/Comments**
		• Length of follow-up: Mean of 12-54 months. • Sponsorship: Stryker Biotech (1 RCT); funding received but source not stated (2 RCTs); no direct funding but benefits may have been received (2 RCTs). • Additional details: Table 23 and surrounding text.
	High	The following outcomes were similar in both treatment groups: length of hospital stay (3 RCTs), fusion (5 RCTs), and ODI outcomes (4 RCTs)
	Low	The following outcomes were similar in both treatment groups: back pain (1 RCT), leg pain (1 RCT), SF-36 physical component subscale scores (1 RCT), neurologic success (1 RCT), or overall success (1 RCT).
	Low	The following outcome was reported as either similar or improved in patients treated with rhBMP-7 compared with ICBG or local autograft: • operative time (similar in 2 RCTs, statistically improved in 1 RCT) • perioperative blood loss (similar in 1 RCT, statistically improved in 1 RCT)
Effectiveness		No studies were identified that evaluated the effectiveness of off-label use of rhBMP-7 in the lumbar spine.
Off-label use of rhBMP-2 in the cervical spine		
Efficacy		***Study characteristics*** • Evidence base: 1 RCT[5]: LoE IIb. There were 33 patients enrolled in the study. • Interventions: Primary one- or two- level ACDF with InFUSE (n = 18) or ICBG (n = 15). RhBMP-2 was used at a dose of 0.6-1.2 mg per patient. • Population: Patients with degenerative cervical disease with radiculopathy and/or myelopathy. • Length of follow-up: 24 months. • Sponsorship: No direct funding but benefits may have been received. • Additional details: Table 30 and surrounding text.
	Low	***Conclusions*** The following outcomes were similar in both treatment groups: operative time, perioperative blood loss, length of hospital stay, fusion, neck pain, SF-36 scores, patient satisfaction, and neurological success. The following outcomes were statistically improved in patients treated with rhBMP-2 compared with ICBG: NDI and arm pain scores.
Effectiveness		***Effectiveness*** • Evidence base: 5 cohort studies[11, 33-36] (including 1 prospective cohort study[11], 3 retrospective cohort studies[33-35], and 1 retrospective case-control database study[36]): all LoE III. Study size ranged from 58-775 patients. • Interventions: Primary or revision single- or multi-level anterior (two studies), posterior (two studies) cervical fusion with rhBMP-2 or iliac crest bone autograft (ICBG) (two studies), allograft and demineralized bone matrix (one study), a combination of autograft and/or allograft materials (one study). One study did not report surgical approach or the details of the control treatment (referred to as "non-BMP"). BMP was used a dose that

Key Question 2: Efficacy and effectiveness		
	SoE	**Conclusions/Comments**
		ranged from 0.9 to 12 mg per patient (when reported). Due to heterogeneity in control treatments and surgical procedures (i.e., approach, use of local autograft or ICBG or allograft, single- versus multilevel design, primary versus revision surgery), we were not able to pool outcomes data. • Population: Two of the cohort studies included patients with DDD; another included patients with DDD, herniated nucleus pulposus, or stenosis. A fourth study treated patients for stenosis, spondylosis, or nonunion from a previous fusion. The fifth study did not report patient diagnoses. • Length of follow-up: 1 – 36 months. • Sponsorship: : No funding received (1 study)[11]; funding received but source not stated (1 study)[36]; no direct funding but benefits may have been received (2 studies)[33, 35], and funding not reported (1 study)[34]. • Additional details: Table 35 and surrounding text. ***Conclusions***
	Low	The following outcome was similar in both treatment groups: perioperative blood loss (3 studies).
	Insufficient	The following outcomes were similar in both treatment groups: operative time (2 studies), ODI outcomes (2 studies), and arm pain (2 studies).
	Insufficient	The following outcome was reported as either similar or improved in patients treated with rhBMP-2 compared with control: • fusion (similar in 1 cohort study, statistically improved in 1 cohort study)
	Insufficient	The following outcomes were reported as either similar or worse in patients treated with rhBMP-2 compared with control: • length of hospital stay (4 studies reported similar outcomes while 1 large study reported longer hospital stays in the rhBMP-2 group compared with the control group) • neck pain (2 studies reported similar outcomes while 1 study reported more rhBMP-2 patients with persistent neck pain at final follow-up). One study was funded but the source was not stated; authors from two studies may have received financial or other benefits related to the study; one study was not funded; funding for the remaining study was not reported.
Off-label use of rhBMP-7 in the cervical spine		
Efficacy		No studies were identified that evaluated the efficacy of off-label use of rhBMP-7 in the cervical spine.
Effectiveness		No studies were identified that evaluated the efficacy of off-label use of rhBMP-7 in the cervical spine.

DBM: demineralized bone matrix; SoE: Strength of Evidence

Summary of evidence for Key Question 3

Key Question 3: Safety		
	SoE	**Conclusions/Comments**
Bone overgrowth		**On-label:** no on-label comparative studies reporting on this outcome.
	Low	**Off-label**: inconsistent results reported • no cases of bone overgrowth in either group (2 RCTS and 2 cohort studies) • incidence of 75% in the rhBMP group vs. 13% in the control (1 RCT), and 21% vs. 8% of spinal levels (1 cohort study)
Osteoclast activity (resorption, osteolysis, graft migration/ loosening/ subsidence)	**Low**	**On-label:** similar risks in both groups, 1.3% vs. 0.0% (FDA pilot and pivotal RCTs for InFUSE)
	Moderate	**Off-label:** similar or possible higher risk for rhBMP • similar with risks ≤ 6% in each group (3 RCTs) • higher risk in the rhBMP-2 group, 62% of spinal levels vs. 10% in the controls (1 cohort study)
Local wound complications, superficial	**Insufficient**	**On-label:** insufficient risk estimates (1 very small pilot study, N = 14)
	Moderate	**Off-label:** similar risks (<10%) in both groups (2 RCTs, 5 cohorts)
Local wound complications, superficial or deep (unspecified)	**Low**	**On-label:** similar risks in both groups, 12.2% vs. 11.5% (FDA pilot and pivotal RCTs for InFUSE)
	Moderate	**Off-label**: similar but variable risks in both groups, 0–20% (4 RCTs, 2 cohort studies)
Local wound complications, deep; surgery for deep wound complications		**On-label:** no on-label comparative studies reporting on this outcome.
	Low	**Off-label:** similar risks in both groups, ≤10% in each group (1 RCT, 4 cohort studies)
Dysphagia/ neck swelling	**Insufficient**	**On-label:** higher risk of "respiratory" complications in rhBMP-2 patients, 8.6% vs. 1.7% (FDA summary on InFUSE)
	Moderate	**Off-label (lumbar):** similar risks of "respiratory" complications in both groups, <7% (3 RCTs, 2 cohorts)
Dysphagia/ neck swelling (cont.)	**Moderate**	**Off-label (cervical)**: higher risks in the rhBMP groups • 35% vs. 9% (pooled) (4 cohort studies) • ~2 fold increase (2 large database studies)
Retrograde ejaculation	**Low**	**On-label:** higher risk in rhBMP-2 groups, 7.9% vs. 1.4% (FDA summary on InFUSE)
	Low	**Off-label**: higher risk in rhBMP-2 groups, 7.2% vs. 0.6% (1 cohort study)
Ileus/bowel obstruction	**Insufficient**	**On-label:** insufficient risk estimates (1 very small pilot study, N = 14)

Key Question 3: Safety		
	SoE	**Conclusions/Comments**
	Insufficient	**On-label:** insufficient risk estimates (1 retrospective cohort study)
Urinary retention	**Insufficient**	**On-label:** insufficient risk estimates (1 very small pilot study, N = 14) **Off-label:** No comparative studies reported on this outcome.
Radiculitis (adverse event)	**Insufficient** **Low**	**On-label**: similar risks in both groups, 23% vs. 22% (FDA SSED for InFUSE) **Off-label**: similar or lower risks in the rhBMP groups • Risks similar for rhBMP-2 compared with controls, 0-2% (1 RCT, 1 cohort study) • Risk lower for rhBMP-7 compared with controls, 6% vs. 13% (1 RCT)
Dural injury or CSF leak	**Insufficient** **High**	**On-label**: similar low risks in both groups, 0% vs. 0.7% (FDA SSED for InFUSE) **Off-label**: similar but variable risks in both groups, 2.4–11% (3 RCTs, 7 cohort studies)
Neurological, unspecified/other (adverse event)	**Insufficient** **High**	**On-label:** similar risks in both groups, 12.5% vs. 15.1% (FDA SSED for InFUSE) **Off-label:** similar but variable risks in both groups, 4.0–26.0% (4 RCTs, 3 cohort studies, FDA SSPB*)
Antibody responses to BMP	**Low** **High** **Low**	**On-label:** similar low risks in both groups, 0.7% vs. 0.8% (1 RCT) **Off-label (rhBMP-2):** similar low risks in both groups, 0–0.7% (4 RCTs) **Off-label (rhBMP-7):** higher risk in rhBMP group, 93.7% vs. 21% (no clinical sequelae) (1 RCT)
Antibody responses to collagen	**Low** **Low**	**On-label:** similar low risks in both groups, 0.7% vs. 0.8% (1 RCT) **Off-label:** similar risks in both groups, 9% vs. 11% (2 RCTs)
Cancer	**Low** **Moderate**	**On-label:** similar low risks in both groups, 0.7% vs. 0.7% (1 pivotal RCT of the FDA SSED for InFUSE) **Off-label:** higher cancer risks in the rhBMP-2 and rhBMP-7 groups at 1, 2, 4, and 5 years; 3.8–16.7% vs. 0.9–7.6% (3 RCTs including the pivotal RCT of the FDA SSED for InFUSE, 1 cohort study)
Cardio/vascular	**Low** **High**	**On-label:** similar but variable risks in both groups, 4.2–10.1% vs. 2.2–12.2%, (1 RCT + largely overlapping FDA SSED) **Off-label:** similar but variable risks in both groups, 3.9–18.3% vs. 2.0–22.1%, (4 RCTs, 3 cohort studies to include the FDA SSPB* which may partially overlap)

Key Question 3: Safety		
	SoE	**Conclusions/Comments**
Deep vein thrombosis	**Low**	**On-label:** similar low risks in both groups, 0% vs. 1.5% (1 RCT)
	Low	**Off-label:** similar but variable risks in both groups, 0–9% versus 1.9–12% (1 RCT, 2 cohort studies)
Death	**Low**	**On-label:** similar low risks in both groups at 24 months, 0% vs. 0.7% (1 RCT + largely overlapping FDA SSED)
	High	**Off-label, lumbar:** similar but variable risks in both groups at 24 to 36 months, 1.6–5.3% vs. 1.7–6.0% (4 RCTs, 2 cohort studies)
	Insufficient	**Off-label, cervical:** higher risk in rhBMP group up to 90 days post-operative, 4.2% vs. 1.7%, *P* = .047 (1 RCT); • causes of death were not reported, no demographic or surgical details provided – thus, significance of this result should be interpreted with caution given an absence of controlling for possible confounding between treatment groups
Revision	**Insufficient**	**On-label:** similar low risks in both groups at 24 months, 0.4% vs. 2.0% (1 integrated analysis)
	High	**Off-label:** similar or lower risks in the rhBMP groups over 17 to 48 months follow-up • Pooled risks similar in both groups, 6.0% vs. 6.2% (7 RCTs) • Lower risks in the rhBMP groups (0–3%) vs. controls (4–10%) (3 cohort studies) • Overall risks were slightly higher with rhBMP-7 use (9.5% vs. 11%) compared with rhBMP2 use (3.8% vs. 4.8%)
Hardware removal	**Low**	**On-label:** similar low risks in both groups at 24 months, 0–1.7% (1 RCT + 1 partially overlapping integrated analysis)
	Moderate	**Off-label:** lower or similar risk in rhBMP groups over 3 to 28 months follow-up • Pooled risks lower in rhBMP-2 group (2.8%) vs. controls (7.2%) at 24 months (4 RCTs) • Risks identical between groups (8.0%) (2 cohort studies)
Supplemental fixation	**Low**	**On-label:** lower risks in rhBMP groups at 24 months, 5.2% vs. 10.8% (2 RCTs: 1 small pilot RCT and one pivotal RCT)
	Moderate	**Off-label (rhBMP-2, lumbar):** lower risk in rhBMP groups, 2.5–6.7% vs. 6.2–9.5% over 2 to 36 months follow-up (2 RCTs, 4 cohorts)
	Low	**Off-label (rhBMP-7, lumbar):** higher risk in rhBMP groups at 24 months, 10% vs. 0% (2 small RCTs)
	Insufficient	**Off-label (rhBMP-2, cervical):** lower risks in rhBMP groups at 30 months, 0% vs. 3.0% (1 large database study)
Reoperation	**Insufficient**	**On-label:** lower risks in the rhBMP-2 group at 24 months, 2.9% vs. 8.0% (1 integrated analysis)
	High	**Off-label:** similar but variable risks in both groups, 0–10% vs. 2.0–10% over 3 to 30 months follow-up (3 RCTs, 3 cohort studies)

Key Question 3: Safety		
	SoE	**Conclusions/Comments**
Fusion at a different spinal level	**Moderate**	**On-label:** no on-label studies reporting on this outcome. **Off-label:** similar but variable risks in both groups, 3.8–5.6% vs. 0–4.0% at 24 to 30 months follow-up (3 RCTs, 2 cohort studies)
Second surgeries (details not reported)	**Insufficient**	**On-label:** no on-label studies reporting on this outcome. **Off-label:** similar but variable risks in both groups, 10.8–15.0% vs. 10.5–21.0% at 24 to 48 months follow-up (2 cohort studies)

*The FDA SSPB for OP-1 was to be considered of low quality because no information was presented for 278/326 patients included in the safety data tables (i.e., study data only presented for the pilot trial, which included 48 patients and presumably is part of the 326 patients evaluated for safety).

Summary of evidence for Key Question 4

Key Question 4: Differential efficacy, effectiveness, and safety		
	SoE	**Conclusions/Comments**
• **Age** • **Sex** • **Smoking status** • **Number of levels treated** • **Complexity of fusions** • **Surgical approach** • **Previous surgeries**	**Insufficient**	We found no strong evidence of the differential effectiveness of spinal fusion using rhBMP-2 or rhBMP-7 versus spinal fusion using ICBG or alternative bone graft substitutes in any subpopulation. Although these eight studies examined outcomes in various subpopulations, none of these studies pre-specified the subgroup analyses, none of the studies performed a test of interaction as the method of subgroup analysis, and some of the studies were inadequately powered to detect differences in treatment effect. In general, fusion without rhBMP tended to have lower risks of complications, while fusion with rhBMP tended to have better radiographic outcomes across most subpopulations examined, although in many cases the differences were small.
• **Baseline functional or pain status** • **Provider type** • **Payor/ beneficiary type**		No studies were identified that evaluated the differential effectiveness of spinal fusion with rhBMP-2 or rhBMP-7 based on baseline function or pain, provider type, or payor/ beneficiary type.

Summary of evidence for Key Question 5

Key Question 5: Cost effectiveness		
	SoE	**Conclusions/Comments**
Cost-effectiveness	**Low**	<u>RhBMP-2 use in lumbar spine (on-label):</u> Conflicting evidence (2 studies): • One study concluded that when analyzed as part of the treatment cost, on-label use of rhBMP-2 results in lower costs per QALY only when it is assumed to be a part of the Medicare reimbursement and no cost differential is calculated. When the cost of BMP is added to the treatment (more common policy), BMP use associated with higher QALYs and higher cost, a common outcome for new technologies. • Another study found that rhBMP-2 use was unlikely to be cost-effective due to higher costs and similar outcomes compared with the control group.
	Low	<u>RhBMP-2 use in lumbar spine (off-label):</u> Off-label use of rhBMP-2 was more cost-effective than ICBG for posterolateral spine fusion in patients ≥ 60 years of age (1 study). No studies were identified that evaluated the cost-effectiveness of spinal fusion with rhBMP-7 in the lumbar or cervical spine or with rhBMP-2 in the cervical spine.

QALY: quality-adjusted life year

1. Appraisal

1.1. Rationale

Recombinant bone morphogenetic proteins (rhBMPs) are currently being used as an adjunct to spinal fusion procedures. To date, two rhBMPs (rhBMP-2 and rhBMP-7) and associated delivery vehicles have received approval from the FDA. InFUSE Bone Graft/LT-CAGE Lumbar Tapered Fusion Device (Medtronic), which is made of rhBMP2 on an absorbable collagen sponge carrier, has received FDA premarket approval for treatment of single level anterior open or anterior laparoscopic spinal fusion procedures in adults with degenerative disc disease (DDD) in the lumbar spine (between L4 and S1) and who have failed conservative care. OP-1 Putty (Stryker), comprised of rhBMP7 and bovine collagen, has received a humanitarian device exemption from the FDA for use in compromised patients undergoing revision posterolateral lumbar spinal fusion for whom autologous bone graft or bone marrow harvest are contraindicated or not expected to result in fusion. In addition, rhBMP2 and rhBMP7 are being used for other off-label uses in the lumbar and cervical spine.

Significant questions remain about the safety, efficacy and effectiveness and cost effectiveness of the use of rhBMP-2 or rhBMP-7 for spinal fusion procedures. Safety questions in particular have recently arisen regarding whether the potential beneficial outcomes of BMP use outweigh their risks; the risks of the intervention, and how often complications arise.

The primary aim of this assessment is to systematically review, critically appraise and analyze research evidence comparing the efficacy, effectiveness and safety of FDA-approved (on-label) and unapproved (off-label) uses of rhBMP-2 and rhBMP-7 in the spine. Available information on the economic impact of this will also be summarized and critically appraised.

1.2. Key Questions

Specific key questions, as formulated by the HCA/Agency, include the following:

When used in patients undergoing spinal fusion:

- What are the expected treatment outcomes of primary single or multilevel lumbar or cervical spinal fusion for degenerative disc disease (DDD), and of revision posterolateral lumbar spinal fusion in compromised patients (i.e., osteoporosis, smoking, diabetes)? Are there validated instruments related to outcomes in patients undergoing these procedures? Has clinically meaningful improvement in outcomes been defined in these patient populations?
- What is the evidence of efficacy and effectiveness of:

a) rhBMP-2 (InFUSE) for on-label lumbosacral spine fusion in patients with DDD?

b) rhBMP-7 (OP-1) for on-label revision posterolateral lumbar spine fusion in compromised (e.g., osteoporosis, smoking, diabetes) patients?

c) rhBMP-2 (InFUSE) for off-label lumbosacral spine fusion?

d) rhBMP-7 (OP-1) for off-label lumbosacral spine fusion?

e) rhBMP-2 (InFUSE) for off-label cervical spine fusion?

f) rhBMP-7 (OP-1) for off-label cervical spine fusion?

Including consideration of perioperative outcomes (including length of surgery) as well as short term and long term:

- Impact on function, pain, radiographic fusion, patient satisfaction, quality of life, activities of daily living and return to work
- Other reported measures

- What is the evidence of the safety of on- or off-label use of rhBMP-2 or rhBMP-7 for spinal fusion compared with spinal fusion using ICBG or alternative bone graft substitutes? Including consideration of:
  - Short- and long term adverse events and complications type and frequency (pain, donor site morbidity, resorption/osteolysis, heterotopic bone formation, graft subsidence, graft migration, dysphagia or respiratory difficulties, elevated antibody responses to BMPs or collagen, wound complications (infection, hematoma, seroma, or dehiscence), local or systemic toxicity, mispositioned graft, neurological complications, retrograde ejaculation, urogenital complications, allergic reactions, mortality, other major morbidity).
  - Revision/re-operation rates

- What is the evidence that on- or off-label use of rhBMP-2 or rhBMP-7 for spinal fusion compared with spinal fusion using ICBG or alternative bone graft substitutes has differential efficacy or safety issues in sub-populations? Including consideration of:
  - Gender
  - Age
  - Baseline functional or pain status
  - Comorbidities (including but not limited to tobacco use, alcohol use, psychological or psychological)
  - Other patient characteristics or evidence-based patient selection criteria
  - Provider type, setting or other provider characteristics
  - Payor/ beneficiary type: including worker's compensation, Medicaid, state employees

- What evidence of cost implications and cost-effectiveness of on- or off-label use of use of rhBMP-2 or rhBMP-7 exists? Including consideration of:
    - Costs (direct and indirect) and cost effectiveness
    - Short term and long term

1.3. Washington State utilization and cost data

State Agency Data for Bone Morphogenetic Protein (BMP) use in Spinal Fusion is presented below. There are limitations to these data including:

- BMP use is reported by ICD-9 procedure code related to a DRG without associated billing; therefore only identifiable in hospital claims. Because of this, it is possible that a) the reported BMP usage underestimates actual utilization and b) reported "non-BMP" fusions include a mix of surgeries with and without BMP.
- No estimates of Medicaid BMP use are available. The non-billable ICD-9 procedure code does not appear to be captured in Medicaid data.
- It is possible that this data underestimates total fusions by omitting those done in non-hospital settings (ASC or outpatient) where BMP use cannot be identified in agency administrative data.
- We did not categorize fusions by number of levels or by auto- or allograft.

Figure 1. Combined Agency Data, BMP use in Spinal Fusion, 2007-2010

Agency/Year	2007	2008	2009	2010	4 years overall
PEB Spinal Fusions					
Total Payments for Spinal Fusion	$5,283,336	$7,388,972	$13,250,116	$13,894,609	$39,817,033
Procedure Count	247	309	434	414	1404
Average Age	57.7	58.8	58.7	58.8	58.5
% Male	45.3%	37.9%	40.3%	41.8%	41.1%
Average length of stay	3.5	3.2	3.1	3.1	3.2
PEB Spinal Fusions reporting BMP use					
Procedure Count (% of total)	14.6%	10.0%	11.5%	10.4%	11.3%
Cervical Procdures (% of BMP)	13.9%	12.9%	12.0%	7.0%	11.3%
Avg Length of Stay (all BMP)	3.6	3.5	3.9	4.0	3.8
L&I Spinal Fusions					
Total Payment for Spinal Fusion	$16,602,621	$19,419,086	$21,386,792	$22,203,502	$79,612,001
Procedure Count	593	647	708	739	2687
Average Age	57.7	58.8	58.7	58.8	58.5
% Male	71.6%	70.8%	70.1%	69.3%	70.5%
Average length of stay	4.0	3.7	3.5	3.5	3.6
L&I Spinal Fusions reporting BMP use					
Procedure Count (% of total)	16.4%	16.8%	14.0%	14.9%	15.4%
Cervical Procdures (% of BMP)	8.2%	10.1%	10.1%	11.8%	10.1%
Avg Length of Stay (all BMP)	4.3	4.2	4.1	4.2	4.2
Medicaid Spinal Fusions					
Total Payment for Spinal Fusion	$6,555,328	$7,497,656	$8,007,877	$2,193,720	$24,254,581
Procedure Count	381	407	435	216	1439
Average Age	46.1	45.0	46.5	46.0	45.8
% Male	51.2%	47.5%	45.7%	49.0%	48.1%
Average length of stay	3.1	3.2	2.8	2.5	3.0
Spinal Fusions reporting BMP use were not identified in Medicaid data					

Figure 2a: PEB Spinal Fusion Costs: BMP fusions vs overall fusions, 2007-2010

PEB	2007	2008	2009	2010	4 years overall
Fusions using BMP					
Total Cost for Fusions	$848,914	$938,088	$1,796,711	$1,392,571	$4,976,284
Patient count	36	31	50	43	159†
Procedure count	36	31	50	43	160
*Average per patient, Primary payer only	$39,763	$51,321	$55,357	$54,828	$51,051
Average per patient	$23,581	$30,261	$35,934	$32,385	$31,297
Median	$22,496	$26,094	$30,747	$35,512	$26,678
Std Dev	$24,949	$40,733	$35,396	$31,004	$33,342
Max	$86,603	$197,573	$157,941	$112,850	$197,573
*Min, Primary payer only	$11,353	$5,410	$16,463	$16,191	$5,410
Min	$0	$1,024	$1,068	$1,100	$0
Spinal Fusions Overall					
Total Cost for Fusions	$5,283,336	$7,388,972	$13,250,116	$13,894,609	$39,817,033
Patient count	245	307	425	409	1356†
Procedure count	247	309	434	414	1404
*Average per patient, Primary payer only	$32,846	$39,985	$48,364	$50,930	$45,498
Average per patient	$21,565	$24,068	$31,177	$33,972	$29,364
Median	$12,855	$13,573	$18,844	$22,139	$17,496
Std Dev	$30,424	$36,555	$41,590	$42,726	$42,001
Max	$240,792	$324,809	$394,072	$376,923	$435,892
Min	$0	$0	$0	$0	$0

* Calculated excluding members with partial payment by other insurance payers.

†Count of unique patients over 4 years

Figure 2b: L&I Spinal Fusion Costs: BMP fusions vs overall fusions, 2007-2010

L&I	2007	2008	2009	2010	4 years overall
Fusions using BMP					
Total Cost for Fusions	$3,567,724	$4,351,499	$3,556,437	$4,193,792	$15,669,452
Patient count	97	109	99	110	402†
Procedure count	97	110	99	110	415
Average per patient	$36,781	$39,922	$35,924	$38,125	$38,312
Median	$28,864	$33,211	$31,780	$32,879	$32,134
Std Dev	$39,217	$21,511	$14,996	$16,581	$26,546
Max	$391,670	$136,790	$105,632	$114,505	$431,667
Min	$16,253	$23	$14,496	$15,946	$14,496
Spinal Fusions Overall					
Total Cost for Fusions	$16,602,621	$19,419,086	$21,386,792	$22,203,502	$79,612,001
Patient count	590	643	703	734	2557†
Procedure count	593	647	708	739	2687
Average per patient	$28,140	$30,201	$30,422	$30,250	$31,135
Median	$24,643	$26,907	$29,105	$29,224	$28,145
Std Dev	$22,201	$20,314	$15,417	$15,215	$21,002
Max	$391,670	$180,008	$147,911	$148,221	$431,667
Min	$7,628	$23	$9,112	$85	$1,005

†Count of unique patients over 4 years

Figure 2c: Medicaid Spinal Fusion Costs: BMP fusions vs overall fusions, 2007-2009

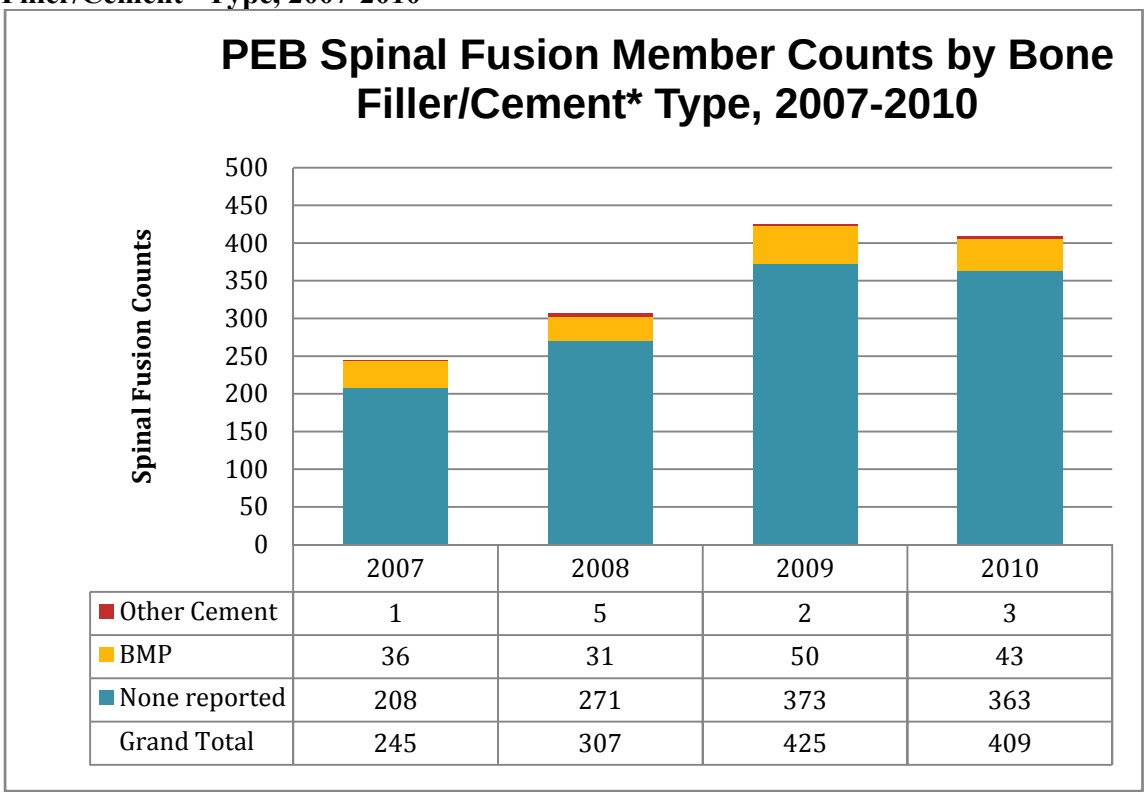
We were unable to distinguish BMP fusions from non-BMP fusions in Medicaid claims.

Medicaid Spinal Fusion Costs	2007	2008	2009	2010*	4 years overall
Spinal Fusions Overall					
Total Cost for Fusions	$6,555,328	$7,497,656	$8,007,877	$2,193,720	$24,254,581
Patient count	371	397	426	213	1363†
Procedure count	381	407	435	216	1439
Average per patient	$17,669	$18,886	$18,798	$10,299	$17,795
Median	$13,843	$14,954	$14,420	$4,783	$13,669
Std Dev	$14,256	$15,432	$15,137	$13,199	$15,601
Max	$94,146	$93,828	$138,497	$83,357	$138,497
Min	$0	$0	$0	$0	$0

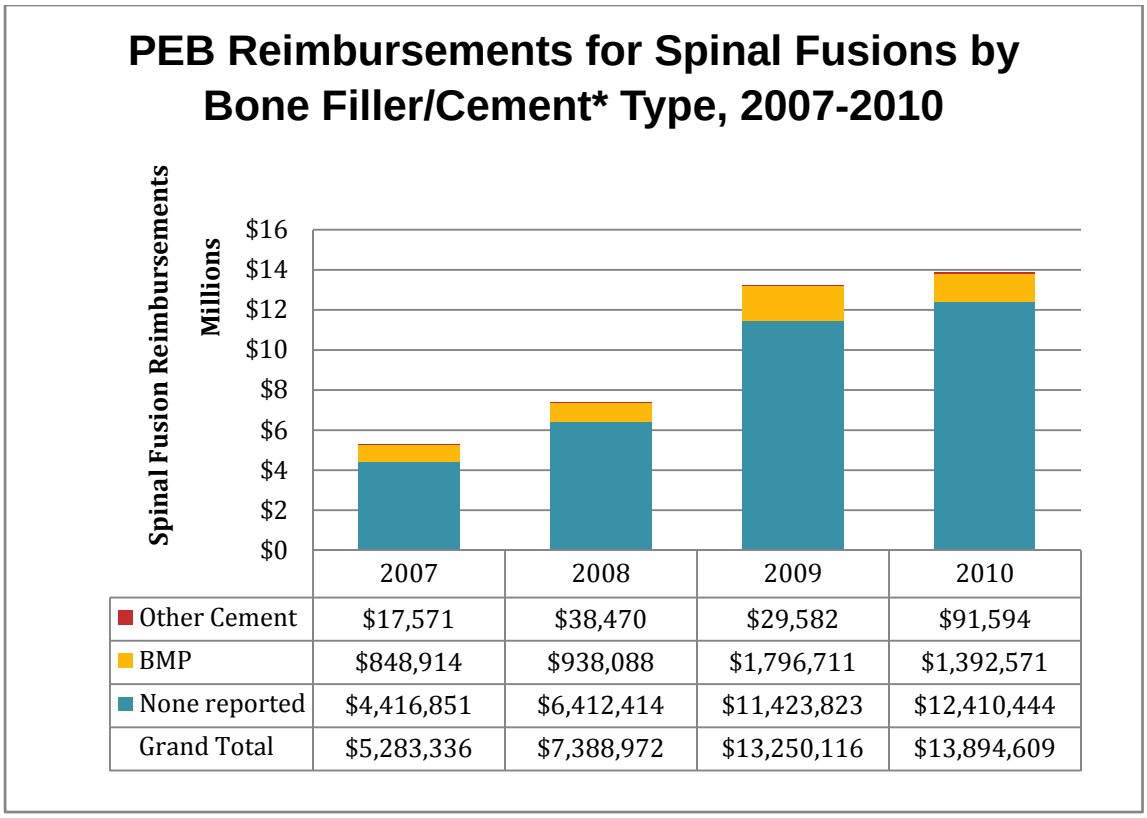
*Spinal fusions required prior authorization as of 4/1/2010

†Count of unique patients over 4 years

Figure 3a: PEB Spinal Fusion Member Counts and Costs by Reported Bone Filler/Cement* Type, 2007-2010



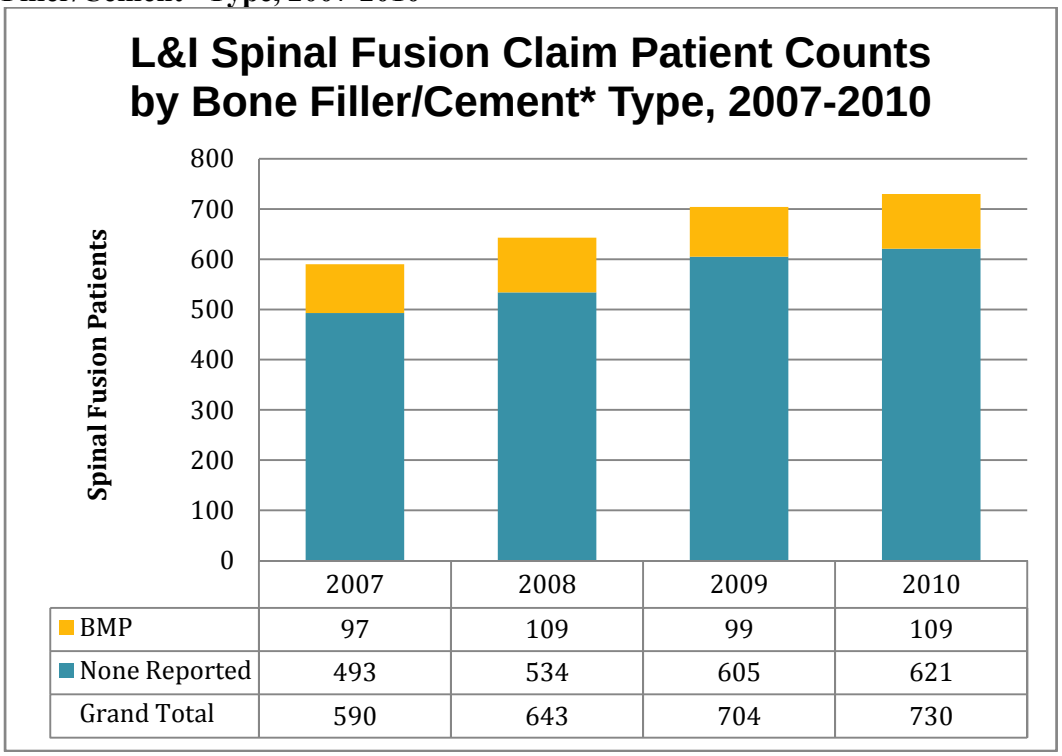
	2007	2008	2009	2010
Other Cement	1	5	2	3
BMP	36	31	50	43
None reported	208	271	373	363
Grand Total	245	307	425	409



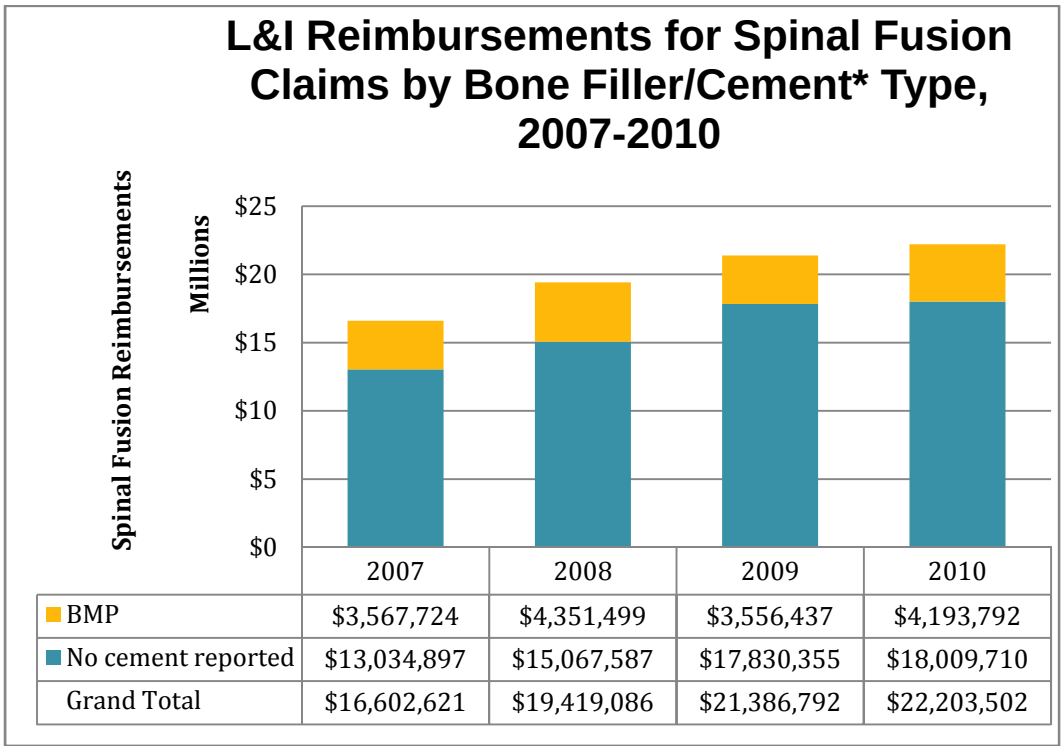
	2007	2008	2009	2010
Other Cement	$17,571	$38,470	$29,582	$91,594
BMP	$848,914	$938,088	$1,796,711	$1,392,571
None reported	$4,416,851	$6,412,414	$11,423,823	$12,410,444
Grand Total	$5,283,336	$7,388,972	$13,250,116	$13,894,609

*see ICD-9 Procedure codes 84.52, 84.55 for description (Related Medical Codes below)

Figure 3b: : L&I Spinal Fusion Member Counts and Costs by Reported Bone Filler/Cement* Type, 2007-2010



	2007	2008	2009	2010
BMP	97	109	99	109
None Reported	493	534	605	621
Grand Total	590	643	704	730



	2007	2008	2009	2010
BMP	$3,567,724	$4,351,499	$3,556,437	$4,193,792
No cement reported	$13,034,897	$15,067,587	$17,830,355	$18,009,710
Grand Total	$16,602,621	$19,419,086	$21,386,792	$22,203,502

*see ICD-9 Procedure codes 84.52, 84.55 for description (Related Medical Codes below)

Figure 4a: PEB Spinal Fusions by type of Fusion, Fusions reporting BMP use, 2007-2010

Fusions using BMP by DRG type	Patient Counts by year					Total Payments for Spinal fusions using BMP by DRG Type					Avg Lnth of Stay
Description and DRG (MSDRG/APDRG)	2007	2008	2009	2010	4 year total	2007	2008	2009	2010	4 year total	
Spinal fusion exc cerv w MCC/CC **(M459, 755)**	0	0	1	2	3	$0	$0	$1,068	$2,200	$3,268	5.3
As above wo CC/ MCC **(M460, 756)**	29	22	40	35	126	$646,065	$403,685	$1,448,064	$1,292,273	$3,790,087	3.8
Ant/Post w CC (M454, 806)	0	2	0	1	3	$0	$255,541	$0	$52,428	$307,969	6.0
Ant/Post wo CC/MCC **(M455, 807)**	1	1	0	0	2	$40,731	$69,781	$0	$0	$110,512	4.0
Spinal Fus ex cerv w curv/malig/infec, w MCC **(M456, 884)**	0	0	2	1	3	$0	$0	$90,641	$1,100	$91,741	16.0
As above w CC **(M457, 884)**	1	0	0	0	1	$992	$0	$0	$0	$992	3.0
As above wo CC/ MCC **(M458, 884)**	0	2	1	1	4	$0	$2,048	$157,941	$1,100	$161,089	3.3
Cervical spinal fusion w CC **(M472, 864)**	0	0	1	0	1	$0	$0	$16,463	$0	$16,463	3.0
Cervical spinal fusion wo CC/MCC **(M473, 865)**	5	4	5	3	17	$161,126	$207,033	$82,534	$43,470	$494,163	1.6

Figure 4b: PEB Spinal Fusions by type of Fusion, All Fusions

All Fusions by DRG	Patient Counts by year					Total Payments for All Spinal Fusions by DRG Type					Avg Lnth of Stay
Description and DRG (MSDRG/APDRG)	2007	2008	2009	2010	4 yr total	2007	2008	2009	2010	4 year total	
Spinal fusion exc cerv w MCC/CC **(M459, 755)**	7	4	5	8	24	$277,825	$4,096	$5,340	$35,180	$322,441	7.2
As above wo CC/MCC **(M460, 756)**	116	147	222	200	685	$2,208,224	$3,354,788	$7,133,950	$8,027,513	$20,724,475	3.5
Ant/Post w MCC **(M453, 806)**	1	3	2	2	8	$992	$144,163	$439,823	$104,612	$689,590	14.3
Ant/Post w CC **(M454, 806)**	6	5	3	5	19†	$150,523	$322,542	$2,136	$199,632	$674,833	5.1
Ant/Post wo CC/MCC **(M455, 807)**	5	8	19	10	42	$184,136	$294,783	$418,308	$308,638	$1,205,865	3.7
Spinal Fus ex cerv w curv/malig/infec, w MCC **(M456, 884)**	1	1	6	8	16	$992	$51,350	$436,406	$595,121	$1,083,869	13.4
As above w CC **(M457, 884)**	7	6	11	9	33	$414,451	$267,609	$503,004	$680,222	$1,865,286	7.2
As above wo CC/MCC **(M458, 884)**	4	9	12	10	35	$417,458	$967,193	$1,040,617	$959,155	$3,384,423	4.6
Cervical spinal fusion w MCC **(M471, 864)**	7	2	6	6	21	$156,440	$2,048	$181,221	$170,347	$510,056	6.0
Cervical spinal fusion w CC **(M472, 864)**	10	14	30	23	77	$202,768	$159,789	$500,839	$406,460	$1,269,856	2.3
Cervical spinal fusion wo CC/MCC **(M473, 865)**	83	110	118	133	444	$1,269,527	$1,820,611	$2,588,472	$2,407,729	$8,086,339	1.4

† Count of unique patients over 4 years

Figure 4c: L&I Spinal Fusions by type of Fusion, Fusions reporting BMP use

Fusions using BMP by DRG type	Patient counts by Year					Total payments for Spinal Fusions using BMP by DRG type					Avg Lnth of Stay
Description and DRG (MSDRG/APDRG)	2007	2008	2009	2010	4 Yr Total	2007	2008	2009	2010	4 Year Total	
Spinal fusion exc cerv w MCC/CC **(M459, 755)**	20	29	18	25	90†	$1,152,572	$1,312,362	$740,562	$1,027,821	$4,233,317	5.4
As above wo CC/MCC **(M460, 756)**	64	64	65	62	252†	$1,985,244	$2,441,784	$2,290,477	$2,257,684	$8,975,189	4.1
Ant/Post w CC **(M454, 806)**	1	2	1	3	7	$56,807	$148,703	$105,632	$158,389	$469,530	6.6
Ant/Post wo CC/MCC **(M455, 807)**	3	3	3	7	16	$142,365	$206,119	$130,970	$460,216	$939,670	5.9
Spinal Fus ex cerv w curv/malig/infec, w MCC **(M456)**	0	0	0	0	0						
As above w CC **(M457)**	0	0	0	0	0						
As above wo CC/MCC **(M458, 884)**	0	0	1	0	1			$23,212		$23,212	
Cervical spinal fusion w CC (**M472, 864)**	3	3	0	3	9	$90,489	$48,624		$60,751	$199,864	2.6
Cervical spinal fusion wo CC/MCC **(M473, 865)**	5	8	10	10	33	$92,768	$193,906	$234,451	$228,932	$750,056	2.3
Spinal procedures w/CC **(836)**			1		1			$31,132		$31,132	3.0
Spinal procedures wo CC **(837)**	1				1	$47,479				$47,479	6.0

† Count of unique patients over 4 years

Figure 4d: L&I Spinal Fusions by type of Fusion, All Fusions

All Fusions by DRG type	Patient counts by Year					Total Payments for All Spinal Fusions by DRG Type					Avg Lnth of Stay
Description and DRG (MSDRG/APDRG)	**2007**	**2008**	**2009**	**2010**	**4 Yr Total**	**2007**	**2008**	**2009**	**2010**	**4 Year Total**	
Spinal fusion exc cerv w MCC/CC **(M459, 755)**	82	92	83	90	326†	$3,541,628	$3,880,712	$3,322,955	$3,523,592	$14,268,887	5.2
As above wo CC/MCC **(M460, 756)**	216	220	264	263	911†	$6,943,759	$7,784,966	$9,179,454	$9,567,237	$33,475,416	4.0
Ant/Post w CC (M454, 806)	8	4	11	17	39†	$645,711	$503,945	$882,007	$958,427	$2,990,090	7.7
Ant/Post wo CC/MCC **(M455, 807)**	9	7	18	19	43	$358,113	$590,363	$872,898	$941,451	$2,762,825	5.0
Spinal Fus ex cerv w curv/malig/infec, w MCC **(M456)**	0	0	0	0	0						
As above w CC **(M457)**	0	0	0	0	0						
As above wo CC/MCC **(M458, 884)**	0	2	3	2	7	$0	$25,267	$116,072	$86,071	$227,411	8.1
Cervical spinal fusion w CC **(M472, 864)**	40	53	54	51	192†	$841,445	$1,197,064	$1,278,884	$1,143,458	$4,460,851	3.1
Cervical spinal fusion wo CC/MCC **(M473, 865)**	229	255	266	288	1005†	$4,057,993	$4,984,575	$5,520,808	$5,906,104	$20,469,480	2.4
Spinal procedures w/CC **(836)**	2	3	3	1	9	$124,235	$125,597	$193,188	$30,589	$473,610	12.3
Spinal procedures wo CC **(837)**	4	7	1	3	15	$89,736	$326,597	$20,526	$46,573	$483,431	6.3

† Count of unique patients over 4 years

Figure 4e. Medicaid Spinal Fusions by type of Fusion, All Fusions (BMP use could not be identified in Medicaid claims)

All Fusions by DRG type	Patient counts by Year					Total Payments for All Spinal Fusions by DRG Type					Avg Lnth of Stay
Description and DRG (MSDRG/APDRG)	2007	2008	2009	2010	4 Yr Total	2007	2008	2009	2010	4 Year Total	
Spinal fusion exc cerv w MCC/CC pag **(M459, 755)**	90	61	47	24	221†	$1,943,259	$1,736,453	$1,380,126	$360,652	$5,420,490	4.1
As above wo CC/MCC **(M460, 756)**	167	88	97	39	380†	$2,623,167	$1,487,670	$1,753,427	$439,895	$6,304,160	2.9
Ant/Post w CC **(M454, 806)**	10	5	14	7	36	$329,926	$317,188	$631,117	$160,668	$1,438,899	11.6
Ant/Post wo CC/MCC **(M455, 807)**	3	5	6	6	20	$76,317	$198,393	$164,578	$92,304	$531,592	3.5
Spinal Fus ex cerv w curv/malig/infec, w MCC **(M456)**	0	0	0	0	0						
As above w CC **(M457)**	0	0	0	0	0						
As above wo CC/MCC **(M458, 884)**	10	24	24	9	66†	$480,788	$1,215,372	$1,256,541	$243,035	$3,195,736	2.6
Cervical spinal fusion w CC **(M472, 864)**	36	83	81	50	247†	$485,421	$1,158,588	$1,116,546	$386,639	$3,147,195	1.6
Cervical spinal fusion wo CC/MCC **(M473, 865)**	64	137	161	79	437†	$616,451	$1,383,991	$1,705,541	$510,525	$4,216,508	5.9

† Count of unique patients over 4 years

Related Medical Codes			
Code Type	**Codes**	**Short Description**	**Additional Info**
ICD9 Procedures			
	84.52	Insertion of recombinant bone morphogenetic rhBMP via collagen sponge, coral, ceramic and other carriers	Procedure/product under review - BMP
	84.55	Insertion of bone void filler, insertion of acrylic cement (PMMA) bone void cement calcium based bone void filler polymethylmethacrylate (excepting vertebroplasty and vertegral augmentation)	Alternative cement used in spinal fusions
APDRG	MSDRG		Main Procedure
	M453	Combined anterior/posterior spinal fusion w MCC	Spinal Fusion
806	M454	Combined anterior/posterior spinal fusion w CC	Spinal Fusion
807	M455	Combined anterior/posterior spinal fusion w/o CC/MCC	Spinal Fusion
	M456	Spinal fus exc cerv w spinal curv/malig/infec or 9+ fus w MCC	Spinal Fusion
	M457	Spinal fus exc cerv w spinal curv/malig/infec or 9+ fus w CC	Spinal Fusion
884	M458	Spinal fus exc cerv w spinal curv/malig/infec or 9+ fus w/o CC/MCC	Spinal Fusion
755	M459	Spinal fusion except cervical w MCC	Spinal Fusion
756	M460	Spinal fusion except cervical w/o MCC	Spinal Fusion
	M471	Cervical spinal fusion w MCC	Spinal Fusion
864	M472	Cervical spinal fusion w CC	Spinal Fusion
865	M473	Cervical spinal fusion w/o CC/MCC	Spinal Fusion
836		Spinal procedures w/CC	Spinal Fusion
837		Spinal procedures wo CC	Spinal Fusion

2. Background

2.1. Bone repair

The skeletal system serves two main functions in the human body: to support and protect vital organs and provide scaffolding for musculature attachment; and to function as a reserve for calcium and phosphate, which are necessary elements for proper metabolic function. Bone is comprised of extracellular matrix (the majority of which is mineralized), collagen, and cells. There are two types of bone in the skeletal system: cortical and trabecular bone. Cortical bone makes up 80% of the skeleton and provides much of the strength associated with the skeleton. The remaining 20% is comprised of trabecular bone. Trabecular bone is found on the inner portions of large flat bones as well as in internal portions of the vertebrae, where it provides mechanical support and serves as a mineral supply. The bone matrix itself is made up of type I collagen fibers and non-collagenous proteins, and hydroxyapatite crystals surround the collagen fibers. Three types of cells necessary for bone growth may be found within the bone matrix: osteoblasts, osteoclasts, and osteocytes. Osteoblasts are cells responsible for the production of new bone, while osteoclasts participate in bone resorption through a process of acidification and proteolysis of the bone matrix. Osteocytes refer to osteoblasts that are trapped within the unmineralized bone matrix, or osteoid[37].

Bone remodeling refers to the continual process by which old bone is replaced by new bone. Unlike many other tissues, which form scar tissue when injured, when a bone is fractured the healing process results in the formation of completely new bone tissue. Fracture healing takes place in four stages. First, a hematoma forms at the fracture site, triggering an inflammatory response that draws repair cells, such as fibroblasts, osteoblasts and endothelial cells. Next, a soft callus develops, consisting of connective tissues and newly formed capillaries. The callus then hardens as the osteoblasts lay down new bone matrix. This bone matrix is randomly arranged and lacks mechanical strength, which results in the final phase of fracture healing, the remodeling phase[38]. During this phase, which can last years, mechanical stresses placed on the bone result in a slow restoration to its original shape[39]. Bone remodeling has numerous systemic regulators, including parathyroid and thyroid hormones, glucocorticoids, calcitriol, calcitonin, estrogens, and androgens[37].

One of the key aspects to successful fracture healing is the formation of new osteoblasts. This process is known as osteoinduction, wherein mesenchymal stem cells (MSC) differentiate to form osteoblasts. Osteoinduction is promoted by three different subgroups of molecules: inflammatory molecules that initiate the repair sequence, angiogenic factors which promote the development of vascular tissue, and molecules that regulate growth and differentiation such as the TGF-ß superfamily. Bone morphogenetic proteins (BMPs) are members of the TGF-ß superfamily and are produced by MSCs and osteoblasts in the bone matrix. BMPs act on mesenchymal osteoprogenitor cells and osteoblasts to promote the development of additional osteoblasts and the release of further growth factors[40].

2.2. The condition

Degenerative disc disease (DDD) has been extensively discussed as an etiology for back pain and tends to appear in the lumbar spines of healthy adults between the ages of 30 and 50 years. This condition occurs when deterioration or damage of the outer ring of the vertebra exposes the nerve root to the contents of the disc. Although the initial tear in the annulus tends to heal, it is left prone to re-injury, resulting in further desiccation of the disc with each episode. Pain can be felt in the back and legs and originates from a number of sources. The injured disc releases inflammatory chemicals, while the torn annulus triggers local prostaglandin release. The arthritic facets can also be a source of pain. Surgical intervention is often undertaken to alleviate this painful condition through fusion of the arthritic joint[41]. Cervical DDD is less frequent, although the treatments are similar[42].

Spondylolisthesis refers to the displacement of one vertebra onto another and may occur in either an anterior or posterior position relative to the vertebra below. The causes of spondylolisthesis include disease (pathologic spondylolisthesis) or fracture (traumatic spondylolisthesis). In the case of isthmic spondylolisthesis, the cause is a stress, lytic fracture, an elongated (but intact) pars, or an acute fracture of the pars interartcularis. Degenerative spondylolisthesis occurs as a consequence of aging and results from deterioration of the spine. The severity of spondylolisthesis is graded based on the extent to which one vertebra has moved relative to the caudal vertebra: grade I is a translation of up to 25%; grade II is 26-50%; grade III, 51-75%; grade IV, 76-100%; and grade V, which is greater than 100%, is classed as spondyloptosis. The majority of spondylolisthesis cases are grade I (75%), and 20% are grade II[43].

Spinal stenosis refers to narrowing of the spinal column or of the neural foramina. The former leads to pressure on the spinal cord, while the latter causes compression of the surrounding nerves. Spinal stenosis causes back and sometimes leg pain. Spinal surgery is used to relieve pressure on the spinal nerves through either laminectomy or decompression[42].

2.3. The technology and its comparators

Autogenous iliac crest bone graft (ICBG) harvested from the posterior iliac crest (hip) has long been the "gold standard" for patients receiving spinal fusion. In addition to being osteogenic, osteoconductive, and osteoinductive, ICBG has the added benefit of being histocompatible and non-immunogenic. However, autogenous bone graft harvesting can result in pain, infection, nerve and artery damage, as well as an increased risk of stress factor at the harvest site. Harvest site pain is the primary motivation to pursue technologies other than autograft[39, 44].

Allograft bone (bone from another person) is also used in spinal fusion surgeries. It is osteoconductive and might be osteoinductive. Drawbacks include the small, unproven risk of infectious disease transmission, immunological reaction, and limited availability. There are a variety of bone graft substitutes on the market, including demineralized bone matrix, which is made from allograft bone, collagen, noncollagenous proteins and growth factors. As result of the

excessive processing it undergoes, demineralized bone matrix (DBM) is the least immunogenic of allograft bone options[39]. Other bone graft substitutes include allograft bone, hydroxyapatite blocks or granules, various calcium-based granules (such as ß-tricalcium phosphate and calcium sulfate), and injectable cements. Each of these potential treatments varies in terms of osteoconductivity and osteoinductivity[45].

BMP products have been developed to serve as substitutes to autograft in spinal fusion procedures that promote the same high rate of fusion observed in patients who undergo spinal arthrodesis with ICBG. Two BMP products have been approved for use in spinal fusion procedures: InFUSE (recombinant human (rh) BMP-2), and OP-1 (rhBMP-7)[3, 22].

InFUSE is manufactured by Medtronic Sofamor Danek, based in Memphis, TN. InFUSE consists of a metallic tapered fusion cage and a bone graft substitute consisting of rhBMP-2 along with a carrier/scaffold for the protein made from bovine Type I collagen that is absorbed into a sponge (ACS, or absorbable collagen sponge) and placed inside the fusion cage. InFUSE was approved by the FDA in July 2002 as part of the Premarket Application (PMA) approval process for the treatment of DDD in the L4-S1 region of the spine, with up to Grade I spondylolisthesis[22]. Although InFUSE was also granted Humanitarian Device Exemption (HDE) in October 2008 for use in the repair of posterolateral lumbar spine pseudarthrosis in patients for whom autologous graft was not feasible, Medtronic voluntarily withdrew HDE approval for InFUSE in early 2010.

Medtronic is also in the processes of seeking premarket approval of a second rhBMP-2 product, AMPLIFY (P050036)[46]. The following indications for use have been proposed by Medtronic in the pre-market approval application: spinal fusion procedures in skeletally mature patients with DDD at one level from L1 – S1 who have failed at least six months of conservative treatment; DDD patients can also have up to Grade 1 spondylolisthesis or retrolisthesis at the involved level. The product has been designed to be implanted via a posterolateral approach and used in conjunction with a metallic posterior spinal fixation system intended for temporary stabilization of the spine. While the form of rhBMP-2 used is identical in both AMPLIFY and InFUSE, AMPLIFY utilizes a different carrier/scaffold than InFUSE, a CRM (compression resistant matrix) that consists of an absorbable bovine Type I collagen with embedded biphasic calcium phosphate granules. AMPLIFY also differs from InFUSE in that a dose of 40 mg is delievered to each per patient compared with 4.2 to 8.4 mg per patient for InFUSE[22, 46]. Concerns about an increased risk of malignancy with AMPLIFY were noted by the FDA Orthopaedic and Rehabilitation Devices Advisory Panel in the July 2010 Executive Summary[46] as well as in a recent safety review by Carragee[47] with the FDA stating that the higher number of cancers in the rhBMP-2 group compared with the control group warrants further investigation.

OP-1 is manufactured by Stryker Biotech in Hopkinton, MA. The product consists of rhBMP-7 and bovine collagen mixed with a sterile saline solution to form a paste, which is placed between the broken ends of a bone during surgery. OP-1 first received FDA approval in October, 2001 for use in as an alternative to autograft in "recalcitrant long bone non-unions". In April 2004, OP-1 was approved under the FDA's HDE for use in revision posterolateral spinal fusion where

use of autograft was contraindicated[3]. HDE is issued for devices or treatments that affect fewer than 4000 individuals in the United States per year.

2.4. Indications and contraindications

InFUSE bone graft is indicated for anterior open or anterior laparoscopic spinal fusion procedures in skeletally mature patients with degenerative disc disease at one level between L4 and S1 (and up to Grade 1 spondylolisthesis at the involved level) and who have failed at least six months of nonoperative care[22]. InFUSE is contraindicated in patients with a known hypersensitivity to rhBMP-2, bovine Type I collagen or to other components of the formulation; in the vicinity of a resected or extant tumor or any active malignancy or patients undergoing treatment for malignancy; in pregnant women; and in patients with an active infection at the operative site[22].

OP-1 is indicated for use as an alternative to autograft in compromised patients requiring revision posterolateral lumbar spinal fusion for whom autologous bone and bone marrow harvest are not feasible or are not expected to promote fusion. Examples of compromising factors include osteoporosis, smoking and diabetes[3]. OP-1 is contraindicated in patients who have a known hypersensitivity to the active substance or to collagen; at or near the vicinity of a resected tumor or in patients with a history of malignancy; in patients who are skeletally immature (<18 years of age or no radiographic evidence of closure of epiphyses); and in pregnant women[3].

2.5. On-label versus off-label use

"On-label use" refers to use of a BMP product in accordance with its FDA-approved indications. These indications are summarized briefly in Table 1, below.

Table 1. FDA-approved indications for BMP products in the spine: on-label uses

Product	Devices	FDA approval	Indication(s)
InFUSE PMA (P00058)	LT-Cage or Inter Fix Threaded Fusion devices	PMA (P000058)[22] (2002)	InFUSE bone graft is indicated for anterior open or anterior laparoscopic spinal fusion procedures in skeletally mature patients with degenerative disc disease at one level between L4 and S1 (and up to Grade 1 spondylolisthesis at the involved level) and who have failed at least six months of nonoperative care.
OP-1 Putty	N/A	HDE (H020008)[3] (2004)	OP-1 is indicated for use as an alternative to autograft in compromised patients requiring revision posterolateral lumbar spinal fusion for whom autologous bone and bone marrow harvest are not feasible or are not expected to promote fusion.

"Off-label use" refers to use of a BMP product not in accordance with its FDA-approved indications. In the last decade since InFUSE received FDA approval for use in anterior spinal

fusion at one level between L4 and S1, the product has been used more and more widely for a variety of spinal fusion procedures[47]. In a 2009 comment to Spine Journal, Vaidya stated the following regarding rhBMP-2 use in spinal fusions[48]: "We have used it in ways that were not originally approved by the FDA because we felt, if it works so well for one indication; why not try it for others. Many of us read early articles on off-label use which showed the results were excellent in the c-spine and in PLIF and TLIF surgery."

Due to the widespread use of BMPs in spinal fusion procedures, we have divided the results up according to the way in which the product was used:

- On-label use of rhBMP-2 in the lumbar spine
- On-label use of rhBMP-7 in the lumbar spine
- Off-label use of rhBMP-2 in the lumbar spine
- Off-label use of rhBMP-7 in the lumbar spine
- Off-label use of rhBMP-2 in the cervical spine
- Off-label use of rhBMP-7 in the cervical spine

Note that the studies used for FDA approval of OP-1 are considered to be off-label, as they are not in accordance with on-label use. Specifically, the studies evaluated primary (rather than revision) fusion; furthermore, the enrolled patients were not necessarily compromised patients for whom autologous bone and bone marrow harvest were not feasible or expected to promote fusion. As state in the FDA SSBP for InFUSE: "Based on a pilot clinical study, OP-1 Putty has demonstrated probable benefit as an alternative to autograft in patients who required a primary uninstrumented fusion for the treatment of degenerative spondylolisthesis. While these data cannot be directly extrapolated to the expected performance of OP-1 Putty in revision posterolateral spinal fusions in the compromised population, there is reason to believe that OP-1 Putty could have a probable benefit in this population"[3].

2.6. Surgical procedures

Spinal fusion is achieved by several different approaches and instrumentations, which are summarized below.

The oldest technique is posterolateral fusion (PLF), which has been in use since the 1950s. In the PLF technique, a bone graft is placed in the posterior lateral gutters between the transverse processes of the levels above and below. This technique can be performed with or without instrumentation. PLF has the advantage of an easy approach, low risk of complications, and surgeon familiarity.

Posterior lumbar interbody fusion (PLIF) is a surgical approach via the posterior side. A laminectomy and partial facetectomy are performed to access the disc space for discectomy, bone graft, and optional cage placement with or without bone. By fusing the degenerative vertebrae, the motion pathology is eliminated with possible resulting pain relief. An incision is made over the vertebrae and paraspinal muscles are dissected to the vertebral facets' lateral margins. A bilateral laminotomy is then performed and, with the inferior articular face exposed,

disc material is removed from the disc space. Pedical screw fixation with rods is often used to provide supplemental fixation.

Transforaminal lumbar interbody fusion (TLIF) is similar to PLIF in method and instrumentation; however, the approach is slightly more lateral, as a complete facetectomy is generally performed in a TLIF. As with PLIF, the bone graft is placed into the disc space with or without a cage.

Anterior lumbar interbody fusion (ALIF) is done by approaching the spinal column anteriorly through an incision in the abdomen, either via mini laparotomy or endoscope. This approach preserves the paraspinal musculature and nerves. The bone graft, which can be placed with or without a cage, receives greater compression, which is thought to aid in fusion. Risks from ALIF include damage to blood vessels and retrograde ejaculation in males[49].

Anterior cervical discectomy and fusion (ACDF) is achieved by approaching the cervical spine through an anterior incision in the neck. The entire disc is removed and the disc space filled with bone graft to fuse the vertebrae.

Autogenous bone graft is frequently harvested from the iliac crest (ICBG). The harvest site is approached from either the anterior or posterior side to extract grafts from the anterior or posterior iliac crest, respectively. Bone grafts can be harvested as weight-bearing struts or as morselized cortical and cancellous bone. Grafts maintain viability for about two hours in normal saline solution and it has been suggested that freeze-drying also maintains the properties of a fresh autograft. Nevertheless, time between procurement and transplant of autograft is crucial to the success of the procedure[50]. Autogenous grafts can also be harvested from the primary surgical site; for example, during spinal fusion, surgeons can make bone chips from bits of transverse processes, facet joints and vertebral bodies[45].

2.7. Mechanism of action

BMPs are growth factors, mainly related to bone and cartilage growth. BMPs belong to the TGFß superfamily of growth factors. Structurally, BMPs are homo or heterodimers linked by disulphide bridges that act by binding to the serine-threonine kinase receptors on the surface of a cell, thus triggering intracellular pathways to activate and influence gene transcription. BMP-2 has been shown (among other things) to be stimulate bone and cartilage formation. Similarly, BMP-7 is involved in bone and cartilage morphogenesis[51]. Bone production is stimulated in two ways, either through the differentiation of mesenchymal stem cells into either osteoblasts directly or cartilage cells, which in turn change into bone cells[39].

In contrast to the TGFß superfamily of growth factors to which they belong, BMPs are differentiation factors, meaning that they cause mesenchymal cells to differentiate, in the case of rhBMP-2 and rhBMP-7 into the bone and cartilage forming cells. Both rhBMP-2 and rhBMP-7 share a similar cell receptors, consisting of one of two Type 1 receptors and a Type 2 BMP receptor[52]. Bone growth and differentiation are induced and regulated through a series of negative feedback loops[52]. This step wise mechanism involves the BMP molecules at different

stages of osteogenesis and most likely includes synergistic relationships between the BMP molecules themselves[53].

2.8. Potential complications and harms

A variety of complications have been suggested to be associated with spinal fusion and the use of BMPs in spinal fusion surgery, a number of which are described below.

Bone overgrowth or uncontrolled bone formation may occur if BMP leaks from the carrier into the disc space, resulting in extradiscal, ectopic, or heterotopic bone formation[54]. Both heterotopic and ectopic bone growth refer to bone that forms in abnormal locations, ectopic referring to more severe levels of ossification the heterotopic[55]. This is thought to occur through leakage of BMP from the carrier[53]. Both ectopic and heterotopic ossification can impinge on surrounding nerve and tissue structures, requiring surgery to correct[56].

On the opposite end, there is also the potential for resorption or osteolysis of the graft and/or endplate or graft subsidence, loosening, or migration. The incidence of resorption with BMP has been attributed to increased osteoclastic activity. While resorption and osteolysis are normal parts of the fusion process, endplate resorption can result in cage migration or subsidence. Cage migration occurs when endplate resorption causes the cage to become loose[57]. According to one recent study, endplate resorption with cage migration resulted in reoperation in 31% (8/26) PLIF cases[58]. Cage migration can also lead to heterotopic bone formation in the spinal canal and neural foramen, resulting in severe pain in the leg or back[57]. Subsidence refers to the subsiding of the interbody cage into the disc space, resulting in the narrowing of that space[57] and has been associated with increased incidence of pseudoarthrosis[59].

Retrograde ejaculation has been reported to occur in some male patients as a complication of ALIF. During the surgical approach to the lower lumbar segments, the aortic and vena cava are manipulated, thus impacting the bladder sphincter control during ejaculation[60]. This results in the propulsion of semen from the posterior urethra into the bladder during ejaculation. RE can result in infertility[61]. One recent cohort study suggested that use of rhBMP-2 in ALIF is associated with increased risk of retrograde ejaculation. In this study, 7.2% of patients undergoing ALIF in L5/S1 with rhBMP-2 had an RE event, as opposed to 0.6% of those in the control[60]. Other urogenital complications that can result from lumbar fusion surgery include postsurgical urinary retention and bowel obstruction.

Other adverse neurologic effects include back and leg pain and new onset radiculitis. These symptoms are frequently triggered by the impingement of ectopic bone on the neural structures. In one case study, revision surgery was conducted on three of the five patients included to remove the ectopic bone and decompress the nerve roots, resulting in only partial alleviation of the neurological symptoms[62].

Neck swelling and respiratory complications may result from cervical spinal fusion. Dysphagia and dysphonia refer to difficulty swallowing and difficulty vocalizing sound, respectively, and are believed to be a result of the local inflammatory response to BMP[54]. The frequency of these

complications prompted the FDA to issue a warning in 2008 about the use of rhBMP in the cervical spine due to life-threatening complications resulting from dysphagia, swelling of neck and throat tissue, and airway compromise [63]. According to the statement, in many cases swelling was severe enough to interfere with breathing, and emergency medical intervention was necessary. In addition, surgery was often necessary to alleviate the swelling. Swelling complications are believed to be more common with the use of BMP in cervical fusion as opposed to lumbar fusion due to proximity of the product to the neck. Furthermore, the soft tissue of the neck is very different than the soft tissue of the lumbar spine, and the affected structures of the neck (i.e., the esophagus and trachea) are more likely to be symptomatic. Dosage, carrier and delivery systems for rhBMP-2 have been optimized for use in ALIF and not optimized for use in the cervical spine[64]. A 2010 systematic review on dysphagia following ACDF stated that in two recent studies, as many as 30% (15/50) and 55% (6/11) of patients receiving BMP developed dysphagia[65].

Local complications have been documented in both lumbar and cervical fusion, ranging from wound infection and dehiscence, to hematoma, seroma, and edema[54]. One case study suggested that the presence of seromas and edemas at the operation site are a result of potential inflammatory effects of rhBMP-2[66] while another study posited that the development of painful seromas at the operation site have been underreported, due to lack of further surgical treatment[67]. Significant deep wound infections and epidural hematomas have also been reported as complications associated with the use of BMP in spinal fusion[68].

Elevated antibodies to BMP and the collagen carriers have been noted with the use of both rhBMP-2[54] and rhBMP-7[2] suggesting an immunogenic response. However, a recent HTA suggested that elevated antibodies were not linked to any adverse effects[69].

There are frequent adverse events associated with iliac crest bone graft harvest. Pain, paresthesias, hematoma, and infection are common complications, occurring in upwards of 50% of cases. Furthermore according to one systematic review, as many as 60% of patients experience long term donor site pain, and between 2% and 5% require reoperation as a result of wound complications[70]. Another health technology assessment evaluating two prospective studies concluded that between 18.3% and 31% of patients experienced pain at 24 months postoperatively. Other complications of the iliac crest graft harvest site may include iliac wing fracture, nerve and vascular injuries at the donor site, seromas, and unsightly scarring[42].

2.9. Clinical guidelines

National Guideline Clearinghouse (NGC)
A search of the National Guidelines Clearinghouse for guidelines that addressed bone morphogenetic protein retrieved 24 potential current guidelines, two provided specific guidance and one was a systematic review. A variety of keyword searches were performed, including "*bone morphogenetic protein,*" "*bone,"* "OP-1," and "*INFUSE.*"

Work Loss Data Institute (2011)[71]:
Low back – lumbar & thoracic (acute & chronic)
A summary provided by the NGC indicates that rhBMP was considered as a treatment for workers with low back pain and was not recommended.

Work Loss Data Institute (2011)[72]:
Neck and Upper back (acute & chronic)
A summary provided by the NGC indicates that rhBMP was considered as treatment for workers with occupational disorders of the neck and upper back. rhBMP was considered and not recommended.

National Institute for Health and Clinical Excellence (NICE)
The National Institute for Health and Clinical Excellence (NICE) provides guidance on health technologies and clinical practice for the National Health Service in England and Wales. A variety of keyword searches were performed, including "*BMP*" and "*bone morphogenetic protein."* No guidelines were found.

PubMed
One guideline was found that addressed bone morphogenetic protein, *Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 16: bone graft extenders and substitutes, 2005*[73]. Even though evidence for a treatment guideline is insufficient, rhBMP-2 in combination with hydroxyapatite (HA) and tricalcium phosphate may be used as a substitute for autograft bone for some cases of posterolateral lumbar fusion (PLF).

NIH Consensus Statement
No consensus statement was found for bone morphogenetic protein.

Professional societies/other (Not indexed in NGC)

American Academy of Orthopaedic Surgeons (AAOS)
No guidelines were found that addressed bone morphogenetic protein.

Google and Google Scholar
A keyword search on terms including "clinical guidelines" AND "BMP" retrieved no guidelines.

No clinical guidelines relating to bone morphogenetic protein were found in the following organizations' resources:

- Agency for Healthcare Research and Quality (AHRQ)
- U.S. Food and Drug Administration (FDA)
- U.S. Army Institute of Surgical Research
- The Clinical Orthopaedic Society
- Guidelines International Library
- The New Zealand Guidelines Group
- New Zealand Accident Compensation Corporation
- American College of Physicians

Table 2. Clinical Guidelines

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence base available	Critical appraisal*	Comments	Primary conclusions
Work Loss Data Institute (2011)[71] *Guideline Summary NGC-8517: Low back – lumbar & thoracic (acute & chronic)*	1993-present	rhBMP	NR	NR		rhBMP was considered and not recommended.
Work Loss Data Institute (2011)[72] *Guideline Summary NGC-8518: Neck and Upper back (acute & chronic)*	1993-present	rhBMP	NR	NR		rhBMP was considered and not recommended.
Resnick (2005)[73] *Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 16: bone graft extenders and substitutes*	1966-November, 2003	rhBMP-2 in ALIF and PLF	% f/u, f/u period NR unless specified • 1 RCT (> 90% f/u, ≥ 48 months); N = 279 • 1 RCT; N = 35 • 1 pilot study (17 months); N = 25 • 1 combined analysis; N = NR	Large RCT: Class I All other studies: LOE III	Only papers with Class III evidence or better were considered.	While evidence for a treatment guideline is insufficient, rhBMP-2 in combination with HA and tricalcium phosphate may be used as a substitute for autograft bone in some cases of PLF. rhBMP-2 is a viable alternative to autografts for interbody fusion procedures.

NR: not reported; RCT: randomized controlled trial; LOE: level of evidence; f/u: follow-up; PLF: posterolateral lumbar fusion; ALIF: anterior lumbar interbody fusion; ACDF: anterior cervical discectomy fusion; HA: hydroxyapatite

* Critical appraisal refers to formal evaluation of individual study quality using criteria such as the Jadad or GRADE

methods of scoring and the determination of overall strength of evidence. Criteria used to evaluate study quality not described [73].

2.10. Previous Systematic Reviews/Technology Assessments

Many of the previous health technology assessments (HTA) and systematic reviews (SR) gave measured support to the use of rhBMP-2 in lumbar fusion surgery in patients for whom iliac crest bone graft (ICBG) was not feasible. Garrison states that rhBMP-2 is more effective then autogenous bone graft for radiographic fusion in the treatment of single level DDD[42]. Feldman stated that in the RCTs reviewed, rhBMP-2 and autologous bone graft were identical in terms of fusion rate, disability, pain, and patient satisfaction[74]. The Washington Department of Labor and Industries Office of the Medical Director concluded that rhBMP-2 and autograft bone resulted in similar levels of fusion; however, rhBMP-2 did not have the disadvantage of donor site pain[75]. The New Zealand Accident Compensation Corporation determined that there is no evidence to favor the use of rhBMP-2 in spinal fusion in terms of fusion rates, but rhBMP-2 does reduce the risk of pain and/or complications associated with autograft procedures[76].

The Agency for Healthcare Research and Quality noted that the two RCTs reviewed for the use of rhBMP-2 in spinal fusion were of insufficient size and duration to give satisfactory data [39]. Carragee determined that the industry sponsored trials were either too small to adequately address safety issues or that adverse effects were not reported, conflicts of interest were either unclear or unreported, and trials were constructed so as to bias against ICBG treatment[47].

All of the previous HTAs and SRs did not support the use of rhBMP-2 in cervical fusion due to lack of evidence for effectiveness of the procedure and high levels of ensuing complications. Only one SR, Mroz, mentioned the use of rhBMP-2 in thoracic fusion, stating that evidence was insufficient to give a recommendation[54].

Most HTAs and SRs cited lack of evidence for the safety of efficacy of rhBMP-7 in fusion of the lumbar spine. Only the Australian Safety and Efficacy Register of New Interventional Procedures – Surgical gave any support for the treatment, stating that it could result in improvements to the fusion process, but noted that studies cited in favor of rhBMP-7 were hampered by small sample sizes and low power [69]. The New Zealand Accident Compensation Corporation also stated that evidence for the use of rhBMP-7 in cervical/thoracic fusion was insufficient [76].

Table 3. Overview of previous technology assessments

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence Base Available	Critical Appraisal*	Comments	Primary conclusions
Agency for Healthcare Research and Quality (2010)[39] *Bone Morphogenetic Protein: The State of Evidence of On-Label and Off-Label Use*	January 1998-February 2010	**On-Label** rhBMP-2 for fusion in lumbar-sacral spine rhBMP-7 for fusion in lumbar spine **Off-Label** rhBMP-2 for fusion in cervical spine rhBMP-2 for fusion in lumbar-sacral spine rhBMP-7 for fusion in lumbar-sacral spine	% f/u NR for all studies, f/u period NR unless specified **On-Label** rhBMP-2/ Lumbar-sacral: • 1 RCT (24 months); N = 279 • 1 RCT (24 months); N =14 • 1 pooled comparative analysis (24 months); N = 679, comprising 4 RCTs, including largest RCT above • rhBMP-7/ Lumbar: • 0 comparative studies • 1 Pilot study (12, 24, 48 months); N = 36 **Off-Label** rhBMP-2/ Cervical • 1 RCT (24 months); N = 33 • 4 non-RCTs (1.5 – 36 months); N = 423 rhBMP-2/ Lumbar-sacral	**On-Label** rhBMP-2 for lumbar sacral: moderate overall evidence, fair study quality for both RCTs, poor quality for pooled analysis. rhBMP-7 for lumbar spine: the strength of the evidence is insufficient **Off-Label** rhBMP-2 for lumbar sacral: fair quality for 4 RCTs, poor quality for 2 RCTs; fair quality for 1 comparative study, poor quality for other 4 comparative studies. Evidence is moderate overall. rhBMP-2 for cervical: Moderate overall evidence. 1 fair quality RCT, and 4 poor quality non-randomized studies. rhBMP-7 for lumbar sacral: Good study quality for largest RCT, poor study quality for other RCTs.	HTA also addresses use of rhBMP in long bone non-unions, acute open shaft tibial fractures, and sinus augmentation. Strength for off-label uses is graded only for settings with more than 1 comparative study.	**On-Label** rhBMP-2/ Lumbar-sacral: No significant adverse events attributed to rhBMP-2; however, the 2 RCTs were of insufficient size and duration to yield satisfactory data pertaining to the number and severity of the adverse effects. In both RCTs the rate of radiographic fusion was similar to autograft bone. Moderate support to clinical benefit for use of rhBMP-2 in avoiding additional procedure of autograft bone harvest and associated adverse events. rhBMP-7/ Lumbar: With no comparative studies identified, the evidence is insufficient as to safety or efficacy of rhBMP-7. **Off-Label** rhBMP-2/ Cervical Insufficient evidence regarding radiographic fusion success or changes in neck disability measures. Moderate evidence for increased cervical swelling and related complications. rhBMP-2/ Lumbar-Sacral Evidence supporting radiographic fusion success is moderate and insufficient as to either efficacy or potential complications from this procedure. Evidence that rhBMP-2 improves other outcomes is low. rhBMP-7/Lumbar sacral

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence Base Available	Critical Appraisal*	Comments	Primary conclusions
			• 6 RCTs (12 – 27 months); N = 832 • 5 non-RCTs (3 – 38 months); N = 331 <u>rhBMP-7/ Lumbar</u> • 1 RCT (≥ 36 months); N = 293 • 1 RCT (12 months); N = 20 • 1 RCT (24 months); N = 19 • 1 RCT (48 months); N = 36			Evidence is insufficient to draw conclusions.
Canadian Agency for Drugs and Technologies in Health: Health Technology Inquiry Service (2009)[77] *InFUSE Bone Grafts for Spinal Fusion Surgery: Clinical Effectiveness and Indications for Use in Canada*	2004 - January, 2009	rhBMP-2 for spinal fusion surgery	% f/u, f/u period, N NR for all studies • 3 RCTs • 1 HTA • 3 observational studies	NR		NR

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence Base Available	Critical Appraisal*	Comments	Primary conclusions
Garrison (2007)[42] *Clinical effectiveness and cost-effectiveness of bone morphogenetic proteins in the non-healing of fractures and spinal fusion: a systematic review*	1800 – January, 2006 (updated search Nov. 2006 for RCTs only) Default start dates for all electronic databases were used.	<u>rhBMP-2/Lumbar</u> Posterolateral and anterior fusion <u>rhBMP-7/Lumbar</u> non-instrumented and instrumented posterolateral fusion <u>rhBMP-2/Cervical</u> anterior discectomy	<u>Lumbar:</u> • 7 RCTs using rhBMP-2 (% f/u NR or 82 – 100%, NR or 24 months); N = 14 – 279 • 4 RCTs using rhBMP-7 (% f/u NR or 79.2 – 100%, 12 months); N = 20 – 40 <u>Cervical</u> 1 RCT for rhBMP-2 (97% f/u, 24 months); N = 43	Overall, studies failed to meet many of the basic quality expectations of an RCT and the overall quality ranged from low to moderate. No specific rating was reported for studies.	HTA also included analyses of tibia fractures and scaphoid non-unions. Results from 14 case series were also presented, but were not included in the analysis due to poor study quality.	rhBMP-2 is more effective than autogenous bone graft for radiographic fusion in the treatment of single level DDD; however there is a lack of evidence on the use of rhBMP-7 in the treatment of patients with degenerative spondylolisthesis with spinal stenosis or spondylolysis. There is limited evidence that rhBMP is associated with greater improvement in clinical outcomes. The use of rhBMP is unlikely to be cost effective for spinal fusion.
Australian Safety and Efficacy Register of New Interventional Procedures-Surgical (2006)[69] *Horizon Scanning Report: OP-1 Putty for posterolateral lumbar fusion*	NR-January, 2005	rhBMP-7 for revision posterolateral (intertransverse) lumbar spinal fusion	% f/u, f/u period NR unless specified • 2 RCTs (6 weeks – 36 months); N = 56 • 1 case series (24 months); N = 12 • 1 case report; N = 9	NR	When this HTA was conducted, the OP-1 Putty had not been approved for use in Australia; however OP-1 in the non-putty form is approved.	While the studies on the use of rhBMP-7 cited in this HTA are generally supportive of the treatment, the studies are limited by small sample sizes and low power. However, the use of the OP-1 Putty could result in considerable improvements to the spinal fusion process.

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence Base Available	Critical Appraisal*	Comments	Primary conclusions
Feldman (2005)[74] *California Technology Assessment Forum: Recombinant Human Bone Morphogenetic Protein-2 for Spinal Surgery and Treatment of Open Tibial Fractures*	NR	rhBMP-2/ Lumbar ALIF, PLIF, and PLF rhBMP-2/ Cervical ACDF	% f/u NR for all studies, f/u period NR unless specified rhBMP-2 Lumbar • 1 RCT; N = 14 • 1 RCT; N = 27 • 1 RCT (24 months); N = 279 • 1 RCT (24 months); N = 67 • 1 RCTs (12, 24 months); N = NR • 1 RCT (3, 6, 12, 24 months); N = 46 • 1 prospective study (6, 12 months); N = 22 • 1 undefined study (6, 12 months); N = 43 rhBMP-2 Cervical • 1 RCT (3, 6, 12, and 24 months); total N = NR, rhBMP-2 arm n = 18	NR	HTA also included analysis of tibial fractures.	rhBMP-2 is equivalent to autologous bone graft for use in spinal fusion. In the RCTs examined, outcomes for both the rhBMP-2 group and the control were similar in terms of disability, pain, fusion rate, and patient satisfaction. However, the use of rhBMP-2 in cervical spine procedures does not meet this criteria and is not recommended due to lack of evidence.

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence Base Available	Critical Appraisal*	Comments	Primary conclusions
Medical Advisory Secretariat: Ontario Ministry of Health and Long Term Care (2004)[78] *Bone Morphogenetic Proteins & Spinal Surgery for Degenerative Disc Disease: An Evidence-Based Analysis*	1966 – November, 2003	rhBMP-2 in ALIF, PLIF, PLF and cervical fusion	% f/u NR for all studies, f/u period NR unless specified <u>rhBMP-2 in ALIF</u> • 1 RCT (24 months); N = 297 • 1 RCT (24 months); N = 134 • 1 RCT (6 months); N = 14 • 1 RCT (24 months); N = 47 <u>rhBMP-2 in PLIF</u> • 1 undefined study; N = 67 <u>rhBMP-2 in PLF</u> • 1 undefined study (12 months); N = 15 • 1 undefined study (17 months); N = 25 <u>rhBMP-2 in Cervical Fusion</u> • 1 undefined study (6 months); N = 33, f/u 6 months	LOE 1: 1 RCT LOE 2: 4 RCTs LOE 2g: 2 RCTs LOE 4b: 1 Case series	rhBMP devices for cervical fusion were not approved for use in Canada at the time the HTA was written.	The evidence reviewed indicates that the InFUSE device appears safe for lumbar fusion. Compared to autologous bone graft, radiologic fusion occurs at a faster rate among rhBMP recipients; however, clinical outcomes are similar with the exception of increased pain at the donor site for autologous bone graft patients. There is little evidence indicating that rhBMP-2 is superior to autologous bone graft in terms of cost effectiveness.

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence Base Available	Critical Appraisal*	Comments	Primary conclusions
Washington Department of Labor and Industries: Office of the Medical Director (2003)[75] *Bone Morphogenetic Protein for the Treatment of Long Bone Fractures and for Use In Spinal Fusion Procedures*	NR	rhBMP-2 ALIF and PLF rhBMP-7 PLF	% f/u NR for all studies, f/u period NR unless specified rhBMP-2/ALIF • 1 RCT (24 months); N = 279 rhBMP-2/PLF • 1 RCT (1.5, 3, 6, 12, and 24 months); N = 27 rhBMP-7/PLF • 1 RCT (12 months); N = 20	NR	HTA also included analyses of the use of rhBMP for tibial and femoral non-unions.	Similar outcomes were experienced by rhBMP and autograft patients; however, rhBMP patients do not experience the donor site pain experienced by autograft patients.

NR: not reported; RCT: randomized controlled trial; LOE: level of evidence; f/u: follow-up; HTA: health technology assessment; ; DDD: degenerative disc disease; PLF: posterolateral lumbar fusion; PLIF: posterior interbody fusion; ALIF: anterior lumbar interbody fusion; ACDF: anterior cervical discectomy fusion
* Critical appraisal refers to formal evaluation of individual study quality using criteria such as the Jadad or GRADE methods of scoring and the determination of overall strength of evidence. Authors assessed the study quality using a method developed by the U.S. Preventive Services Task Force (USPSTF) and the overall body of evidence using a modified GRADE method [39]. Level of evidence criteria in Medical Advisory Secretariat format (Level 1: Large randomized controlled trial, systematic reviews of RCTs; Level 2: Small randomized controlled trial; Level 2g: Small randomized controlled trial unpublished but reported to an international scientific meeting; Level 4b: Case series (multi-site)) [78]. Authors assessed studies' methodological quality using the York Centre for Reviews and Dissemination (CRD) criteria; RCTs were evaluated for: randomization method, allocation concealment, blinding of outcome assessors, similar prognostic baselines, well-defined inclusion/exclusion criteria, ITT (intention to treat) analysis, and number of patients lost to follow-up; non-RCTs were evaluated for: explicit population definition, similar prognostic baselines, assessment of outcomes and number of patients lost to follow-up[42].

Table 4. Overview of previous systematic reviews.

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence base available*	Critical appraisal†	Comments	Primary Conclusions
Carragee, (2011)[47] *A critical review of recombinant human bone morphogenetic protein-2 trials in spinal surgery: emerging safety concerns and lessons learned*	1995-2010 for MEDLINE, government, and admin-istrative databases	rhBMP-2 ALIF, PLF, PLIF, and ACDF	10 studies reported in 13 articles (% f/u, f/u period NR) total N = 780	Level of evidence rated for complications, morbidity, and mortality associated with rhBMP-2 in FDA data and subsequent publications. Criteria used not reported.	Intent of the SR is to compare the conclusions regarding safety and related efficacy published in industry-sponsored trials with information and data available from FDA publications.	No rhBMP-2-associated adverse events were reported in any industry-sponsored studies. A review of FDA publications revealed unpublished adverse events and internal inconsistencies. An estimate of 10% - 50% adverse events associated with rhBMP-2 use in spinal fusion is suggested, depending upon the surgical approach. Estimates of rhBMP-2 safety in the industry-sponsored publications is underestimated; either the studies were too small to adequately assess safety or the adverse events were not reported. In the industry-sponsored studies, conflicts of interest were either not reported or were unclear. The original estimate of ICBG harvesting morbidity was based on invalid assumptions and faulty methodology resulting in inflated benefit and underestimated morbidity. The control group methodology (for PLIF and PLF procedures) was potentially handicapped by significant design bias against the controls.

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence base available*	Critical appraisal†	Comments	Primary Conclusions
Mroz (2010)[54] *Complications related to Osteobiologics Use in Spine Surgery: A Systematic Review*	1990 – June 2009	ALIF, PLIF, TLIF, PLF	% f/u, f/u period, N NR for all studies rhBMP-2 17 studies rhBMP-7 5 studies 1 large registry study did not differentiate between rhBMP-2 and rhBMP-7	The strength of evidence for types of complications: high for lumbar spine, low for cervical spine, very low for thoracic spine. The strength of the evidence for rates of complications: moderate for lumbar spine, low for cervical spine, very low for thoracic spine.	This SR also investigated the possibility of a dose response relationship associated with complications following the use of rhBMP.	rhBMP-2 is not recommended for ventral cervical spine surgery until well-designed and well-executed studies can demonstrate its clinical efficacy. There are also concerns about the use of rhBMP-2 in PLIF due to potential complications. There is insufficient data to make a recommendation for either cervical or thoracic fusion.
Ryken (2009)[79] *Techniques for cervical interbody grafting*	1966 – 2007	ACDF or anterior cervical vertebrectomy and fusion	% f/u NR for all studies, f/u period NR unless specified • 1 systematic review (4 studies); N = 218 • 1 RCT (3, 6, 12, and 24 months); N = 33 • 1 prospective study (≥ 3 months); N = 20 • 1 retrospective review; N = 24 • 1 retrospective review; N = 151 • 1 retrospective review; N = 234	Overall Class II quality of evidence, "C" strength of recommendation based on: 1 LOE II/III study 2 LOE II studies 3 LOE III studies	This guideline also addresses other cervical interbody graft techniques.	Current evidence does not support the routine use of rhBMP-2 for cervical fusion. Although rhBMP-2 does promote fusion, rates of complications high compared to a standard approach.

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence base available*	Critical appraisal†	Comments	Primary Conclusions
Agarwal (2009)[68] *Osteoinductive bone graft substitutes for lumbar fusion: a systematic review*	1950 – April 2009	rhBMP-2/On-label ALIF rhBMP-2/Off-label PLF, PLIF rhBMP-7 PLF	% f/u NR for all studies rhBMP-2/On-label • 1 RCT (24 months); N = 14 • 1 RCT (24 months); N = 279 • 1 RCT (24 months); N = 131 • 1 retrospective controlled study (24 months); N = 36 • 1 prospective controlled study (24 months); N = 75 rhBMP-2/off-label • 1 RCT (24 months); N = 27 • 1 RCT (24 months); N = 98 • 1 RCT (24 months); N = 67 • 1 prospective controlled study (24 months); N = 52 rhBMP-7 • 1 RCT (12 months); N = 20 • 1 RCT (12 months); N = 20 • 1 RCT (48 months); N = 36	rhBMP-2/On-Label • 3 RCTs: study quality score 2 • Retrospective controlled study and prospective controlled study: study quality score NR rhBMP-2/Off-label • 1 RCT (N = 98): study quality score 1 • 2 RCTs: study quality score 2 •Prospective controlled study: study quality score NR rhBMP-7 • 2 RCTs: study quality score 2 • 1 RCT (N = 36): study quality score 3	SR compared the efficacy/ safety of osteoinductive bone graft substitutes with autografts and allografts for lumbar fusion.	rhBMP-2 is effective in reducing radiographic nonunion compared with AIBG; however, there is no difference in radiographic nonunion when rhBMP-7 is compared with AIBG. Neither rhBMP-2 nor rhBMP-7 demonstrated significant improvement in ODI score, possibly due to the studies being underpowered.

Assessment (year)	Lit search dates	Procedure(s) evaluated	Evidence base available*	Critical appraisal†	Comments	Primary Conclusions
New Zealand Accident Compensation Corporation (2006)[76] *Evidence based review: Bone Morphogenetic Proteins- 2 and 7*	October 2003 – March 2006	<u>rhBMP-2</u> ALIF, PLF, anterior cervical discectomy and interbody fusion, anterior vertebrectomy <u>rhBMP-7</u> Non-instrumented PLF, PLF	N represents total N for combined study categories and meta-analysis <u>rhBMP-2 in lumbar fusion</u> • 3 RCTs (12 – 24 months); N = 272 • 1 meta-analysis of 3 trials (24 months); N = 679 • 6 comparative studies (NR or 17 – 24 months); N = 452 • 6 comparative studies (NR or 17 – 24 months); N = 452 • 5 non-comparative/case series/prospective case studies (NR or 6 – 12 months); N = 402 <u>rhBMP-2 in cervical fusion</u> • 1 RCT (24 months); N = 33 • 3 prospective or retrospective non-comparative studies: (NR or 3 – 13 months): N = 195 <u>rhBMP-7 in lumbar fusion</u> • 3 RCTs (12 – 36 months); N = 90 • 2 non-comparative studies (12 months); N = 17	<u>rhBMP-2 in lumbar fusion</u> • 3 RCTs: LOE 1- to 1+ • 1 meta-analysis: LOE 1- • 6 comparative studies: LOE 3 to 1+ • 5 non-comparative/case series/prospective case studies: LOE 3 <u>rhBMP-2 in cervical fusion</u> • 1 RCT: LOE 1- • 3 prospective or retrospective non-comparative studies: LOE 3 <u>rhBMP-7 in lumbar fusion</u> • 3 RCTs: LOE 1- to 1+ • 2 non-comparative studies: LOE 3 <u>rhBMP-7 in cervical fusion</u> • 2 non-comparative studies: LOE 3	At the time of writing, rhBMP-2 and rhBMP-7 were not approved for use in New Zealand	No evidence to support claims that rhBMP is superior to autograft in achieving bone fusion or improving quality of life. However, the use of rhBMP eliminates pain/complications associated with harvesting bone for autografts. Evidence that rhBMP-2 is better at achieving long-term pain is not strong compared to bone autograft. Evidence supporting the use of rhBMP-7 for cervical/thoracic fusion seems scant and unconvincing.

NR: not reported; RCT: randomized controlled trial; LOE: level of evidence; f/u: follow-up; SR: systematic review; PLF: posterolateral lumbar fusion; PLIF: posterior interbody fusion; ALIF: anterior lumbar interbody fusion; ACDF: anterior cervical discectomy fusion; TLIF: transforaminal lumbar interbody fusion; AIBG: autologous iliac crest bone graft; ICBG: iliac crest bone graft; FDA: U.S. Food and Drug Administration
*Two of the three RCTs using rhBMP-7 for lumbar fusion possibly comprised the same patients [76].
†Critical appraisal refers to formal evaluation of individual study quality using criteria such as the Jadad or GRADE
methods of scoring and the determination of overall strength of evidence. Level of evidence was assessed using the GRADE method [54]. The quality of the studies is based on a three-class system as described in Matz[80]: Class I (well-designed RCTs), Class II (RCTs with design problems or well-designed cohort studies), and Class III (case series or poorly designed cohort studies); the RCT was assigned two LOE grades: Class II for fusion assessment, Class III for other outcomes assessments due to non-blinding; the strength of the recommendation used the SIGN (Scottish Intercollegiate Guidelines Network) system where a "C" recommendation is defined as a body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results, or extrapolated evidence from studies rated as 2++ [79-81]. Critical appraisal of clinical trials used the Jadad scale, with scores ranging from 0 (very poor) to 5 (rigorous) [68]. Methodological quality of studies assessed using the SIGN (Scottish Intercollegiate Guidelines Network) system: LOE 1+ (well conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias), LOE 1- (meta-analyses, systematic reviews of RCTs or RCTs with a high risk of bias), LOE 3 (non-analytic studies) [76, 81].

2.11. Medicare and Representative Private Insurer Coverage Policies

There are currently no coverage policies for bone morphogenetic protein published from the Centers for Medicare and Medicaid Services (CMS). Coverage policies from selected bell-weather payers are somewhat consistent for coverage of these procedures. In general, coverage policies are similar for rhBMP-7, for patients in whom an autograft is unfeasible; one payer policy does not address coverage for rhBMP-7. There is some variation among payers regarding the use of rhBMP-2, with some payers covering multiple levels, off-label use, or a different area of the lower spine. Table XX provides an overview of policy decisions.

- **Medicare**

  Medicare does not have a National Coverage Determination (NCD) for bone morphogenetic protein. Local Coverage Determinations (LCDs) do not exist at this time.

- **AETNA**[82]

  rhBMP-7

  AETNA considers the use of rhBMP-7 (OP-1) as medically necessary as an alternative to autograft in spinal fusion, where the use of an autograft is unfeasible for any of the following reasons:

  - The member has received a previous autograft and is not a candidate for further autograft procedures because the tissue is not longer available,
  - There is insufficient autogenous tissue for the intended purpose, or
  - The member is deemed an unacceptable candidate for any of the following reasons: age over 65 years; excessive risk of anatomic disruption (including fracture) from harvesting autograft from donor site; member's bone is of poor quality (osteoporosis); obesity; or presence of morbidity (infection, or fracture) preventing harvesting at autograft donor site.

  rhBMP-7 has no proven value in persons with any of the following contraindications:

  - A history of malignancy,
  - Known hypersensitivity to rhBMP-7 or to collagen,
  - Persons who are skeletally immature (less than 18 years of age or no radiographic evidence of closure of epiphyses), or
  - Pregnant women.

  rhBMP-2

AETNA considers the InFUSE Bone Graft (rhBMP-2)/LT-CAGE Lumbar Tapered Fusion Device as medically necessary for spinal fusion procedures in skeletally mature patients with degenerative disc disease (DDD) only for a single level from L4 – S1, in persons who meet all of the following criteria:

- rhBMP-2/device is to be implanted via an anterior approach;
- Member does not have greater than Grade I spondylolisthesis at the involved level,
- Member has DDD confirmed by patient history and radiographic studies,
- Member has had at least 6 months of non-operative treatment prior to treatment with the rhBMP-2/device, and
- Use of autograft or cadaveric allograft is unfeasible for any of the reasons listed for rhBMP-7 above.

rhBMP-2 is considered to be experimental and investigational for all other indications, including use in multiple levels.

- **CIGNA**[83]

Either rhBMP-2 or rhBMP-7 can be used alone or in combination with autografts, allografts, or ceramic/polymer-based synthetic bone graft substitutes.

rhBMP-7

CIGNA considers rhBMP-7 (OP-1) medically necessary when provided in accordance with the FDA Humanitarian Device Exemption (HDE) specifications in the revision of posterolateral lumbar fusion surgery when an autograft is not feasible in patients who have a failed previous spinal fusion surgery and are not candidates for an autograft due to conditions including osteoporosis, diabetes, or smoking.

rhBMP-7 is contraindicated in patients with:

- an allergy to OP-1 or collagen
- a history of malignancy, existing tumor, or previous history of cancer
- skeletal immaturity
- pregnancy.

rhBMP-2

CIGNA considers rhBMP-2 medically necessary in combination with a fusion device for single-level anterior interbody lumbar fusion surgery.

The use of rhBMP-2 has been approved by the FDA for skeletally mature patients with DDD only for a single level from L4 – S1, no more than Grade 1 spondylolisthesis at involved level, and who has had a failure of at least 6 months of nonoperative therapy.

rhBMP-2 is contraindicated in the following conditions:

- hypersensitivity to rhBMP-2, bovine Type 1 collagen, or other components of the formulation
- active infection or resected or extant tumor at the operative site
- allergy to titanium or titanium alloy
- possible or confirmed pregnancy.

o Health Net [84]

rhBMP (rhBMP-2 and rhBMP-7) is considered investigational and therefore not medically necessary for indications other than those listed below, including multiple levels of spinal fusion or thoracic or cervical fusion.

rhBMP-7

Health Net considers OP-1 (rhBMP-7) medically necessary as an alternative to autologous bone graft in "compromised" patients (diabetes, smokers, and osteoporosis) undergoing revision of a prior spinal fusion and should be used cautiously in patients with the following conditions:

- Autoimmune diseases,
- Receiving chemotherapy, radiation, immunosuppressive or steroid therapy, or
- Plan to become pregnant within one year of the procedure or are nursing.

rhBMP-2

Health Net considers InFUSE Bone Graft (rhBMP-2) medically necessary in conjunction with an LT-cage lumbar tapered fusion for spinal fusion procedures in skeletally mature patients with DDD at one level from L4-S1 after a failure of at least 6 months of non-operative therapy that could include physical therapy, pain management, spinal braces and spinal injections.

rhBMP-2 and rhBMP-7 are investigational and therefore not medically necessary in persons with any of the following contraindications:

- Known hypersensitivity to bovine Type 1 collagen, titanium or any of the material contained in the devices,
- Skeletal immaturity (under 18 years of age or no radiographic evidence of closure of epiphyses),
- Active infection at the operative site,
- Pregnancy, or
- Used in the vicinity of a resected or extant tumor, in patients with any active malignancy or patients undergoing treatment for a malignancy.

- **Premera Blue Cross (Washington and Alaska)**[85]

rhBMP-7

Premera considers rhBMP-7 (OP-1) medically necessary as an alternative to autograft in compromised patients (osteoporosis, tobacco use, or diabetes) requiring non-instrumented revision posterolateral intertransverse lumbar spinal fusion, for whom autologous bone and bone marrow harvest are not feasible or are not expected to promote fusion (FDA approved under HDE).

rhBMP-2

Premera considers rhBMP-2 (InFUSE) medically necessary for anterior spinal interbody fusion procedures at one or more levels or for instrumented posterolateral intertransverse spinal fusion procedure at one or more levels, in conjunction with an FDA-approved interbody fusion device, at one or more levels in skeletally mature patients with DDD from L2-S1. Patients should have failed at least six months of conservative treatment.

rhBMP-7 and rhBMP-2 are considered investigational for all other indications, including:

- posterolateral interbody spinal fusion,
- posterior or transforaminal interbody fusion, or
- as initial treatment or revision of non-instrumented posterolateral intertransverse spinal fusion not meeting the criteria listed above.

rhBMP-7 and rhBMP-2 are contraindicated in patients who:

- are pregnant,
- might be allergic to any of the materials in the devices,
- have an infection near the area of surgical incision,
- currently have/have had a tumor removed from the area of implantation, or
- are skeletally immature.

- **Presbyterian Healthcare Services** [86]

rhBMP-7

Presbyterian Healthcare Services policy does not address the use of rhBMP-7.

rhBMP-2

Presbyterian covers the use of InFUSE Bone Graft (rhBMP-2)/LT-Cage Lumbar Tapered Fusion Device in treating vertebra in the lower spine when all of the following indications are met:

- Skeletally mature patients with DDD at one level from L4 to S1,
- Patients does not have greater than Grade 1 spondylolisthesis at the involved level,
- Patients have had at least six months of nonoperative treatment prior to treatment with rhBMP-2, and
- rhBMP-2/device is to be implanted via an anterior open or an anterior laparoscopic approach.

The following are contraindications to rhBMP-2/device:

- The device should not be used in patients with a known hypersensitivity to recombinant human Bone Morphogenetic Protein-2, bovine Type 1 collagen or to other components of the formulations,
- The device should not be used in the vicinity of a resected or extant tumor,
- The device should not be used in patients who are skeletally immature (<18 years of age or no radiographic evidence of epiphyseal closure),
- The device should not be used in pregnant or nursing women. Women of childbearing potential should be advised to not become pregnant for one year following implantation of the device,
- The device should not be implanted in patients with an active infection at the operative site, and
- The device should not be implanted in patients with an allergy to titanium or titanium alloy.

Table 5. Overview of payer policies.

Payer (year)	Lit search dates	Evidence base available	Policy	Rationale/comments
Centers for Medicare and Medicaid Services (CMS)	N/A	N/A	No NCDs or LCDs for the region that includes Washington State	N/A
Aetna (2011)[82] *Clinical Policy Bulletin: Bone and Tendon Graft Substitutes and Adjuncts Number 0411, last review 05/27/2011*	NR	40 references provide the basis of this policy, with descriptions of the following spine-related sources: • 2 RCTs • 1 Summary statement • 1 SR • 1 ER • 1 Undefined study • 1 Technology Assessment	<u>rhBMP-7</u> Osteogenic protein-1 (OP-1) implant (rhBMP-7) is medically necessary as an alternative to autograft in spinal fusion where the use of an autograft is considered to be unfeasible for any of the following reasons: • Member has received a previous autograft and is not a candidate for further autograft procedures, or • There is insufficient autogenous tissue for the intended purpose, or • Member is deemed an unacceptable candidate for specified reasons. rhBMP-7 is considered to be experimental and investigational for all other indications. rhBMP-7 has no proven value in persons with any of the following contraindications: • A history of malignancy, • Known hypersensitivity to rhBMP-7 or to collagen, • Persons who are skeletally immature, or • Pregnant women. <u>rhBMP-2</u> rhBMP-2 is medically necessary for spinal fusion procedures in skeletally mature patients with DDD only for a single level from L4 – S1, in persons who meet all of the following criteria: • rhBMP-2/device is to be implanted via an anterior approach; • Member does not have greater than Grade I spondylolisthesis at the involved level, • Member has DDD confirmed by patient history and radiographic studies, • Member has had ≥ 6 months of non-operative	There were conflicting results from two RCTs and a Summary Statement in the use of rhBMP-7 for posterolateral fusion. One RCT found that rhBMP-7 did not perform better than autografts in bony fusion, another RCT found rhBMP-7 a safe and effective alternative to autografts for degenerative spondylolisthesis and symptomatic spinal stenosis, and the Summary Statement could not conclude that rhBMP-7 ensured fusion. One SR found that the use of an rhBMP results in slightly greater, yet statistically significant, efficacy than traditional techniques. A Technology Assessment found that the only use of rhBMP-2 to meet its criteria was a one-level anterior lumbar interbody spinal fusion for symptomatic one-level DDD of L4-S1. One ER found that although fusion occurred at a faster rate for patients receiving rhBMP-2, clinical outcomes did not differ among treatment groups. ICD-9 codes for rhBMP-7 if selection criteria met: 732.6 – 732.9. ICD-9 codes for rhBMP-2 if selection criteria met: 722.51, 722.52, 756.11, 756.12.

Payer (year)	Lit search dates	Evidence base available	Policy	Rationale/comments
			treatment prior to treatment with the rhBMP-2/device, and • Use of autograft or cadaveric allograft is unfeasible for any of the reasons listed for rhBMP-7 above. rhBMP-2 is considered to be experimental and investigational for all other indications, including use in multiple levels.	
Cigna (2011)[83] *Bone Graft Substitutes for Use in Bone Repair Number 0118, last review 1/15/2008*	NR	• Unspecified number of case series, RCTs, literature reviews • 4 HTAs • 1 ER	<u>rhBMP-7</u> rhBMP-7 (OP-1) is medically necessary when used per FDA Humanitarian Device Exemption (HDE) specifications in the revision of posterolateral lumbar fusion surgery when an autograft is not feasible in patients who have a failed previous spinal fusion surgery and are not candidates for an autograft due to conditions including osteoporosis, diabetes, or smoking. rhBMP-7 is contraindicated in patients with: • an allergy to OP-1 or collagen, • a history of malignancy, existing tumor, or previous history of cancer, • skeletal immaturity, or • pregnancy. <u>rhBMP-2</u> rhBMP-2 is medically necessary in combination with a fusion device for single-level anterior interbody lumbar fusion surgery. Use of rhBMP-2 has been approved by the FDA for skeletally mature patients with DDD only for a single level from L4 – S1, no more than Grade 1 spondylolisthesis at involved level, with a failure of ≥ 6 months of nonoperative therapy. rhBMP-2 is contraindicated in the following conditions: • hypersensitivity to any components of the formulation, • active infection or resected or extant tumor at the operative site, • allergy to titanium or titanium alloy, and • possible or confirmed pregnancy.	Evidence in published peer-reviewed literature indicates that rhBMPs, when used in an FDA-approved manner, are at least as effective as autogenous ICBG in achieving spinal fusion. Evidence from four HTAs and one ER was summarized in this policy: • On-label use: one HTA found moderate support for clinical benefit of rhBMP-2 in lumbar-sacral spinal fusion, but insufficient evidence to evaluate use of rhBMP-7 for spinal fusion. Another HTA found that rhBMP-2 meets its criteria when patients who meet FDA use criteria. A third HTA found that rhBMP-2 is more effective than autografts in fusion for single-level DDD. • Off-label use: one HTA found moderate evidence that rhBMP-2 improved fusion success, moderate evidence that rhBMP-2 use in anterior cervical spinal fusion results in increased rate of complications, and insufficient evidence to draw any conclusions regarding the off-label use of rhBMP-7 in lumbar fusion. • Another HTA found that patients who received rhBMP in spinal fusion had similar clinical outcomes when compared with patients receiving an autograft. One ER found that although fusion occurred at a faster rate for patients receiving rhBMP-2, clinical outcomes did not differ among treatment groups. ICD-9 codes for rhBMP if selection criteria met: 715.90 – 715.98, 722.52, 722.73, 722.83, 724.9, 806.4

Payer (year)	Lit search dates	Evidence base available	Policy	Rationale/comments
			Either rhBMP-2 or rhBMP-7 can be used alone or in combination with autografts, allografts, or ceramic/polymer-based synthetic bone graft substitutes. rhBMP is not covered for any other indication, including the following indications considered experimental, investigational, or unproven: • when used for spinal fusion procedures other than single-level anterior spinal fusion (rhBMP-2), and • for the treatment of cervical spine conditions (rhBMP-2 or rhBMP-7)	
Health Net (2011) [84] *National Medical Policy: Bone Morphogenetic Protein Number NMP243, last review 02/2011*	NR	68 total references form the basis of this policy, including descriptions of: • 2 RCTs • 1 Retrospective review • 2 Undefined	<u>rhBMP-7</u> OP-1 (rhBMP-7) is medically necessary as an alternative to autologous bone graft in compromised patients undergoing revision of a prior spinal fusion and should be used cautiously in patients: • With autoimmune diseases, • Receiving chemotherapy, radiation, immunosuppressive or steroid therapy, and • Who plan to become pregnant within one year of the procedure or are nursing. <u>rhBMP-2</u> InFUSE Bone Graft (rhBMP-2) is medically necessary in conjunction with an LT-cage lumbar tapered fusion for spinal fusion procedures in skeletally mature patients with DDD at one level from L4-S1 after a failure of ≥ 6 months of non-operative therapy. rhBMP-2 and rhBMP-7 are investigational and not medically necessary in persons with any of the following contraindications: • Known hypersensitivity to any of the material contained in the devices, • Skeletal immaturity, • Active infection at the operative site, • Pregnancy, • Used in the vicinity of a resected or extant tumor or in patients with any active malignancy.	• 2 RCTs demonstrated that patients receiving rhBMP-2 experience shorter hospital stays and less blood loss than ICBG patients (one RCT included one-level open anterior lumbar fusion, the other RCTs type of fusion NR). One of these RCTs also reported a higher fusion rate among rhBMP-2 patients. • One observational study found that the effectiveness of rhBMP-2 was not dependent upon the number of levels treated or the surgical approach in TLIF surgeries. This study also reported no complications or bone overgrowth. • One retrospective review found that high doses of rhBMP-2 in anterior cervical fusion resulted in a significant rate of complications in patients. ICD-9 codes for rhBMP if selection criteria met: 715.90 – 715.99, 724.9.

Payer (year)	Lit search dates	Evidence base available	Policy	Rationale/comments
			rhBMP is considered investigational and not medically necessary for other indications, including multiple levels of spinal fusion or thoracic or cervical fusion.	
Premera, Blue Cross (2010) [85] *Bone Morphogenetic Protein, Number 7.01.100, last review 10/12/10*	2004 – June 2010	38 total references provide the basis for this policy, with descriptions of: • 1 Case control • 1 Case series • 2 Meta-analysis • 1IDE trial • 1 Investigational protocol • 7 RCTs • 2 Retrospective review • 2 SR • 4 Undefined	<u>rhBMP-7</u> rhBMP-7 (OP-1) may be considered medically necessary as an alternative to autograft in compromised patients requiring non-instrumented revision posterolateral intertransverse lumbar spinal fusion, for whom autologous bone and bone marrow harvest are not feasible or are not expected to promote fusion (FDA approved under HDE). <u>rhBMP-2</u> rhBMP-2 (InFUSE) is considered medically necessary for anterior spinal interbody fusion procedures or for instrumented posterolateral intertransverse spinal fusion procedure, in conjunction with an FDA-approved interbody fusion device, at ≥ 1 levels in skeletally mature patients with DDD from L2-S1. Patients should have failed at least six months of conservative treatment. rhBMP-7 and rhBMP-2 are considered investigational for all other indications, including: • posterolateral interbody spinal fusion; • posterior or transforaminal interbody fusion; • initial treatment or revision of non-instrumented posterolateral intertransverse spinal fusion not meeting the criteria listed above. rhBMP-7 and rhBMP-2 are contraindicated in patients who: • are pregnant, • might be allergic to any of the materials in the devices, • have an infection near the area of surgical incision, • currently have/have had a tumor removed from the area of implantation, • are skeletally immature.	The majority of trials were designed to assess the equivalence of rhBMP and autograft therapies, not the superiority of any one treatment. <u>rhBMP-7</u> • Evidence from several RCTs and a meta-analysis indicates there is no difference between rhBMP-7 and autograft in lumbar fusion or instrumented posterolateral intertransverse lumbar fusion regarding ODI, fusion, or overall success. <u>rhBMP-2</u> <u>Lumbar:</u> • Regarding ALIF/PLF/PLIF: a meta-analysis found a lower risk of non-union and ODI scores not significantly improved in rhBMP-2 patients. An SR reported that potential complications raised concerns regarding its routine use for PLIF. Three RCTs reported mixed results: one study reported no difference in fusion success, ODI, or back pain, the other two studies reported higher fusion rates and improved pain with rhBMP-2. • Regarding PLIF/TLIF: One RCT reported high radiographic success and ODI score with rhBMP-2. A retrospective review and other studies reported varying rates of ectopic bone formation with rare symptomatic neurologic involvement. • Regarding instrumented posterolateral intertransverse lumbar fusion: Several comparative studies reported no difference in pain, ODI or SF-36 scores, and higher fusion rates with rhBMP-2, although this study had a high LTF. One prospective study found high rate of fusion with off- and on-label use of rhBMP-2. One RCT reported no difference in ODI, SF-26, or pain scores, and higher fusion

Payer (year)	Lit search dates	Evidence base available	Policy	Rationale/comments
				and lower second surgery rates with high doses of rhBMP-2. Cervical • Although several comparative studies showed successful fusion or greater endplate resorption and subsidence in rhBMP-2 patients, one of these studies also showed high levels of adverse effects, such as swelling and dysphagia. rhBMP-2 is not recommended for use in cervical spine. ICD-9 codes for rhBMP if selection criteria is met: 722.52
Presbyterian Healthcare Services (2010) [86] *InFUSE Bone Graft (recombinant Human Bone Morphogenetic Protein-2 Number MPM 9.4, last review 07/28/10*	NR	5 total references form the basis of this policy, with no descriptions of any of the studies.	rhBMP-7 NR rhBMP-2 For treating vertebra in the lower spine all of the following indications must be met: • Skeletally mature patients with DDD at one level from L4 to S1, • Patients do not have greater than Grade 1 spondylolisthesis at the involved level, • Patients have had ≥ six months of nonoperative treatment, and • Device is to be implanted via an anterior open or an anterior laparoscopic approach. The FDA has determined the following contraindications: • Device should not be used in patients with a known hypersensitivity to components of the formulations, • Device should not be used in the vicinity of a resected or extant tumor, • Device should not be used in patients who are skeletally immature, • Device should not be used in pregnant or nursing women, and • Device should not be implanted in patients with an active infection at the operative site.	ICD-9 codes for rhBMP-2 if selection criteria is met: 715.90-715.98, 722.51, 722.52, 722.73, 722.83, 724.9, 756.11, 756.12, 806.4.

NR: not reported
LTF: Loss to Follow-up
ICBG: Iliac Crest Bone Graft
DDD: Degenerative Disc Disease
HDE: Humanitarian Device Exemption
ER: Evidence Review
SR: Systematic Review
ALIF: Anterior Lumbar Interbody Fusion
PLIF: Posterior Lumbar Interbody Fusion
PLF: Posterolateral Fusion
TLIF: Transforaminal Lumbar Interbody Fusion

3. The Evidence

3.1. Methods of the Systematic Literature Review

3.1.1. Inclusion/exclusion

The focus of this systematic review is on- and off-label use of rhBMP-2 and rhBMP-7 in the lumbar and cervical spine. Inclusion and exclusion criteria are summarized in Table 6.

Table 6. Summary of inclusion and exclusion criteria

Study Component	Inclusion	Exclusion
Participants	Patients with back and/or leg or neck pain	◆ Skeletally immature patients (< 18 years of age) ◆ Pregnancy ◆ History of tumor in the implantation site ◆ Infection at the implantation site
Intervention	◆ FDA-approved ("on-label") and -unapproved ("off-label") implantation of rhBMP-2 (InFUSE) or rhBMP-7 (OP-1) in the lumbar or cervical spine	◆ Implantation of rhBMP-2 or rhBMP-7 into sites other than the spine ◆ Spine fusion not using rhBMP-2 or rhBMP-7
Comparators	◆ Placebo ◆ Standard care ◆ Physical therapy ◆ Autograft bone, allograft bone, bone marrow, demineralized bone matrix, stem cells, and/or other bone substitutes used to enhance bone remodeling)	
Outcomes	Perioperative outcomes Short- and long- term: ◆ Functional outcomes ◆ Pain ◆ Radiographic fusion ◆ Patient satisfaction ◆ Quality of life ◆ Activities of daily living ◆ Return to work ◆ Complications/Adverse events (safety) ◆ Reoperation (safety) ◆ Prognostic factors	◆ Non-clinical outcomes
Study Design	◆ Reliability/validity studies for question 1. ◆ Comparative studies for questions 2-4. ◆ Case series and case reports designed to evaluate adverse events for question 3. ◆ Formal economic studies will be sought for question 5	◆ Non-clinical studies
Publication	◆ Studies published in English in peer reviewed journals, published HTAs or publically available FDA reports	◆ Abstracts, editorials, letters ◆ Duplicate publications of the same study which do not report on different

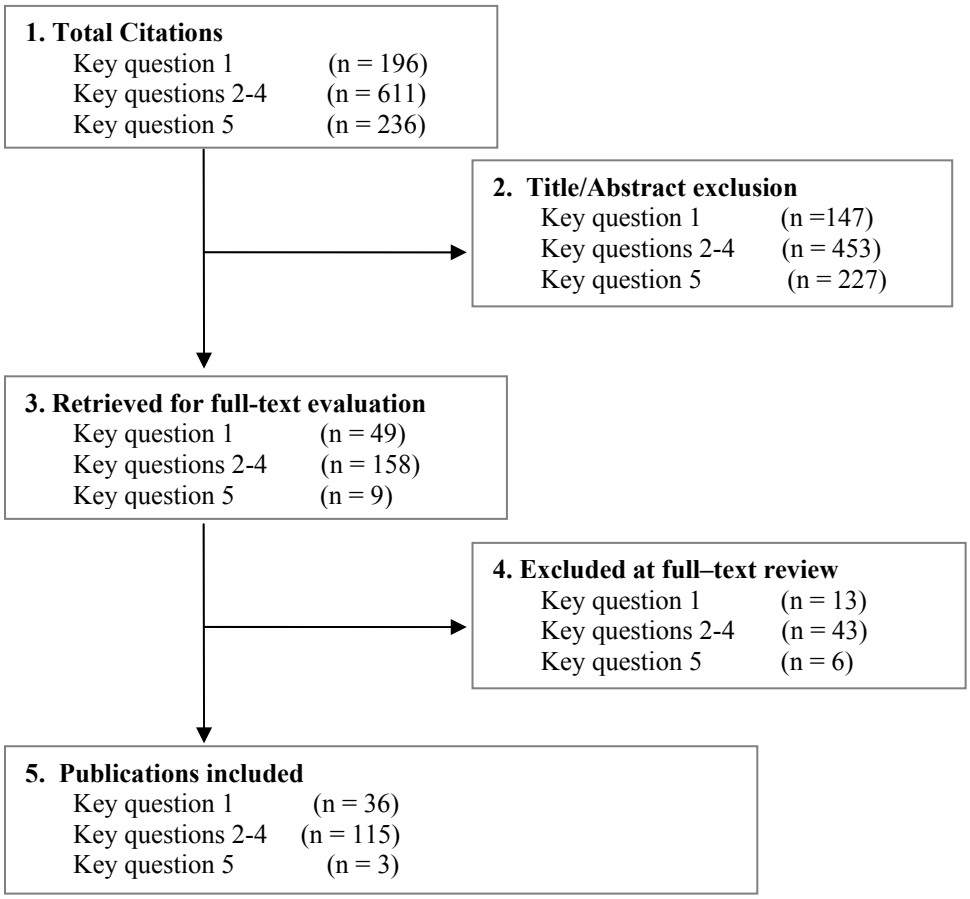
Study Component	Inclusion	Exclusion
	◆ Full formal economic analyses (e.g. cost-utility studies) published in English in HTAs or in a peer-reviewed journal published after those represented in previous HTAs.	outcomes ◆ Single reports from multicenter trials ◆ Studies reporting on the technical aspects of BMP use in fusion surgery ◆ White papers ◆ Narrative reviews ◆ Articles identified as preliminary reports when results are published in later versions ◆ Incomplete economic evaluations such as costing studies

3.1.2. Data sources and search strategy

The clinical studies included in this report were identified using the algorithm shown in Appendix A. The search took place in four stages. The first stage of the study selection process consisted of a comprehensive literature search using electronic means and hand searching. We then screened all possible relevant articles using titles and abstracts in stage two. This was done by two individuals independently. Those articles that met a set of *a priori* retrieval criteria based on the criteria above were included. Any disagreement between screeners that were unresolved resulted in the article being included for the next stage. Stage three involved retrieval of the full text articles remaining. The final stage of the study selection algorithm consisted of the selection of those studies using a set of a priori inclusion criteria, again, by two independent investigators. Those articles selected form the evidence base for this report.

Electronic databases searched included PubMed, EMBASE, CINAHL, ClinicalTrials.gov, CRISP, HSTAT, *The Cochrane Library*, EconLIT, PsychINFO, AHRQ, and INAHTA for eligible studies, including health technology assessments (HTAs), systematic reviews, primary studies and FDA reports. The databases were searched from inception through mid-September, 2011. For key question 2 and comparative studies in key question 3, the search performed in the AHRQ HTA on BMP was accepted and used[39]; this search went into February, 2010 and we performed an additional search to identify relevant studies published from January 2010 through mid-September, 2011. Reference lists of all eligible studies were also searched. The search strategies used for PubMed and EMBASE, are shown in Appendix B. Figure 1 shows a flow chart of the results of all searches for included primary studies. Articles excluded at full-text review are listed in Appendix C.

Figure 1. Flow chart showing results of literature search



3.1.3. Data extraction

Reviewers extracted the following data from the clinical studies: study population characteristics, study type, study period, patient demographics and preoperative diagnoses, study interventions, follow-up time, study outcomes (pain, patient satisfaction, global perceived effect, health-related quality of life, anxiety and depression, function, medication usage, and "success"), adverse events (reoperation, device-related complications, and other complications or side effects. An attempt was made to reconcile conflicting information among multiple reports presenting the same data. For comparative studies for key question 2, data abstraction from the recent AHRQ HTA on BMP[39] was accepted and used; thus we did not re-abstract efficacy or effectiveness data from the studies included in that report. We did re-abstract safety data, however. For economic studies, data related to sources used, economic parameters and perspectives, results, and sensitivity analyses were abstracted. Detailed abstraction tables may be found in Appendix F.

3.1.4. Study quality assessment: Level of evidence (LoE) evaluation

The method used by Spectrum Research, Inc. (SRI) for assessing the quality of evidence of individual studies as well as the overall quality of evidence incorporates aspects of the rating scheme developed by the Oxford Centre for Evidence-based Medicine[87], precepts outlined by the Grades of Recommendation Assessment, Development and Evaluation (GRADE) Working Group[88], and recommendations made by the Agency for Healthcare Research and Quality (AHRQ)[89].

Details of the Level of Evidence (LoE) grades are found in Appendix E. Each clinical/human study chosen for inclusion was given a LoE rating based on the quality criteria listed in Appendix D. Standardized abstraction guidelines were used to determine the LoE for each study included in this assessment. Reasons for LoE grading for each study are provided in Appendix E following the tables.

Details on the grading of economic studies are found in Appendix D.

Following the assessment of the quality of each individual study included in the report and of the data available to answer each key question, an overall "strength of evidence" for the relevant question is determined (see section 5.0 of the report). The method and descriptions of overall strength are adapted for diagnostic studies from a system described by the GRADE Working Group[88] for the development of clinical guidelines. Details are provided in Appendix D.

3.2. Quality of Literature available

3.2.1. Quality of studies retained

We initially found 1043 citations using the search strategy in Appendix B.

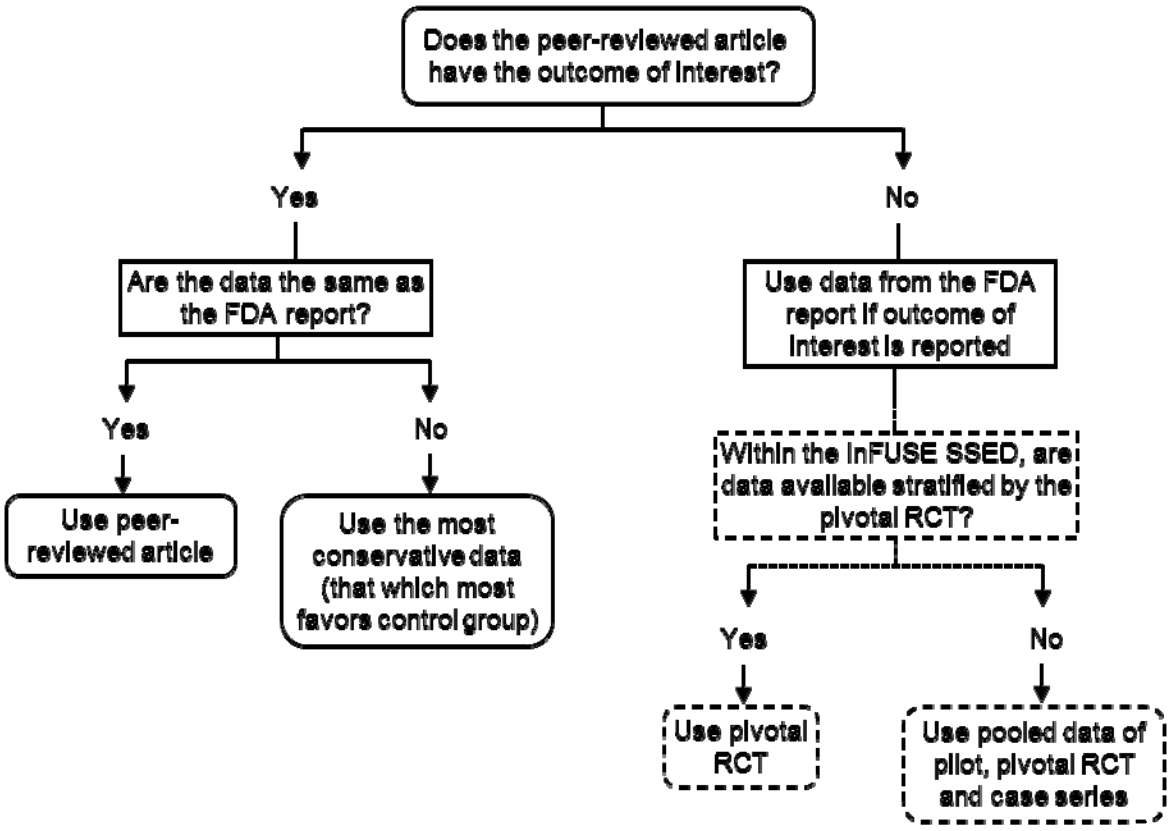
For Key Question 1 we identified a total of 36 studies.

For Key Question 2 on efficacy/effectiveness we found a total of 14 RCTs (LoE I: 0 studies; LoE IIa: 1 study; LoE IIb: 13 studies) and 15 cohort studies (LoE II: 2 studies; LoE III: 13 studies).

For Key Question 3 on safety we found 2 additional subsets of RCTs included in Key Question 2 that presented more safety data [1, 2]), 12 cohort studies (LoE II: 1 study; LoE III: 11 studies), 33 case series, and 16 case reports. All studies included in Key Question 2 were evaluated for safety in Key Question 3. Thus Key Question 3 evaluates 14 RCTs (+ 2 additional subset analyses from these studies), 27 cohort studies, 33 case series, and 16 case reports.

Note that for Key Question 3, we identified three publically available FDA reports, all three of which had peer-reviewed publications associated. In general, we used the safety data from the peer-reviewed publication when possible. If the peer-reviewed publication did not have the outcome of interest, we looked to the FDA report. If the peer-reviewed publication and the FDA report both provide data on the outcome of interest, but the data are reported differently, we used the most conservative data; that is, we used the report in which the comparison most favored the control group, Figure 2. For the InFUSE FDA SSED, we noted that most of the data presented were a compilation of three different datasets: a small pilot study, a large pivotal RCT and a single arm case series. When using the FDA InFUSE SSED, if the data were segregated by dataset, we used the pivotal RCT. If not, we then looked to the pooled data that included all three datasets.

Figure 2. Algorithm to assist in determining which data to use from overlapping peer-reviewed publications and FDA safety reports.



To address outcomes following fusion with versus without BMP in special populations (Key Question 4), we included 8 cohort studies (LoE II: 1 study; LoE III: 7 studies).

For Key Question 5, we identified 3 studies that met our inclusion criteria.

Tables summarizing the level of evidence can be found in APPENDIX E.

4. Results

4.1. Key Question 1: Expected treatment outcomes, validated instruments, and clinically meaningful improvement

What are the expected treatment outcomes of lumbar or cervical spinal fusion? Are there validated instruments related to outcomes in patients undergoing these procedures? Has clinically meaningful improvement in outcomes been defined in these patient populations?

Summary: We identified four outcome measures commonly used in the comparative studies in this HTA. Only one, the SF-36, was evaluated for validity in spinal fusion patients. One study demonstrated criterion validity and internal consistency of the SF-36 in patients undergoing lumbar fusion[77]. Responsiveness and other aspects of validity were not examined in this study. The other three outcome measures (the ODI, the NDI, and pain assessed by a VAS) have been shown to have a degree of validity, reliability, and responsiveness in various spine populations, some of which might be eligible for fusion. The minimal clinically important difference (MCID) was defined in fusion patients for the ODI in several studies[38, 82, 90] (MCID is variously defined as 10 – 22.9, depending upon the study population and calculation method) and pain as measured by the VAS in one study[82] (MCID = 18 – 19). However, there is some cause for concern surrounding the definition of MCID for ODI, including the wide range and variability of reported values, the feasibility of the MCID being able to detect an improvement considered important to patients, and the various calculation methods of the MCID.

What are the expected treatment outcomes?

Expected treatment outcomes were identified from the comparative studies included in KQ2 efficacy/effectiveness. The most common outcomes measures used (see Figure 3) include:

- ***Oswestry Disability Index (ODI)***

  The Oswestry Disability Index (ODI) consists of 10 items measuring the level of pain interference with physical activities such as lifting, walking, sitting, sleeping, standing, sex life, traveling, social life, and personal care [90-92]. It has a maximum score of 100, with higher scores indicating greater disability.

- ***Neck Disability Index (NDI)***

  The Neck Disability Index (NDI) is a modification of the ODI, with five scales from the ODI (pain intensity, personal care, lifting, driving, and sleeping) and five new scales (reading, headaches, concentration, work, and recreation)[93]. It has a maximum score of 100, with higher scores indicating greater disability.

- ***Pain (Visual Analogue Scale - VAS)***

Pain is often measured using a Visual Analogue Scale (VAS) of a 100 mm length [94-96]. It typically a minimum score of 0 mm (no pain) and a maximum score of 100 mm (maximum amount of pain). In the comparative studies included in this report, pain was sometimes measured on a scale of 0-20 to account for pain frequency and intensity, each of which are measured on separate 0-10 scales.

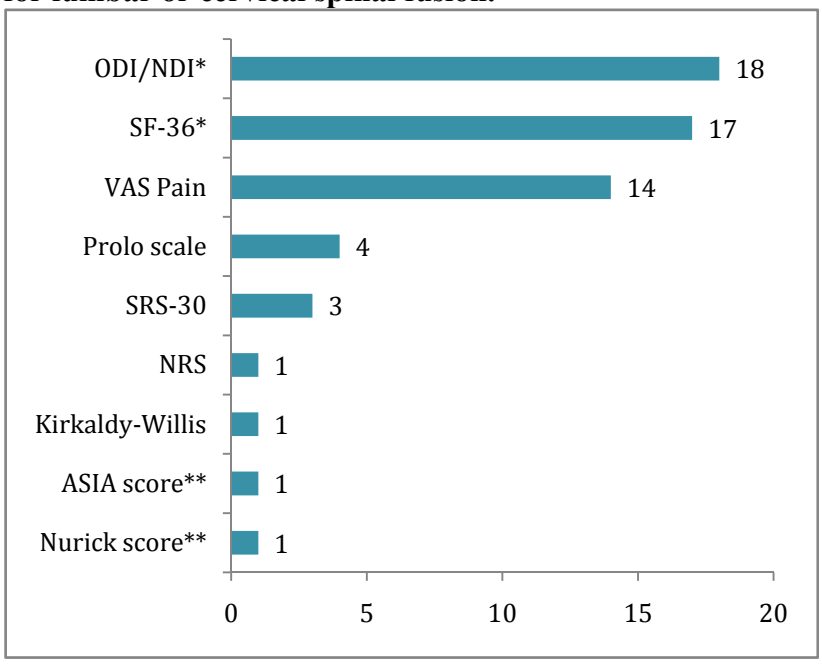
- ***Medical Outcomes Study Short-Form 36 (SF-36)***

  The Medical Outcomes Study Short-Form 36 (SF-36) is a self-administered questionnaire with 36 items assessing patient responses in eight health domains: physical functioning, mental health, emotional role limitations, social functioning, vitality, general health, physical role limitations, and bodily pain[94, 97]. Each domain or subscale has a maximum score of 100, with lower scores indicating greater disability[94]. This instrument was originally developed in a U.S. population of 2474 patients with a variety of medical problems[98].

  An additional outcome measure, spinal radiographic assessment (e.g., X-ray, CT, or computerized tomography, scan) of fusion, is a surrogate outcome in spinal fusion studies (as opposed to the "gold standard" of surgical exploration[99-101]). It is a low-cost method, has widespread availability, and a long history of use. However, it has limitations as a reliable test for determining the presence or absence of solid fusion[101]. It is not known exactly how accurate plain radiographs are in determining whether a spine is successfully fused, there is variation in the radiographic criteria used to determine whether a spine is fused or not, and no standard agreed-upon method exists to evaluate the successful fusion in the spine. In addition, it is unclear how radiographic measures correspond to clinical outcomes in patients[102, 103]. Most of the RCTs included in this HTA defined successful fusion on plain radiograph as presence of bridging trabecular bone between the transverse processes and absence of motion (≤ 3 mm of translation and < 5º angular motion on flexion-extension views)[10, 16-20, 104], with some studies adding the absence of radiolucent lines over 50% or more of the implant surfaces to the preceding definition[1, 6-9, 12, 21, 25, 26]. A few studies defined successful fusion on CT when there was presence of bridging trabecular bone[5] growing through both cages[6, 21].

Table 7 presents a description of these and other less commonly reported measures from the comparative studies included in this HTA.

Figure 3. Frequency of outcomes measures used in comparative studies utilizing rhBMP for lumbar or cervical spinal fusion.



*Includes lumbar and cervical studies
**Cervical study only
ODI indicates Oswestry Disability Scale; NDI, Neck Disability Index; SF-36, Short form-36; VAS, Visual Analogue Scale; SRS-30, Scoliosis Research Society-30; NRS, Numerical Rating System; ASIA score, American Spinal Injury Association.

Table 7. Description of outcomes instruments used in comparative studies evaluating BMP use in the spine.

Outcome measure	Clinician or patient reported	Instrument type	Components	Score range	Interpretation

Outcome measure	Clinician or patient reported	Instrument type	Components	Score range	Interpretation
ASIA scale (American Spinal Injury Association)[105]	Clinician	Spine	Sensory and motor function	A – E	A = Complete: no sensory or motor function is preserved in the sacral segments S4-S5. B = Sensory Incomplete: sensory but not motor function is preserved below the neurological level and includes the sacral segments S4-S5. C = Motor Incomplete. Motor function is preserved below the neurological level, and more than half of key muscle functions below the single neurological level of injury have a muscle grade less than 3. D = Motor Incomplete. Motor function is preserved below the neurological level, and at least half (half or more) of key muscle functions below the NLI have a muscle grade > 3. E = Normal. Motor and sensory function are normal.
			Muscle function	0 – 5	0 = Total paralysis 1 = Palpable or visible contraction 2 = Active movement, full ROM with gravity eliminated 3 = Active movement, full ROM against gravity 4 = Active movement, full ROM against gravity, moderate resistance in a muscle-specific position 5 = Normal: active movement, full ROM against gravity, full resistance in a muscle-specific position expected from an otherwise unimpaired person 5* = Normal: active movement, full ROM against gravity, sufficient resistance to be considered normal if identified inhibiting factors were not present
Kirkaldy-Willis criteria[106]	Patient	Generic	Overall clinical results	Poor to Excellent	Poor: no return to work Fair: reduction in working capacity, taking lighter job or working part-time; might have occasional recurrence of pain requiring absence from work for 1 – 2 weeks once or twice a year Good: return to normal work with possible restriction in other activities; after heavy work might have recurrent back pain requiring short rest (a few days) Excellent: return to normal work and other activities with little to no complaint

Outcome measure	Clinician or patient reported	Instrument type	Components	Score range	Interpretation
NDI (Neck Disability Index)[93, 94]	Patient	Neck	Pain intensity Personal care Lifting Reading Headaches Concentration Work Driving Sleeping Recreation	0 – 50 or 0 – 100*	Higher scores = greater disability
NRS (Numerical Rating System)[107]	Patient	Generic	Pain If composite score: 10 points for pain intensity and 10 points for pain duration/frequency	0 – 10 or 0 – 20 if composite score	No pain: 0 Mild pain: 1 – 3 Moderate pain: 4 – 6 Severe pain: 7 – 10
Nurick scale[108]	Patient	Myelopathy severity (cervical spine)	Ambulatory ability	0 - 5	Grade 0: signs or symptoms of root involvement but without evidence of spinal cord disease Grade 1: signs of spinal cord disease but no difficulty in walking Grade 2: slight difficulty in walking that did not prevent full-time employment Grade 3: difficulty in walking that prevented full-time employment or the ability to perform all housework but that was not severe enough to require someone else's help to walk Grade 4: able to walk with someone else's help or the aid of a frame Grade 5: chair bound or bedridden
ODI (Oswestry Disability Index, or Oswestry Low Back Pain Disability Questionnaire) (version 2.0) [94, 109]	Patient	Back	Pain intensity Personal care Lifting Walking Sitting Standing Sleeping Sex life Social life Travelling	0–100*	Higher scores = greater disability

Outcome measure	Clinician or patient reported	Instrument type	Components	Score range	Interpretation
Prolo Scale (modified)[110]	Patient	Back	Pain: unbearable to no pain Functional status: total incapacity to ability to do everything Economic status: unable to do tasks at home to ability to work at heavy capacity or previous job Medication: ≥ 10 hydrocodone tablets or equivalent to no or occasional hydrocodone	4 – 20 (5 points for each domain)	lower scores = greater disability
SF-36 (Short Form 36 health survey questionnaire)[94, 98]	Patient	Generic	8 subscales (# items) Physical functioning (10) Role limitations due to physical health problems (4) Bodily pain (2) General health (5) Vitality (4) Social functioning (2) Role limitations due to emotional problems (3) Mental health (5)	0–100 for each subscale (total score not used)	Lower score = greater disability
SRS-22 (Scoliosis Research Society)[94] SRS-30 (Scoliosis Research Society)[111]	Patient	Disease specific	Satisfaction with surgery Mental health Pain Function/activity Self image/appearance (22 or 30 items total)	0 – 100	Lower score = greater disability
VAS pain (Visual Analogue Scale)[94, 96]	Patient	Back	Pain and disability in 2 indices (# items) Subjective (15) Objective (10)	0–10 or 0-100 measured using 100 mm visual analog scale	Higher score = greater disability

ROM: range of motion

* ODI and NDI: Each of the ten subscales is scored on a scale of 0–5 points; the total score is then doubled for a final score ranging from 0–100 points.

Are there validated instruments related to outcomes in patients undergoing these procedures?

Of the four most common outcomes measures used in assessing efficacy and effectiveness of rhBMP, only the SF-36 was validated in patients undergoing spinal fusion. The other three common outcomes measures (the ODI, the NDI, and pain assessed via a VAS) have been shown to have some degree of validity, reliability and responsiveness in various spine populations, some of which would be eligible for fusion.

Oswestry Disability Index (ODI)

The ODI has been validated against numerous disease-specific and generic instruments, has been translated into numerous languages, including Danish, Norwegian, and Japanese, and has been validated in French, German, Greek, and Finnish[94]. The ODI has been found to be valid, reliable, and responsive for measuring changes in functional status in patients receiving treatment for spinal complaints[90, 92, 94].

A review summarized the validity, reliability, and responsiveness of the ODI as follows:[92]

- Face and content validity have been demonstrated in several studies with various analyses[112-114]. Construct validity has been demonstrated by a correlation with a pain measure (VAS), the McGill Pain Questionnaire, and the SF-36[115-118].
- Reliability has been met in several studies in the general spine population. Internal consistency has been demonstrated, with Cronbach's alpha values ranging from 0.71 to 0.87 in three studies[113, 114, 119]. Test-retest reliability has also been demonstrated, although the correlation of scores differs with the time interval between testing[91, 114, 116].
- Responsiveness was demonstrated in several studies with an acceptable ROC index in a population of less severely affected patients[112, 120]. The review reports that the ROC index has not been calculated in more severely affected patients.

Neck Disability Index (NDI)

Validity of the NDI was studied in two studies in patients with neck pain. Content validity was demonstrated in one study of 48 patients with neck pain by conducting peer-review and patient feedback sessions[93] and in another study of 40 patients with neck pain by asking the patients about the relevance of the questionnaire to their disorder[121]. The latter study also included 18 patients with no neck pain. In one of the aforementioned studies, construct validity was demonstrated with a nearly normal distribution of the frequency of the scores[93]. However, poor construct validity was demonstrated in a study of 38 patients with cervical radiculopathy[122]. The study of 48 patients with neck pain also demonstrated criterion (concurrent) validity by calculating a correlation between the changes in NDI pre- and posttreatment scores and those of an improvement in activity level (on a VAS scale) on a subset of patients[93].

Reliability and responsiveness were evaluated in two studies, one in a population of patients with neck pain[93] and one in a population with cervical radiculopathy[122]. With respect to reliability, Vernon et al. studied 48 patients with neck pain and demonstrated a high degree of test-retest reliability and internal consistency with high correlation (Pearson's $r = 0.89$, $p \leq .05$) and alpha (0.80) coefficients[93]. An intraclass correlation coefficient 0.68 (95% CI, 0.30 – 0.90) was calculated by Cleland et al. to demonstrate adequate test-retest reliability in a study of 38 patients with cervical radiculopathy[122].

With respect to responsiveness, Vernon et al. demonstrated moderate responsiveness by correlation with a self-report measure of clinical improvement[93]. In contrast, the responsiveness was not demonstrated by Cleland et al. with either an area under the ROC curve (AUC) analysis or a correlation of the NDI change scores with the global rating of change (GROC) and numerical pain rating scale (NPRS)[122].

Pain (VAS)

The VAS (visual analogue scale) for pain was found to have construct validity in a study of 230 patients with chronic, disabling musculoskeletal disorders to include the spine, by correlating the pain VAS with the Pain Disability Questionnaire (PDQ) and ODI[123]. Another study of 289 patients with chronic low back pain demonstrated criterion validity, using a patient global assessment of improvement[124].

Reliability was demonstrated with a high degree of intraobserver reproducibility (intraclass correlation coefficient = 0.966) in one study of 19 patients with chronic back pain[96].
Two studies in patients with chronic low back pain assessed responsiveness. In one study of 289 patients, responsiveness was demonstrated using an ROC curve analysis and an effect size calculation[124]. Another study of 230 patients with chronic, disabling musculoskeletal disorders also demonstrated responsiveness using an effect size calculation[123].

SF-36

One study was found that assessed the SF-36 instrument in patients undergoing lumbar fusion[115]. This study presented correlation coefficients between each domain of the SF-36 and the Oswestry Low Back Pain Disability and Low Back Outcome scores as evidence of criterion validity (*r*s ranged from 0.40 to 0.78), but the authors state that there is no evidence that these two scales represent a gold standard for low back pain disability. Internal consistency was also demonstrated with high alpha coefficients. Responsiveness and other aspects of validity were not examined in this study.

In non-fusion patients, one cervical and one lumbar study appraised the validity of the SF-36. One study of 88 patients with cervical spondylotic myelopathy (CSM) demonstrated construct

validity of the SF-36 by confirming a hypothesized relationship between selected SF-36 subscales and various myelopathy scales[97]. Another study of 620 patients with lumbar disc prolapse, lumbar canal stenosis, or cervical spondylotic radiculomyelopathy demonstrated construct, convergent, discriminate, and predictive validity by calculating correlations between the specific and generic scores[125].

Reliability was demonstrated in one study of 88 patients with CSM with an internal consistency analysis (Cronbach alpha ranging from 0.79 – 0.91 for the eight domains)[97].

Responsiveness was met in two studies in the patients with lumbar or cervical pain. A study of 970 patients with lumbar pain/leg symptoms (due to herniated disc, spinal stenosis, or spondylosis) demonstrated responsiveness using an ROC curve analysis and effect size calculations[126]. Another study of lumbar and cervical patients demonstrated moderate to good responsiveness with a standardized response mean (SRM) analysis[125].

Has clinically meaningful improvement been defined in these patient populations?
The Minimal Clinically Important Difference (MCID) is the smallest change in an outcome measure that is important to a patient[127].

Oswestry Disability Index (ODI)
No single agreed-upon MCID value for the ODI has been reported. The FDA required an ODI change of 15 points in a study comparing lumbar fusion with total disc arthroplasty[9]. Others have used values ranging from 4 to 18.4 points[92, 128-131]. Some contend that the MCID should have a context-specific value such as a 15% change rather than a fixed number[130]. Other issues surrounding the definition of MCID for ODI are cause for concern. In a study of 294 patients with severe chronic low back pain, Hagg et al.[132] calculated an MCID for ODI improvement to be 10 units, with a 95% tolerance interval (TI) of 10 units. The authors concluded that because of the imprecision of the instrument, the outcome measure might not be responsive enough to detect an improvement that is considered important to patients. Another study examined MCID in a population of 45 patients undergoing transforaminal lumbar interbody fusion (TLIF) for degenerative lumbar spondylolisthesis-associated back and leg pain[127]. This study calculated MCID values for the ODI ranging from 11 – 22.9, depending upon the calculation method.

Neck Disability Index (NDI)
No studies were found that assessed the MCID in the spinal fusion population. However, the MCID has been reported in three studies of patients with neck pain, with MCID values of 7.0, 7.5, and 19 points[122, 133, 134].

Pain (VAS)

MCID pain scores suffer from similar problems as the ODI. The range of absolute values for pain have been reported from 2 to 29 points[131, 132, 135, 136]. In a 2008 review of change scores for pain and functional status in patients with low back pain, Ostelo et al. noted that studies estimating MCID used different time intervals for test-retest, applied different external criteria to define "important", employed different statistical techniques to calculate MCID, and provided little or no theoretical or empirical justification for study design, anchor or method used in estimating MCID[137]. Nevertheless, for patients with chronic LBP, 20 mm improvement for pain has been recommended by some as the minimal clinically important difference[4].

SF-36

No studies were found that examined the MCID of the SF-36 in any spine population. An HTA reported an MCID of 5 based on a previous study; the study population was not defined[131].

4.2. Key Question 2: Efficacy and effectiveness

What is the evidence of efficacy and effectiveness of:

a) rhBMP-2 (InFUSE) for on-label lumbosacral spine fusion in patients with DDD?
b) rhBMP-7 (OP-1) for on-label revision posterolateral lumbar spine fusion in compromised (e.g., osteoporosis, smoking, diabetes) patients?
c) rhBMP-2 (InFUSE) for off-label lumbosacral spine fusion?
d) rhBMP-7 (OP-1) for off-label lumbosacral spine fusion?
e) rhBMP-2 (InFUSE) for off-label cervical spine fusion?
f) rhBMP-7 (OP-1) for off-label cervical spine fusion?

Including consideration of perioperative outcomes (including length of surgery) as well as short term and long term:

- Impact on function, pain, radiographic fusion, patient satisfaction, quality of life, activities of daily living and return to work
- Other reported measures

4.2.1. rhBMP-2: on-label use (lumbar spine)

EFFICACY

Summary

Studies:

Two LoE IIb RCTs were identified that met our inclusion criteria. Study size ranged from 14 to 279 patients. These two studies served as the pilot and pivotal trials in the 2002 FDA Summary of Strength and Effectiveness Data (SSED) for InFUSE (P000058), and both studies were sponsored by Medtronic. Patients were followed for 24 months. Because the studies were similar in design, we were able to pool outcomes data from both studies. Patients with DDD, radiculitis, and/or up to 25% spondylolisthesis and who were refractory

to conservative care underwent primary single-level open anterior lumbar fusion with either rhBMP-2/ACS (InFUSE) (n = 154) or iliac crest bone autograft (ICBG) (n = 139). RhBMP-2 was used at a dose ranging from 4.2 to 8.4 mg per patient. Additional details are available in Table 8 and the surrounding text.

Outcomes:

Perioperative outcomes (Table 9):

Operative time: The mean length of operative time was similar in both groups (1.6 versus 2.0 hours for rhBMP-2 versus ICBG) (2 RCTs). The strength of this evidence is *low*.

Blood loss: The mean perioperative blood loss was lower in the rhBMP-2 group compared with the ICBG group (108.9 versus 153.3 mL) (2 RCTs). The strength of this evidence is *low*.

Length of hospital stay: The mean length of hospital stay was similar in both groups (3.0 vs. 3.3 days for rhBMP-2 versus ICBG) (2 RCTs). The strength of this evidence is *low*.

Fusion (Table 10): The percentages of patients with successful fusion were similar in both treatment groups at all follow-ups. The strength of this evidence is *low*. By 24 months, 94.2% of rhBMP-2 and 88.5% of ICBG patients had successful fusion.

ODI (Table 11): ODI outcomes were similar between groups at all reported follow-ups (2 RCTs). The strength of this evidence is *low*. At 24 months, 84.4% of rhBMP-2 and 82.0% of ICBG patients had ODI "success", which was defined as improvement from baseline by at least 15%. Mean score improvements at 24 months were 29.6 and 23.7 points, the difference between which is not considered clinically meaningful.

Pain (Table 11): There were not clinically meaningful differences between groups in back and leg pain VAS scores as reported by one RCT (N = 279). The strength of this evidence is *low*. The percentage of patients with back pain "success", which was defined as improvement by > 3 points from baseline, was similar between groups at all follow-ups between 1.5 and 24 months. At 24 months, 75% and 79% of patients in the rhBMP-2 and ICBG groups had achieved back pain "success", while the mean VAS score improvement was 8.5 and 8.2 points, respectively. Similarly, the proportion of patients with leg pain "success" was similar between groups at 24 months (80% versus 74%, respectively. The 24-month mean improvement in leg pain VAS scores was similar at 1.5 months and identical in both groups by 24 months (6.2 points).

SF-36: function (Table 11): There was no difference in function as measured by the SF-36 between groups as reported by one RCT (N = 14). The strength of this evidence is *low*. The mean improvement in SF-36 physical function subscale scores was similar in both groups at all reported follow-ups between 3 and 24 months (38 vs. 37 for rhBMP-2 vs. ICBG, respectively at 24 months).

Patient satisfaction (Table 11): Patient satisfaction was similar in both groups at 24 months (2 RCTs). The strength of this evidence is *low*.

Work status (Table 11): There was no difference in work status between groups as reported by two RCTs. The strength of this evidence is *low*. At 24 months, the percentage of patients who were working was similar in both groups: 67.5% (104/154) compared with 56.1% (78/139) of patients in the pooled rhBMP-2 and ICBG groups, respectively.

Neurological status (Table 12): As reported by one RCT (N = 279), the percentage of patients with neurological success was similar between groups at all reported follow-ups (3 to 24 months). The strength of this evidence is *low*. At 24 months, 83% and 84% of those in the BMP and control groups had neurological success.

Detailed study characteristics:

Two randomized controlled trials[6, 21] were identified that evaluated rhBMP-2 for on-label use in the lumbar spine. Detailed abstraction tables may be found in Appendix F. The study characteristics are summarized in Table 8.

Table 8. On-label use of rhBMP-2 in the lumbar spine: RCT study overview

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Notes
Pooled data[6, 21] (2 RCTs)	InFUSE (n = 154) (4.2-8.4 mg/pt)	LT-CAGE	Anterior 1 level/pt primary	DDD, radiculitis	24 mos. 84% (246/293)	IIb	Medtronic*†	Pilot + pivotal trials for InFUSE (PMA P000058)
	ICBG (n = 139)	LT-CAGE						
Individual study details								
Boden (2000)[21]	InFUSE (n = 11) (4.2-8.4 mg/pt)	LT-CAGE	Anterior 1 level/pt primary	DDD, ≤ 25% spondylo.	24 mos. 100%	IIb	Medtronic*	Pilot trial for InFUSE (PMA P000058)
	ICBG (n = 3)	LT-CAGE						
Burkus (2002)[6]	InFUSE (n = 143) (4.2-8.4 mg/pt)	LT-CAGE	Anterior 1 level/pt primary	DDD, radiculitis	24 mos. 83%	IIb	NR† (Medtronic according to Carragee[47])	Pivotal trial for InFUSE (PMA P000058)
	ICBG (n = 136)	LT-CAGE						

DDD: degenerative disc disease; f/u: follow-up; InFUSE: rhBMP-2 applied to an absorbable collagen sponge (ACS) (Medtronic); NR: not reported
According to Carragee et al. (2011)[47], the authors had the following financial relationship with Medtronic:
* totaling $21,025,000.
† ranging from $21,121,000 - $23,581,000 plus an additional $1,500,000 for Burkus.

The RCTs by Boden (2000)[21] and Burkus (2002)[6] constitute the pilot and pivotal trials, respectively, used for FDA approval of InFUSE (Medtronic). Both studies received a level of evidence (LoE) grade of IIb. Patients in both trials underwent primary single-level anterior lumbar fusion with an LT-CAGE Tapered Lumbar Fusion Device (Medtronic) filled with either InFUSE (rhBMP-2 on an absorbable collagen sponge (ACS); Medtronic) or iliac crest bone autograft (ICBG). Because the studies were similar in design, we were able to pool outcomes data. Both studies enrolled patients with degenerative disc disease (DDD) who were unresponsive to nonoperative treatment for six months or more. Methods by which randomization was performed were not described: Boden et al., (2000)[21] randomized patients using the "marginal balancing method", which was not described; Burkus et al., (2002)[6] did not describe the method of randomization. Neither study described how treatment allocation concealment was achieved. There were differences in patient weight between groups in the Boden RCT[21] that were not controlled for (the mean weight of the control group patients was 45 pounds heavier than that of the BMP group), otherwise baseline characteristics and preoperative scores were similar between groups. Blinding of surgeons and patients was not possible due to inherent treatment differences however all primary outcomes were evaluated in a blind manner or were patient-reported. There were no obvious differences in the application of the co-interventions. Neither study had an explicit statement that data were analyzed in accordance with the intention to treat principle: the data in the Boden RCT[21] appear to have been handled this way; in the Burkus RCT[6], it appears that data from patients classified as failures (i.e., had to undergo device removals, revisions, or supplemental fixations) were not reported after they had failed the treatment. All randomized patients were followed for 24 months, and complete follow-up data were available in 84% (246/293) (range, 83-100%) of all patients. Both studies were funded by Medtronic, and according to Carragee's recent systematic review[47], the authors' financial relationship with the company totaled over $21 million.

Detailed outcomes

Perioperative outcomes (Table 9)

- Operative time (2 RCTs)
  The mean operative time was similar in both treatment groups (1.6 vs. 2.0 hours)[6, 21].

- Blood loss (2 RCTs)
  The average blood loss was lower in the BMP group compared with the ICBG group (108.9 vs. 153.3 mL)[6, 21] and again is likely attributable to the additional procedure undertaken in the control group necessary to harvest the autograft.

- Length of stay (2 RCTs)
  The mean length of hospital stay was similar in both groups (3.0 days for the InFUSE group and 3.3 days for the ICBG group)[6, 21].

Table 9. Efficacy of on-label use of rhBMP-2 in the lumbar spine: perioperative outcomes

Outcome measure	Author	Results		p-value
Operative time		**Mean (hours)**		
	Boden (2000)	**rhBMP-2 (n = 11)**	**ICBG (n = 3)**	
		1.9 ± 0.2	3.3 ± 0.6	.006
	Burkus (2002)	**rhBMP-2 (n = 143)**	**ICBG (n = 136)**	
		1.6	2.0	NR
Blood loss		**Mean ± SD (mL)**		
	Boden (2000)	**rhBMP-2 (n = 11)**	**ICBG (n = 3)**	
		95 ± 31	167 ± 111	NR
	Burkus (2002)	**rhBMP-2 (n = 143)**	**ICBG (n = 136)**	
		110	153	NR
Mean length hospital stay		**Mean ± SD (days)**		
	Boden (2000)	**rhBMP-2 (n = 11)**	**ICBG (n = 3)**	
		2.0 ± 0.6	3.3 ± 1.4	NR
	Burkus (2002)	**rhBMP-2 (n = 143)**	**ICBG (n = 136)**	
		3.1	3.3	NR

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Radiographic outcomes (Table 10)

- **Fusion** was considered successful[6, 21]:
  - On plain radiograph when:
    - there was less than 5 degrees of angular motion of flexion-extension film,
    - there was less than 3 mm translation (Burkus 2002 only), and
    - there was an absence of radiolucent lines over 50% or more of the implant surfaces.
  - On CT when:
    - there was evidence of the presence of continuous trabecular bone growing through both cages

- Successful fusion (2 RCTs):
  The percentage of patients with successful fusion was similar in both groups at all follow-ups[6, 21] (rhBMP-2 vs. ICBG):
  - 6 months: 97.4% (150/154) vs. 95.7% (133/139) (rhBMP-2 vs. ICBG)
  - 12 months: 97.4% (150/154) vs. 92.1% (128/139)
  - 24 months: 94.2% (145/154) vs. 88.5% (123/139)

- **Time to fusion** was not reported by either of the two RCTs.

Table 10. Efficacy of on-label use of rhBMP-2 in the lumbar spine: radiographic outcomes

Outcome measure	Author	Time	Results		p-value
Fusion success (see text for definition)	Boden (2000)		**% (n)**		
			rhBMP-2 (n = 11)	**ICBG (n = 3)**	
		3 mos.	91% (10)	67% (2)	NR
		6 mos.	100% (11)	67% (2)	NR
		12 mos.	100% (11)	67% (2)	NR
		24 mos.	100% (11)	67% (2)	NR
	Burkus (2002)		**rhBMP-2 (n = 143)**	**ICBG (n = 136)**	
		6 mos.	97% (139)	96% (131)	NR
		12 mos.	97% (139)	93% (126)	NR
		24 mos.	94% (134)	89% (121)	NR

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Pain (Table 11)

- ODI (2 RCTs)
  Mean score improvement: As reported by two RCTs, the mean ODI (Oswestry Disability Index) score improvement from baseline was similar in both groups at all follow-ups (rhBMP-2 vs. ICBG)[6, 21]. The ODI is discussed in more detail in Key Question 1; scores range from 0 to 100.
  - 3 months: 19.2 vs. 14.5 for rhBMP-2 (n = 154) vs. ICBG (n = 139)
  - 6 months: 24.1 vs. 20.2
  - 12 months: 27.7 vs. 25.6
  - 24 months: 29.6 vs. 23.7

  "Success": The percentage of patients with ODI "success" (defined as improvement from baseline by at least 15%) was also similar in both groups at all reported follow-ups (rhBMP-2 vs. ICBG)[6, 21].
  - 12 months: 85.7% (132/154) vs. 85.6% (119/139) (rhBMP-2 vs. ICBG)
  - 24 months: 84.4% (130/154) vs. 82.0% (114/139)

- Back pain (1 RCT)
  Mean score improvement: Back pain scores were reported by one RCT (Burkus 2002)[6]. As evaluated on a 20-point numerical rating scale (consisting of the sum of the scores for pain intensity (0-10) and pain frequency/duration (0-10)) there was no difference between groups in the mean score improvement from baseline. For back pain at 24 months, there was a mean score improvement of 8.5 points in the rhBMP-2 group versus 8.2 in the ICBG group, a difference which is not clinically meaningful. In addition, Boden et al. (2000) reported one episode of back pain that occurred at 6 monhts in the rhBMP-2 group and none in the control group (9% (1/11) versus 0% (0/3), respectively)[21].

  "Success": Similarly, there was no difference between groups in the percentage of patients = who achieved "successful" back pain relief, defined as an improvement of at

least 4 (of 20) points. At 24 months, 75% of patients in the rhBMP-2 group had successful back pain relief compared with 79% of patients in the ICBG group ($P \geq .05$)[6].

- Leg pain (1 RCT)
  Mean score improvement: Leg pain was also reported by one RCT (Burkus 2002)[6]. Patients in both groups had identical mean leg pain scores at 24 months (6.2 in both groups). Like back pain, leg pain was evaluated on a 20-point numerical rating scale (consisting of the sum of the scores for pain intensity (0-10) and pain frequency/duration (0-10)).

  "Success": There was no difference in the percentage of patients between the rhBMP-2 and ICBG groups who had leg pain "success" (79% vs. 74%, respectively; $P \geq .05$), which was defined as an improvement of 4 or more (of 20) points on the NRS scale.[6]

Function (Table 11)

- SF-36 physical function subscale (1 RCT)
  Mean score improvement: One small RCT reported no difference in the mean improvement in SF-36 physical function subscale scores from baseline between groups at 24 months follow-up (38 vs. 37 for rhBMP-2 vs. ICBG, respectively)[21]. The SF-36 is discussed in more detail in Key Question 1; scores range from 0 to 100, with lower scores indicating greater disability.

Patient satisfaction and quality of life (Table 11)

- SF-36 general health perception subscale (1 RCT)
  Mean score improvement: One small RCT reported no difference in the mean improvement of SF-36 general health perception subscale scores between groups at 24 months follow-up (5 vs. 8 for rhBMP-2 vs. ICBG, respectively)[21]. The SF-36 is discussed in more detail in Key Question 1; scores range from 0 to 100, with lower scores indicating greater disability.

- Patient satisfaction (2 RCTs)
  - There was no difference in overall patient satisfaction at 24 months as reported by Burkus (2002) (81% (116/143) vs. 80% (109/136) for rhBMP-2 vs. ICBG, respectively)[6]. When asked whether they would undergo the same procedure again, 82% of patients in the rhBMP-2 group said they would compared with 77% of patients in the ICBG group.
  - Boden et al. reported that 100% (11/11) of patients in the rhBMP-2 group rated their outcome as excellent compared with 33% (1/3) patients in the control group (33% rated outcome as good and fair, respectively (1 patient each))[21].

Social function or mental health outcomes were not reported by either of the two RCTs.

Work status (Table 11) (2 RCTs)

- At 24 months, slightly more patients in the pooled rhBMP-2 groups (67.5% (104/154) were working, compared with 56.1% (78/139) of patients in the pooled ICBG groups[6, 21].
- Burkus (2002) reported that the median number of days to return to work was similar in both treatment groups (64 vs. 65 for rhBMP-2 and ICBG, respectively; $P \geq .05$)[6].

Neurological status (Table 12) (1 RCT)

- Neurological success was evaluated by one RCT, and was defined as maintenance or improvement in each of four neurologic measurements (motor function, sensory function, deep tendon reflexes, and sciatic tension signs). There was no difference between treatment groups in the percentage of patients with neurologic success at any follow-up point[6].

Table 11. Efficacy of on-label use of rhBMP-2 in the lumbar spine: patient-reported outcomes

Outcome measure	Author	Time	Results		p-value
ODI Range: 0 – 100			**Mean score improvement (from baseline) (points)**		
	Boden (2000)		**rhBMP-2 (n = 11)**	**ICBG (n = 3)**	
		1.5 mos.	NR	NR	
		3 mos.	9	35	NR
		6 mos.	12	-18	NR
		12 mos.	22	7	NR
		24 mos.	25	8	NR
	Burkus (2002)		**rhBMP-2 (n = 143)**	**ICBG (n = 136)**	
		1.5 mos.	12	14	NS
		3 mos.	20	14	NS
		6 mos.	25	21	NS
		12 mos.	28	26	NS
		24 mos.	30	24	NS
ODI "success" (≥ 15% improvement from baseline score)			**% (n)**		
	Boden (2000)		**rhBMP-2 (n = 11)**	**ICBG (n = 3)**	
		1.5 mos.	NR	NR	
		3 mos.	55% (6)	0% (0)	NR
		6 mos.	64% (7)	67% (2)	NR
		12 mos.	91% (10)	67% (2)	NR
		24 mos.	91% (10)	67% (2)	NR
	Burkus (2002)		**rhBMP-2 (n = 143)**	**ICBG (n = 136)**	
		1.5 mos.	NR	NR	
		3 mos.	NR	NR	NS
		6 mos.	NR	NR	NS
		12 mos.	85% (122)	86% (117)	NS
		24 mos.	84% (120)	82% (136)	NS

Back pain NRS	Burkus (2002)	**Mean score improvement (from baseline) (points)**		
Range: 0 – 20		**rhBMP-2 (n = 143)**	**ICBG (n = 136)**	
	1.5 mos.	6.5	7.3	NS
	3 mos.	7.1	7.1	NS
	6 mos.	7.2	7.2	NS
	12 mos.	7.8	7.7	NS
	24 mos.	8.5	8.2	NS
Back pain "success" (> 3 point improvement from baseline score)	Burkus (2002)	**% (n)**		
		rhBMP-2 (n = 143)	**ICBG (n = 136)**	
	1.5 mos.	77% (110)	76% (103)	NS
	3 mos.	74% (106)	78% (106)	NS
	6 mos.	78% (112)	72% (98)	NS
	12 mos.	79% (113)	73% (99)	NS
	24 mos.	75% (107)	79% (107)	NS
Leg pain NRS	Burkus (2002)	**Mean score improvement (from baseline) (points)**		
Range: 0 – 20		**rhBMP-2 (n = 143)**	**ICBG (n = 136)**	
	1.5 mos.	5.0	4.1	NS
	3 mos.	5.7	5.7	NS
	6 mos.	6.2	6.2	NS
	12 mos.	6.2	5.9	NS
	24 mos.	6.2	6.2	NS
Leg pain "success" (> 3 point improvement from baseline score)	Burkus (2002)	**% (n)**		
		rhBMP-2 (n = 143)	**ICBG (n = 136)**	
	1.5 mos.	NR	NR	
	3 mos.	NR	NR	
	6 mos.	NR	NR	
	12 mos.	72% (103)	73% (99)	NS
	24 mos.	80% (114)	74% (101)	NS
SF-36	Boden (2000)	**Mean score improvement (from baseline) (points)**		
Physical function subscale		**rhBMP-2 (n = 11)**	**ICBG (n = 3)**	
Range: 0 – 100	3 mos.	10	13	NR
	6 mos.	18	27	NR
	12 mos.	27	37	NR
	24 mos.	38	37	NR
SF-36 General health perception subscale		**Mean score improvement (from baseline) (points)**		
Range: 0 – 100	Boden (2000)	**rhBMP-2 (n = 11)**	**ICBG (n = 3)**	
	3 mos.	6	-2	NR
	6 mos.	0	16	NR
	12 mos.	2	5	NR
	24 mos.	5	8	NR
Work status		**% of patients working (n)**		
	Boden (2000)	**rhBMP-2 (n = 11)**	**ICBG (n = 3)**	

	3 mos.	NR	NR	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	91% (10)	67% (2)	NS
Burkus (2002)		**rhBMP-2 (n = 143)**	**ICBG (n = 136)**	
	3 mos.	38% (54)	28% (38)	NS
	6 mos.	51% (73)	46% (63)	NS
	12 mos.	55% (79)	50% (68)	NS
	24 mos.	66% (94)	56% (76)	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Table 12. Efficacy of on-label use of rhBMP-2 in the lumbar spine: physician-reported outcomes

Outcome measure	Author	Time	Results		p-value
Neurological success			**% of patients with neurologic success (n)**		**p-value**
	Burkus (2002)		**rhBMP-2 (n = 143)**	**ICBG (n = 136)**	
Maintenance or improvement of preoperative scores for four neurological measurements		3 mos.	84% (120)	77% (105)	NS
		6 mos.	78% (112)	81% (110)	NS
		12 mos.	82% (117)	85% (116)	NS
		24 mos.	83% (119)	84% (114)	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Social function, mental health, quality of life outcomes were not reported by either of the two RCTs.

EFFECTIVENESS

Summary

Studies:

One integrated analysis (LoE II) met our inclusion criteria[23]. The analysis retrospectively pooled data from 679 patients that was reported in three studies, including one RCT (reported in efficacy) (n = 279), one case series (n = 22), and one unpublished study (n = 378). *Note that patients from the Burkus RCT make up 41% of the population reported in the integrated analysis (279/679 patients); thus results in this section partially overlap with those reported above*. A weakness of this integrated analysis is that more than half of the data (56%) were taken from an unpublished cohort study. Patients with DDD and radiculitis who were refractory to conservative care underwent primary single-level open (41%) or laparoscopic (59%) anterior lumbar fusion with either rhBMP-2/ACS (InFUSE) (n = 277) or iliac crest bone autograft (ICBG) (n = 402). Patients were followed for 24 months. Additional details are available in Table 13 and the surrounding text.

Outcomes:

Perioperative outcomes (Table 14): The following three outcomes were better in the rhBMP-2 group than in the ICBG group: operating time (1.8 versus 2.7 hours), blood loss (127 versus 193 mL), and length of hospital stay (2.2 versus 3.1 days). The strength of this evidence is *low*.

Fusion (Table 14): The percentage of patients with successful fusion was similar between groups at 6, 12, and 24 months. The strength of this evidence is low. By 24 months, 94% of the BMP and 89% of the ICBG patients had successful fusion.

ODI (Table 14): At six months and later, there were no clinically meaningful differences between groups in ODI outcomes. The strength of this evidence is *low*. At three months, the BMP group did have clinically meaningful improvement in ODI scores while the ICBG group did not (mean scores improvement of 31 versus 5 points, respectively). At six, 12, and 24 months, however, both groups had clinically meaningful improvement, and the differences between groups were not clinically meaningful (mean score improvement at 24 months: 31 versus 26 points, respectively).

SF-36: pain and function (Table 14): SF-36 pain and function outcomes were similar between groups. The strength of this evidence is *low*. While the mean score improvements in SF-36 pain index and physical component subscale scores were higher in the rhBMP-2 versus the ICBG group at all follow-ups, it is unlikely that these differences are clinically meaningful. The mean score improvement at 24 months for the pain index subscale was 39 (BMP) versus 33 (ICBG) points, and for the physical component subscale it was 16 (BMP) versus 12 (ICBG) points.

Work status (Table 14): A similar percentage of patients had returned to work by 24 months (75% versus 65% for rhBMP-2 versus ICBG, respectively). However, those in the rhBMP-2 group returned to work a median of 55 days sooner than those in the control group. The strength of this evidence is *low*.

Detailed study characteristics

Table 13. On-label use of rhBMP-2 in the lumbar spine: cohort study overview

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Notes
Burkus (2003)[23]	InFUSE (n = 277) (dose NR)	LT-CAGE	Anterior (open or lapro-scopic)	DDD, radiculitis	24 mos. 85% (574/	II	NR* (Medtronic according to Carragee[47])	Data sources: (1) Burkus 2002 RCT[6] (n = 279) (2) Kleeman

	ICBG (n = 402)	LT-CAGE	1 level/pt primary		679)			2001[24] case series (n = 22) (3) Unpublished data set from a third study

DDD: degenerative disc disease; f/u: follow-up; InFUSE: rhBMP-2 applied to an absorbable collagen sponge (ACS) (Medtronic); NR: not reported
* According to Carragee et al. (2011)[47], the authors had a financial relationship with Medtronic ranging from $22,732,000 - $23,192,000 plus an additional $1,500,000 for Burkus.

The Burkus et al. (2003) retrospective integrated analysis[23] pooled data from the Burkus et al. (2002) RCT (n = 279)[6] (reported above), a small case series (n = 22)[24], and an unpublished cohort study (n = 378). Patients from the Burkus trial make up 41% of the population reported in the integrated analysis (279/679 patients); thus results from this study overlap in part with those reported above. Patients in this randomized trial underwent open anterior lumbar fusion[6]. Patients from the nonrandomized unpublished trial underwent laparoscopic surgery and make up 56% of the represented population (378/679). Inclusion and exclusion criteria were identical for all studies, except that there was no minimum ODI score required for enrollment in the unpublished trial[23]. The study received a LoE grade of II. It appeared that all primary outcomes were either patient-reported or evaluated by a blinded interpreter. Co-interventions were applied equally, and there was complete follow-up of 85% of patients. The authors summarized twenty different baseline characteristics and controlled for the seven that were significantly different between treatment groups (age, previous back surgery, use of non-narcotic medications, use of weak narcotic medications, use of muscle relaxant medications, preoperative low back pain score on the ODI, and preoperative SF-36 physical component score) using analysis of covariance. The primary weakness of this integrated analysis is that more than half of the data were taken from an unpublished cohort study.

Table 14. Effectiveness of on-label use of rhBMP-2 in the lumbar spine: results from one cohort study

	Treatment groups	Perioperative (mean ± SD)	Radiographic	Pain	Function	Work status
Burkus (2003) LoE II	InFUSE/LT-CAGE (n = 277) vs. ICBG/ LT-CAGE (n = 402)	OR time 1.8 ± 0.8 vs. 2.7 ± 1.3 hrs ($P < .001$) Blood loss 127 ± 295 vs. 193 ± 414 mL ($P = 0.024$) Hospital LOS 2.2 ± 1.7 vs. 3.1 ± 3.2 ($P < .001$)	Successful fusion (%, n): 6 mos: 95% (263) vs. 96% (386) 12 mos: 96% (266) vs. 93% (374) 24 mos: 94% (260) vs. 89% (358) ($P = .022$)	ODI (mean score improvement): 3 mos: 31 vs. 5 ($P = .0041$) 6 mos: 26 vs. 20 ($P = .0053$) 12 mos: 30 vs. 23 ($P = .0013$) 24 mos: 31 vs. 26 ($P = .0023$) SF-36 pain index subscale (mean score improvement): 3 mos: 27 vs. 20 ($P = .0002$) 6 mos: 32 vs. 24 ($P = .0002$) 12 mos: 36 vs. 29 ($P = .0002$) 24 mos: 39 vs. 33 ($P = .0008$)	SF-36 physical component subscale (mean score improvement): 3 mos: 9 vs. 5 ($P = .0015$) 6 mos: 12 vs. 8 ($P = .0004$) 12 mos: 14 vs. 10 ($P = .0003$) 24 mos: 16 vs. 12 ($P = .0007$)	Return to work† (%, n): 24 mos: 75% (103) vs. 65% (109) (P = NS) rhBMP-2 group returned to work a median of 55 days sooner than patients in the ICBG group (adjusted $P = .0156$)

LOS: length of stay
*Fusion was considered successful: (1) on plain radiograph when: there was less than 5 degrees of angular motion of flexion-extension film, there was less than 3 mm translation, and there was an absence of radiolucent lines over 50% or more of the implant surfaces; or (2) on CT when there was evidence of the presence of continuous trabecular bone growing through both cages.
† Return to work: patients who were working preoperatively had returned to work.

4.2.2. rhBMP-7 on-label use

No studies were identified that evaluated on-label use of rhBMP-7 (OP-1).

Note that while OP-1 has received a humanitarian device exemption (HDE) from the FDA (H020008) for "use as an alternative to autograft in compromised patients [i.e., osteoporotic, smokers, diabetics] requiring revision posterolateral (intertransverse) lumbar spinal fusion, for whom autologous bone and bone marrow harvest are not feasible or are not expected to promote fusion,"[3] the pilot and pivotal trials evaluated primary (not revision) posterolateral fusion patients and these trials are therefore not in accordance with on-label use of OP-1.

4.2.3. rhBMP-2 off-label use: lumbar spine

EFFICACY

Summary

Studies:

Six RCTs met our inclusion criteria. One study received an LoE grade of IIa, and the remaining five studies were graded LoE IIb. Study size ranged from 27 to 463 patients. Patients with DDD, radiculitis, and/or up to 25% spondylolisthesis who were refractory to conservative care underwent primary single- (or in one study, multi-) level posterior (four studies), anterior (one study), or posterolateral (one study) lumbar fusion with either rhBMP-2/ACS (InFUSE) or iliac crest bone autograft (ICBG). Due to heterogeneity in surgical procedures (i.e., approach, use of ceramic granules, use of cage versus allograft dowel versus no device, single- versus multilevel design), we were not able to pool outcomes data from the six studies. Patients received BMP in a variety of forms: rhBMP-2/CRM; InFUSE; and AMPLIFY. Doses of rhBMP-2 varied and ranged from 4.2 to 40 mg per patient (when reported). Patients were followed for 17 (mean) to 24 months. Five RCTs were sponsored by Medtronic and one by a Norton Healthcare grant. Additional details are available in Table 15 and the surrounding text.

Outcomes:

Perioperative outcomes (Table 15):

Operative time: Operative times were similar or improved with rhBMP-2 compared with control treatment (6 RCTs). The strength of this evidence is *moderate*. Three of the studies reported similar mean operative times for both groups, with the difference between groups ranging from 0.1 to 0.4 hours). The other three reported statistically shorter operative time for patients receiving rhBMP2 compared with ICBG (difference between groups ranged from 0.4 to 0.5 hours).

Blood loss: Mean perioperative blood loss volumes were similar or improved (ie., lower) with rhBMP-2 compared with control treatment (6 RCTs). The strength of this evidence is *moderate*. Four studies reported similar mean blood loss between groups, with differences ranging from 5 to 123 mL. Two studies reported statistically lower mean blood loss in the rhBMP-2 group, with differences ranging from 98 to 106 mL, although it is unclear whether these differences are clinically meaningful.

Length of hospital stay: There was no difference between groups in the mean length of hospital stay (5 RCTs). The strength of this evidence is *high*. Mean length of stay ranged from 2.9 to 4.1 days in the rhBMP-2 groups and from 3.3 to 5.2 days in the control group.

Fusion (Table 17): Fusion rates were similar or improved with rhBMP-2 versus ICBG treatment (6 RCTs). While there may be a slight improvement in fusion success in those treated with rhBMP-2 compared with ICBG, it is not clear that the differences are clinically meaningful. The strength of this evidence is *moderate*. With six studies reporting, the proportions of patients with successful fusion at 24 months ranged from 86-100% in the rhBMP-2 group and from 67-89% in the ICBG group (the latter excludes one study with only 5 patients in the control group; otherwise the range is 40-89% for the ICBG group). More specifically, outcomes were reported as similar or improved: three studies reported that the percentage of patients with successful fusion was similar between treatment groups at all follow-ups, while another three studies reported statistically meaningful improvements in the rhBMP-2 versus the ICBG groups.

ODI (Table 18): In general, ODI outcomes were similar between groups at all reported follow-ups in terms of both the percentage of patients with ODI "success" (defined as improvement from baseline scores by at least 15% to 20%, depending on the study) (reported by three RCTs) and the mean ODI score improvement from baseline (reported by six RCTs). The strength of this evidence is *high*. Regarding ODI "success", all three studies reporting statistically similar incidences of "success" at last follow-up (17-24 months), with the proportion of patients with ODI "success" ranging from 2 to 21% higher in the BMP. Regarding mean score improvement, in five RCTs there was clinically meaningful improvement (ie., improvement by at least 15 points) in both treatment groups at 24 months, but the difference between the groups was not clinically meaningful. In one RCT (N = 46), the BMP group had clinically meaningful improvement in their mean ODI scores while the ICBG group did not; however, the difference in mean score improvement was only 2 points (15 versus 13 points, respectively).

Pain (Table 18): Back pain outcomes were clinically similar in five RCTs and clinically improved in one RCT with rhBMP-2 treatment at final follow-up of 17 to 24 months. The strength of this evidence is *moderate*. Mean improvements in back pain scores at 17-24 months ranged from 3.1 to 9.6 points in the rhBMP-2 groups and from 3 to 8 points in the ICBG groups; differences between groups ranged from 0.1 to 4.5 points (6 RCTs). One RCT (N = 67) reported clinically and statistically improved 24 month scores in rhBMP-2 versus the ICBG groups (9 versus 4.5 (of 20 possible) points, respectively). Clinically meaningful improvement is considered to be 20mm on a 100 cm scale[4], which translates to 4 points on the 20 point scales used here. Regarding leg pain, all six RCTs reported similar clinical improvements in leg pain VAS scores between groups at final follow-up of 17 to 24 months. The strength of this evidence is *high*. Mean VAS score improvements in leg pain ranged from 3.6 to 9.3 points in the rhBMP-2 groups and from 3.1 to 7.2 points in the ICBG groups; differences between groups ranged from 0.5 to 2.1 points.

SF-36: function (Table 18): Outcomes were similar between groups; the strength of this evidence is *high*. All six RCTs reported improvements in the SF-36 physical component subscale for both the rhBMP-2 and ICBG treatment groups, which ranged from 7 to 15 points in the rhBMP-2 and 7 to 17 points in the ICBG groups at final follow-up (17 to 24 months). Regarding the differences between groups in mean score improvement, five RCTs reported no differences between treatment groups at 17 to 24 months, while one RCT reported that the rhBMP-2 group had statistically better improvements compared with the ICBG group at 6, 12, and 24 months ($P \leq .02$). In this case, however, it is unlikely that the small score differences reported between treatment groups (of 6, 8, and 3 points at each follow-up, respectively) are clinically meaningful.

Patient satisfaction: Patient satisfaction was similar between treatment groups at 17 and 24 months as reported by two RCTs (N = 27 – 67). The strength of this evidence is *low*.

Work status: There were no differences in work status between treatment groups as reported by four RCTs. The strength of this evidence is *high*. Two RCTs reported that similar percentages of patients between groups were working or had returned to work in both treatment groups at 24 months follow-up, and two other RCTs reported no difference in mean time to return to work between treatment groups.

Neurological status: There were no differences in neurological success between groups at 24 months as reported by one RCT (N = 67). The strength of this evidence is *low*.

Overall success: The composite measure of "overall success" was reported by one RCT (N = 41) and defined as a combination of successful fusion, ODI success, an absence of severe adverse events, an absence of secondary surgical procedures at the index level, and maintenance or improvement in neurological status. The incidence of "overall success" was similar between groups at 24 months. The strength of this evidence is *low*.

Detailed study characteristics:

Six randomized controlled trials[1, 7-9, 12, 25, 26] were identified that evaluated off-label use of rhBMP-2 in the lumbar spine. All studies compared rhBMP-2 with ICBG. Detailed data abstraction tables can be found in Appendix F. The study characteristics are summarized in Table 15.

Table 15. Off-label use of rhBMP-2 in the lumbar spine: RCT study overview

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Notes
Boden (2002)[7]	BMP2 (n = 11) (40 mg/pt)	CRM /TSRHSS/ CBD	PLF 1 level/pt primary	DDD, grade ≤ 1 spondylo., radiculitis	mean 17 (12-27) mos. 93% (25/27)	IIb	Yes (sponsor NR)* (Medtronic according to Carragee[47])	Pilot trial for device which has not received FDA approval
	BMP2 (n = 11) (40 mg/pt)	CRM/CBD						
	ICBG (n = 5)	TSRHSS/CBD						
Dawson (2009)[25]	InFUSE (n = 25) (12 mg/pt)	CRM/ instrum.	PLF 1 level/pt primary	DDD, grade ≤ 1 spondylo., radiculitis	24 mos. 87% (40/46)	IIb	Medtronic‡	Pilot study for InFUSE/ Mastergraft device which has not received FDA approval; HDE voluntarily withdrawn by Medtronic in 2010
	ICBG (n = 21)	instrum.						
Dimar (2009)[8]	AMPLIFY (n = 239) (40 mg/pt)	CRM/ instrum.	PLF 1 level/pt primary	DDD, grade ≤ 1 spondylo., radiculitis	24 mos. 89% (410/ 463)	IIb	Medtronic‡	AMPLIFY (Medtronic) = rhBMP-2 matrix
	ICBG (n = 224)	instrum.						
Glassman (2008)[26]	InFUSE (n = 50) (dose NR)	instrum.	PLF mean 2 levels/pt (range NR) primary	DDD, stenosis, spondylo., instabil., adj. segment degen.	24 mos. 94% (100/ 106)	IIb	Norton Healthcare (grant)	Age > 60 years 2 pts from each arm excluded perioperatively
	ICBG (n = 52)							
Burkus (2005)[12]/ **(2006)**[1]	InFUSE (n = 79) (8-12 mg/pt)	cortical bone dowels (CBD)/ instrum.	ALIF 1 level/pt primary	DDD, grade ≤ 1 spondylo.	24 mos. 96% (126/ 131)	IIa	Medtronic†	
	ICBG (n = 52)							
Haid (2004)[9]	InFUSE (n = 34) (4.2-8.4 mg/pt)	interbody fusion cages	PLIF 1 level/pt	DDD, grade ≤ 1 spondylo., radiculitis	24 mos. 94%	IIb	NR§ (Medtronic according to	Study halted after prelim CT scans showed

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Notes
	ICBG (n = 33)		primary		(63/67)		Carragee[47])	bone growth posterior to cages; not restarted.

ALIF: anterior lumbar interbody fusion; CRM: compression resistant matrix (i.e., ceramic granules: mixture of hydroxyapatite + tricalcium phosphate, percentage varies by study); CBD: cortical bone dowels (allograft); DDD; degenerative disc disease; f/u: follow-up; InFUSE: rhBMP-2 applied to an absorbable collagen sponge (ACS) (Medtronic); NR: not reported; PLIF: posterior lumbar interbody fusion; PLF: posterolateral lumbar fusion; spondylo.: spondylolisthesis; TSRHSS: Texas Scottish Rite Hospital Spinal System pedicle screw instrumentation

According to Carragee et al. (2011)[47], the authors had a financial relationship with Medtronic ranging from:
* $565,000–1,474,000
† $1,705,000–2,167,000 plus an additional $1,500,000 for Burkus
‡ $12,101,000–16,021,000 (sum accounts for Glassman but not Dimar), plus an additional $1,500,000 for Burkus
§ $5,844,000–5,854,000, plus an additional $1,500,000 for Burkus

- Boden (2002)[7] conducted a small RCT in which N = 27 patients were randomized 1:2:2 to receive primary single-level posterolateral lumbar fusion (PLF) using threaded cortical allograft dowels filled with (1) rhBMP-2, ceramic granules, and Texas Scottish Rite Hospital (TSRH) pedicle screw instrumentation (n = 11); (2) rhBMP-2 and ceramic granules (n = 11); or (3) ICBG with the TSRH instrumentation. The primary patient diagnosis was DDD with or without radiculitis with grade 0 or 1 spondylolisthesis; all patients had been unresponsive to conservative care for six months or longer. Ninety-three percent (25/27) of patients were followed for a mean of 17 months (range, 12-27 months). The study was funded by Medtronic. Boden (2002) received a LoE grade of IIb. No information was provided regarding random sequence generation or how allocation concealment was ensured. While there was no mention of following the intention to treat principle, data appear to have been handled in accordance with this method. The study did not adequately control for possible confounding: 40% of patients in the autograft group had diabetes compared with 0% in either rhBMP-2 treatment group (P = .036), and the differences not controlled for. All outcomes were either patient-reported or evaluated in a blinded manner.

- Dawson (2009)[25] randomized 46 patients with DDD to receive single-level primary instrumented PLF with InFUSE (rhBMP2/ACS) (n = 25) wrapped around ceramic granules or ICBG (n = 21). Outcomes were reported at 24 months; complete follow-up data were available for 87% of patients. The study was funded by Medtronic. Dawson (2009) received a LoE grade of IIb. Randomization was stratified by site and achieved with a fixed block size of four. No information was provided regarding allocation concealment; two patients in each treatment group chose not to participate in the study, though whether the patients were aware of their treatment allocation was not reported. The study used a modified intent-to-treat analysis in which data for patients who had failed was carried forward from the last available follow-up prior to failure. Credit was

not given for controlling for confounding between groups, as baseline scores for outcome measures (i.e., ODI, pain) were not reported. While the authors used regression analysis to control for differences in demographics (Workers' Comp, litigation, previous spinal surgery), this does not appear to have been done to control for potential differences in preoperative scores. All outcomes were either patient-reported or evaluated in a blinded manner.

- Dimar (2009)[8] conducted a large RCT in which a total of 463 patients underwent single-level primary instrumented PLF with AMPLIFY (rhBMP2 matrix consisting of 15% hydroxyapatite and 85% ß-tricalcium phosphate particles, Medtronic Sofamor Danek) (n = 239) or ICBG (n = 224). All enrolled patients had DDD that was refractory to at least six months conservative care. The study was sponsored by Medtronic. Dimar (2009) received a LoE grade of IIb. Randomization was performed at a central location and achieved by stratification by site with a fixed block size of 4; although treatment allocation was concealed using sealed envelopes with sequential numbers, there was no mention that the envelopes were opaque. This study used an "as-treated" rather than the intention to treat analysis. The study received partial credit for possible confounding as there was only one statistically meaningful baseline difference (spinal litigation) in the 15+ baseline characteristics or preoperative data reported; however differences in spinal litigation rates between groups was not controlled for or discussed. All outcomes were either patient-reported or evaluated in a blinded manner.

- Glassman (2008)[26] enrolled 106 patients over 60 years of age with DDD, spondylolisthesis, stenosis, deformity, instability, postdecompression revision, or adjacent level fusion. Patients received single- or multilevel (mean of 2 levels treated) instrumented PLF with either InFUSE (rhBMP2/ACS) (n = 50) or ICBG (n = 52). Two patients were excluded from each treatment group in the immediate postoperative period (1 perioperative death, 1 procedure extended beyond L1-S1, 1 patient did not undergo fusion, and 1 patient who refused any postoperative follow-up after changing physicians). The study was funded by a grant from Norton Healthcare. This study received an LoE grade of IIb. No information was provided regarding random sequence generation or allocation concealment. Although the authors stated that data were analyzed according to the intention to treat principle, patients who failed treatment and required revision had data from their last observation carried forward. Regarding equal treatment in both groups, while the mean number of levels fused was similar between groups, the authors did not report the number of patients in each group that underwent 1-, 2-, 3- etc. level fusion. The study did not control for possible cofounding; while there were no differences in baseline characteristics, there was a difference in preoperative leg pain scores between groups that was not controlled for. All outcomes were either patient-reported or evaluated in a blinded manner.

- Burkus (2005)[12]/(2006)[1] randomized 131 patients with DDD to undergo single-level primary instrumented anterior lumbar interbody fusion with cortical bone dowels filled with either InFUSE (rhBMP-2/ACS) (n = 79) or ICBG (n = 52). Outcomes were reported at 24 months, at which point complete follow-up data were available for 96%

of patients. The study was sponsored by Medtronic. Burkus (2005) received a LoE grade of IIa. Patients were randomized using a statistical program (SAS), which produced sequentially numbered envelopes for each enrollment site. The authors clearly stated that surgeons remained blinded to the randomization schedule. There was no explicit statement that data were analyzed in accordance with the intention to treat principle; data from patients classified as failures (i.e., had to undergo device removals, revisions, or supplemental fixation) were not reported after they had failed the treatment. The study did not adequately control for possible cofounding; while there were no differences in baseline characteristics, there was a difference in preoperative back pain scores between groups (P = .039) that was not controlled for. All outcomes were either patient-reported or evaluated in a blinded manner.

- Haid (2004)[9] enrolled 67 patients with DDD and radiculitis who were unresponsive to at least 6 months' nonoperative treatment. Patients were randomized to receive single-level primary posterior lumbar interbody fusion (PLIF) using cylindrical interbody cages filled with either InFUSE (rhBMP-2/ACS) or morselized autograft (n = 33). Complete data were available for 94% of patients at the 24-month follow-up. The study has been reported to have been sponsored by Medtronic[47]. The study received a LoE grade of IIb. No information was provided regarding random sequence generation or allocation concealment. Data appear to have been handled in accordance with the intention to treat principle, although there was no explicit statement indicating that this method was used for data analysis. Baseline characteristics and preoperative scores were similar between treatment groups, and all outcomes were either patient-reported or evaluated in a blinded manner.

Detailed outcomes

Perioperative outcomes (Table 16)

- Operative time (6 RCTs)
  Overall, rhBMP-2 use was associated with similar or shorter operating times compared with ICBG. Half of the studies reported no difference in mean operative time between groups[7, 9, 25], while the other half reported statistically shorter operative time for patients receiving rhBMP2 compared with ICBG ($P \leq .024$)[1, 8, 12, 26]. This increased operative time in the ICBG group is likely due to the additional time it takes to harvest the autograft. Mean operative time in the rhBMP-2 groups ranged from 1.4 to 4.1 hours, and for the ICBG groups from 1.9 to 4.5 hours.

- Blood loss (6 RCTs)
  Similarly, use of rhBMP-2 resulted in similar or less blood loss compared with use of ICBG. Four studies reporting similar average blood loss between groups[7, 9, 25, 26] and two studies reporting lower mean blood loss in the rhBMP-2 group[1, 8, 12]. Average blood loss ranged from 87 to 670 milliliters in the rhBMP-2 groups and from 185 to 675 milliliters in the ICBG group.

- Length of stay (5 RCTs)

There was no difference between groups in the mean hospital length of stay as reported by five RCTs[1, 7-9, 12, 25]. Mean length of stay ranged from 2.9 to 4.1 days in the rhBMP-2 groups and from 3.3 to 5.2 days in the control group.

Table 16. Efficacy of off-label use of rhBMP-2 in the lumbar spine: perioperative outcomes

Outcome measure	Author	Results		p-value
Operative time		**Mean ± SD (hours)**		
	Boden (2002)	**rhBMP-2 (n = 22)***	**ICBG (n = 5)**	
		2.9 ± 0.3	3.1 ± 0.4	NR
	Dawson (2009)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
		2.7 ± 0.7	2.8 ± 0.8	NS
	Dimar (2009)	**rhBMP-2 (n = 239)**	**ICBG (n = 224)**	
		2.5 ± 0.09	2.9 ± 1.0	< .001
	Glassman (2008)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
		4.1 ± 0.6	4.5 ± 1.0	.024
	Burkus (2005)/(2006)	**rhBMP-2 (n = 79)**	**ICBG (n = 52)**	
		1.4	1.9	< .001
	Haid (2004)	**rhBMP-2 (n = 34)**	**ICBG (n = 33)**	
		2.6	3.0	NS
Blood loss		**Mean ± SD (mL)**		
	Boden (2002)	**rhBMP-2 (n = 22)***	**ICBG (n = 5)**	
		455 ± 117	430 ± 82	NS
	Dawson (2009)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
		329 ± 212	452 ± 210	NS
	Dimar (2009)	**rhBMP-2 (n = 239)**	**ICBG (n = 224)**	
		343 ± 265	449 ± 302	< .001
	Glassman (2008)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
		670 ± 487	675 ± 456	NS
	Burkus (2005)/(2006)	**rhBMP-2 (n = 79)**	**ICBG (n = 52)**	
		87	185	< .001
	Haid (2004)	**rhBMP-2 (n = 34)**	**ICBG (n = 33)**	
		323	373	NS
Length of hospital stay		**Mean ± SD (days)**		
	Boden (2002)	**rhBMP-2 (n = 22)***	**ICBG (n = 5)**	
		3.7 ± 0.5	4.4 ± 0.5	NS

Dawson (2009)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
	4.0 ± 1.4	4.1 ± 1.1	NS
Dimar (2009)	**rhBMP-2 (n = 239)**	**ICBG (n = 224)**	
	4.1 ± 2.3	4.0 ± 1.9	NS
Glassman (2008)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
	NR	NR	
Burkus (2005)/(2006)	**rhBMP-2 (n = 79)**	**ICBG (n = 52)**	
	2.9	3.3	NS
Haid (2004)	**rhBMP-2 (n = 34)**	**ICBG (n = 33)**	
	3.4	5.2	NS (.065)

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation
* Boden (2002): results pooled for both rhBMP-2 groups (i.e., with (n = 11) or without (n = 11) instrumentation).

Radiographic outcomes (Table 17)

- Definitions of successful fusion varied by study, but in general encompassed the following criteria as evaluated on plain radiograph or CT[1, 7-9, 12, 25, 26]:
    - Presence of bridging trabecular bone between the transverse processes, and
    - Absence of motion (≤ 3 mm of translation and < 5º angular motion on flexion-extension views, and
    - Absence of radiolucent lines through the fusion mass or around more than half of the graft.

- Successful fusion (6 RCTs):
  Overall, rhBMP-2 use was associated with similar or higher rates of fusion success compared with ICBG use. Three studies reported considerably higher 24-month fusion rates in the rhBMP-2 group compared with the ICBG group[1, 7, 8, 12], while another three reported similar rates between the groups[9, 25, 26]. While there may be a slight improvement in fusion success in those treated with rhBMP-2 compared with ICBG, it is not clear that the differences are clinically meaningful. In general, rates of successful fusion at 24 months follow-up ranged from 86-100% in the rhBMP-2 group and from 67-89% in the ICBG group (when the one study with only 5 patients in the control group is excluded; otherwise the range is 40-89%).

  Glassman (2008) also reported that the average CT fusion grade was statistically higher in the rhBMP-2 group compared with the ICBG group at 24 months (4.3 ± 1.3 vs. 3.8 ± 0.9, respectively; $P = .030$)[26]. Fusion was graded on a scale from 1 (no fusion) to 5 (solid bilateral fusion). Burkus (2005) noted that 18% (14/79) patients in the rhBMP-2 group had localized areas of bone remodeling in the vertebral body next to the femoral ring allograft between 3 and 12 months but were resolved by 24 months[12].

- **Time to fusion** was not reported by any of the six RCTs[1, 7-9, 12, 25, 26].

Table 17. Efficacy of off-label use of rhBMP-2 in the lumbar spine: radiographic outcomes

Outcome measure	Author	Results		p-value
Fusion success		**% (n)**		
(see text for definition)	Boden (2002)	**rhBMP-2 (n = 20/22)***	**ICBG (n = 5)**	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	100% (20)	40% (2)	< .03
	Dawson (2009)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
	6 mos.	91% (23)	58% (12)	.032
	12 mos.	89% (22)	67% (14)	NS
	24 mos.	95% (24)	67% (14)	NS
	Dimar (2009)	**rhBMP-2 (n = 239)**	**ICBG (n = 224)**	
	6 mos.	79% (189)	65% (146)	.002
	12 mos.	88% (210)	83% (186)	NS
	24 mos.	96% (229)	89% (199)	.014
	Glassman (2008)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	86% (43)	71% (37)	NR
	Burkus (2005)/(2006)	**rhBMP-2 (n = 79)**	**ICBG (n = 52)**	
	6 mos.	96% (76)	85% (44)	.047
	12 mos.	99% (78)	89% (46)	.035
	24 mos.	98% (77)	76% (40)	< .001
	Haid (2004)	**rhBMP-2 (n = 34)**	**ICBG (n = 33)**	
	6 mos.	93% (32)	93% (31)	NS
	12 mos.	85% (29)	92% (30)	NS
	24 mos.	92% (31)	78% (26)	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation
* Boden (2002): results pooled for both rhBMP-2 groups (i.e., with (n = 11) or without (n = 11) instrumentation). Radiographic outcomes were available for 20/22 patients.

Pain (Table 18)

- ODI (3-6 RCTs)
  Mean score improvement in ODI scores were similar between groups at 24 months as reported by five RCTs[1, 8, 9, 12, 25, 26]; one RCT reported no difference between groups at final follow-up of 17 months[7]. Mean score improvement in the rhBMP-2 groups at final follow-up ranged from 15-33 points and in the ICBG groups from 13-27 points[1, 7-9, 12, 25, 26]. The ODI is discussed in greater detail in Key Question 1; scores range from 0-100 points..

  "Success": There was no difference between groups in the percentage of patients with ODI "success" (defined as improvement from baseline by at least 15% (or 20% in Dawson's RCT) at 17-24 months postsurgery as reported by three RCTs[7, 9, 25]. Rates of

ODI success at 17-24 months ranged from 69-91% in the rhBMP-2 group and from 56-80% in the ICBG group.

- Back pain (6 RCTs)
  Mean score improvement: Back pain intensity and frequency/duration were clinically similar or improved with rhBMP-2 treatment at 17-24 months. Four studies reported similar improvement in back pain scores between treatment groups at 17-24 months,[7, 8, 25, 26] while two studies reported statistically better improvements in back pain scores in the rhBMP-2 versus the ICBG groups[1, 9, 12]. Of the two studies that reported statistically significant improvements in back pain, only one demonstrated clinically meaningful improvements in rhBMP-2 versus the ICBG groups (9 versus 4.5 (of 20 possible) points, respectively)[9]. Back pain was evaluated on a 20-point numerical rating scale (consisting of the sum of the scores for pain intensity (0-10) and pain frequency or duration (0-10)). Mean improvements in back pain scores at 17-24 months ranged from 3.1-9.6 points in the rhBMP-2 groups and from 3-8 points in the ICBG groups.

  "Success": Back pain success was not evaluated by any of the six RCTs.

- Leg pain (6 RCTs)
  Mean score improvement: Treatment with rhBMP-2 was associated with clinically similar leg pain intensity and frequency/duration compared with treatment with ICBG at 17-24 months. Five RCTs reported similar mean score improvement between groups at 17-24 months[7-9, 25, 26]; one RCT reported statistically greater improvement in the rhBMP-2 group at 6, 12, and 24 month postoperation[1, 12], however, the difference between groups was not clinically meaningful. Leg pain was evaluated on a 20-point numerical rating scale (consisting of the sum of the scores for pain intensity (0-10) and pain frequency/duration (0-10)).

  "Success": Leg pain success was not evaluated by any of the six RCTs.

Function (Table 18)

- SF-36 physical component subscale (6 RCTs)
  Patients who received rhBMP-2 had similar improvements in SF-36 physical component subscale scores at 17-24 months follow-up. Five studies reported no differences between treatment groups at 17-24 months[7-9, 25, 26], while one study reported statistically better outcomes in the rhBMP-2 group compared with the ICBG group at 6, 12, and 24 months ($P \leq .02$) [1, 12], however, the small difference in scores between groups are not likely to be clinically meaningful. Mean SF-36 physical component subscale score improvements at 17-24 months follow-up ranged from 7-15 points in the rhBMP-2 groups, and from 7-17 points in the ICBG groups.

- SF-36 physical function subscale (1 RCT)
  Dawson et al., (2009) reported similar mean SF-36 physical function score improvement between groups at 24 months (36 versus 18 for rhBMP-2 (n = 25) versus ICBG (n = 21), respectively; $P \geq .05$)[25].

Patient satisfaction and quality of life

- Patient satisfaction (2 RCTs)
  Patient satisfaction was similar between treatment groups at 17 and 24 months as reported by two RCTs[7, 9].
    - Boden et al. (2002) reported that 82% (18/22) of patients in the pooled rhBMP-2 groups and 60% (3/5) of those in the ICBG group were considered their surgical outcome to be good or excellent[7].
    - Haid et al. (2004) found that in the rhBMP-2 group, 72% (25/34) of patients reported that they were definitely or mostly satisfied with their surgery outcomes compared with 80% (26/33) of patients in the ICBG group[9].

Social function or mental health outcomes were not reported by any of the six RCTs.

Work status (4 RCTs)

Percent of patients working

Two RCTs reported that similar percentages of patients between groups were working or had returned to work in both treatment groups at 24 months follow-up[8, 25]:

- Dawson (2004): 35% (8/23) versus 30% (6/20) working (rhBMP-2 versus ICBG (p-value not reported))[25].
- Dawson (2004): 100% (6/6) versus 67% (6/9) had returned to work (rhBMP-2 versus ICBG (p-value not reported))[25]
- Dimar (2009): 42% (87/207) versus 48% (89/184) had returned to work (rhBMP-2 versus ICBG ($P \geq .05$))[8].

Average days to return to work

Two RCTs reported no difference in mean time to return to work between treatment groups[1, 9, 12]:

- Burkus (2005)/(2005): 89 versus 96 days (rhBMP-2 (n = 79) versus ICBG (n = 52)) ($P \geq .05$)[1, 12].
- Haid (2004): 43 versus 137 (rhBMP-2 (n = 34) versus ICBG (n = 52)) ($P \geq .05$)[9].

Neurological status (1 RCT)

- Haid et al. (2004) reported that at 24 months, 100% of patients in both groups had neurological success[9]. Neurological success was defined as success (not defined) in each of four neurologic measurements (straight leg raise and motor, sensory, and reflex functions).

Overall success (1 RCT)

The primary outcome of interest as reported in the Dawson (2009) RCT was the composite measure of "overall success", which was defined as a combination of successful fusion, ODI success, an absence of severe adverse events, an absence of secondary surgical procedures at the index level, and maintenance or improvement in

neurological status. At 24 months, there was not a statistically meaningful difference in the rates of overall success between treatment groups (81% (17/21) versus 55% (11/20) for rhBMP-2 versus ICBG, respectively; $P = .345$)[25].

Physician impression of outcome (1 RCT)

Boden et al. (2004) reported that the physician impression of patient outcome was good or excellent for 82% (18/22) of patients in the pooled rhBMP-2 groups and for 60% (3/5) of those in the ICBG group[7]. Both patient and patient satisfaction were identical in this study.

Table 18. Efficacy of off-label use of rhBMP-2 in the lumbar spine: patient-reported outcomes

Outcome measure	Author	Results		p-value
ODI		**Mean score improvement (from baseline) (points)**		
Range: 0 – 100	Boden (2002)	**rhBMP-2 (n = 22)***	**ICBG (n = 5)**	
	1.5 mos.	~11	~10	NS
	3 mos.	~20	~15	NS
	6 mos.	~23	~17	NS
	17 mos.	~21	~25	NS
	Dawson (2009)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
	1.5 mos.	NR	NR	
	3 mos.	NR	NR	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	28	23	NS
	Dimar (2009)	**rhBMP-2 (n = 239)**	**ICBG (n = 224)**	
	1.5 mos.	NR	NR	
	3 mos.	NR	NR	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	~27	~27	NS
	Glassman (2008)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
	1.5 mos.	NR	NR	
	3 mos.	14	13	NS
	6 mos.	18	17	NS
	12 mos.	19	18	NS
	24 mos.	15	13	NS
	Burkus (2005)/(2006)	**rhBMP-2 (n = 79)**	**ICBG (n = 52)**	
	1.5 mos.	NR	NR	
	3 mos.	NR	NR	
	6 mos.	32	26	.031
	12 mos.	33	27	NS
	24 mos.	33	27	NS

	Haid (2004)	**rhBMP-2 (n = 34)**	**ICBG (n = 33)**	
	1.5 mos.	NR	NR	
	3 mos.	NR	NR	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	30	25	NS
ODI "success" (≥ 15% (or > 20%†) improvement from baseline score)		**% (n)**		
	Boden (2002)	**rhBMP-2 (n = 22)***	**ICBG (n = 5)**	
	1.5 mos.	63% (14)	80% (4)	NS
	3 mos.	84% (18)	60% (3)	NS
	6 mos.	86% (19)	80% (4)	NS
	17 mos.	82% (18)	80% (4)	NS
	Dawson (2009)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
	24 mos.	91% (23)†	70% (15)†	NS
	Dimar (2009)	**rhBMP-2 (n = 239)**	**ICBG (n = 224)**	
		NR	NR	
	Glassman (2008)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
		NR	NR	
	Burkus (2005)/(2006)	**rhBMP-2 (n = 79)**	**ICBG (n = 52)**	
		NR	NR	
	Haid (2004)	**rhBMP-2 (n = 34)**	**ICBG (n = 33)**	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	69% (24)	56% (19)	NS
Back pain Range: 0 – 20		**Mean score improvement (from baseline) (points)**		
	Boden (2002)	**rhBMP-2 (n = 22)***	**ICBG (n = 5)**	
	1.5 mos.	~7	~7	NS
	3 mos.	~8.5	~5	NS
	6 mos.	~8	~4	NS
	17 mos.	~8	~5	NS
	Dawson (2009)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
	1.5 mos.	NR	NR	
	3 mos.	NR	NR	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	9.6	7.2	NS
	Dimar (2009)	**rhBMP-2 (n = 239)**	**ICBG (n = 224)**	
	1.5 mos.	NR	NR	
	3 mos.	NR	NR	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	~9	~8	NS

Glassman (2008)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
1.5 mos.	4.3	4.0	NS
3 mos.	NR	NR	
6 mos.	4.1	4.0	NS
12 mos.	4.1	3.9	NS
24 mos.	3.1	3.0	NS

Burkus (2005)/(2006)	**rhBMP-2 (n = 79)**	**ICBG (n = 52)**	
1.5 mos.	NR	NR	
3 mos.	NR	NR	
6 mos.	9.2	7.7	.006
12 mos.	9.2	7.3	.007
24 mos.	8.6	7.1	.032

Haid (2004)	**rhBMP-2 (n = 34)**	**ICBG (n = 33)**	
1.5 mos.	NR	NR	
3 mos.	NR	NR	
6 mos.	NR	NR	
12 mos.	NR	NR	
24 mos.	9	4.5	.009

Leg pain
Range: 0 – 20

Mean score improvement (from baseline) (points)

Boden (2002)	**rhBMP-2 (n = 22)***	**ICBG (n = 5)**	
1.5 mos.	~5.5	~7	NS
3 mos.	~6.5	~3	NS
6 mos.	~4	~3	NS
17 mos.	~6	~4	NS

Dawson (2009)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
1.5 mos.	NR	NR	
3 mos.	NR	NR	
6 mos.	NR	NR	
12 mos.	NR	NR	
24 mos.	9.3	7.2	NS

Dimar (2009)	**rhBMP-2 (n = 239)**	**ICBG (n = 224)**	
1.5 mos.	NR	NR	
3 mos.	NR	NR	
6 mos.	NR	NR	
12 mos.	NR	NR	
24 mos.	~9.3	~7.2	NS

Glassman (2008)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
1.5 mos.	4.6	4.1	NS
3 mos.	NR	NR	
6 mos.	4.4	4.2	NS
12 mos.	3.8	3.9	NS
24 mos.	3.6	3.1	NS

Burkus	**rhBMP-2 (n = 79)**	**ICBG (n = 52)**

	(2005)/(2006)			
	1.5 mos.	NR	NR	
	3 mos.	NR	NR	
	6 mos.	7.7	7.3	.043
	12 mos.	7.5	6.2	.011
	24 mos.	6.8	4.9	.011
	Haid (2004)	**rhBMP-2 (n = 34)**	**ICBG (n = 33)**	
	1.5 mos.	NR	NR	
	3 mos.	NR	NR	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	7.7	6.5	NS
SF-36 Physical component subscale Range: 0 – 100		**Mean score improvement (from baseline) (points)**		
	Boden (2002)	**rhBMP-2 (n = 22)***	**ICBG (n = 5)**	
	6 mos.	~8	~2	NR
	17 mos.	~10	~17	NR
	Dawson (2009)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
	6 mos.	NR	NR	
	12 mos.	NR	NR	
	24 mos.	13	10	NS
	Dimar (2009)	**rhBMP-2 (n = 239)**	**ICBG (n = 224)**	
	6 mos.	~13	~9	
	12 mos.	~13	~10	
	24 mos.	~13	~10	NS
	Glassman (2008)	**rhBMP-2 (n = 25)**	**ICBG (n = 21)**	
	6 mos.	8	9	NS
	12 mos.	10	10	NS
	24 mos.	7	7	NS
	Burkus (2005)/(2006)	**rhBMP-2 (n = 79)**	**ICBG (n = 52)**	
	6 mos.	14	9	.001
	12 mos.	16	11	.003
	24 mos.	15	12	.015
	Haid (2004)	**rhBMP-2 (n = 34)**	**ICBG (n = 33)**	
	6 mos.	~12	~6	NR
	12 mos.	~14	~6	NR
	24 mos.	~14	~11	NR

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation
* Boden (2002): results pooled for both rhBMP-2 groups (i.e., with (n = 11) or without (n = 11) instrumentation).
† Dawson (2009) considered ODI success to be more than 20% improvement from the baseline score.

EFFECTIVENESS
Summary

Studies:
Eight cohort studies met our inclusion criteria. Of these eight studies, there were two prospective cohort studies, one prospective case control study, three retrospective cohort studies, and two retrospective cohort studies with historical controls. One prospective cohort study received an LoE grade of II, while the remaining seven studies were graded LoE III. Study size ranged from 36 to 126 patients. Patients were followed for a mean of 9 to 39 months. Six of the cohort studies included patients with DDD, radiculitis; some of these studies also include those with up to grade 1 or 2 spondylolisthesis, scoliosis, instability, nonunion, or adjacent segment degeneration. One study treated patients with symptomatic pseudarthrosis following previous PLIF for DDD; one study evaluated patients with scoliosis with degeneration distal to a prior long idiopathic scoliosis fusion site. Patients underwent primary or revision single- or multi-level anterior (two studies), posterior (two studies), transforaminal (one study), or posterolateral (three studies) lumbar fusion with rhBMP-2 or iliac crest bone autograft (ICBG), allograft chips, or local or rib autograft. Doses of rhBMP-2 varied and ranged from 3 to 36 mg per patient (when reported). Due to heterogeneity in control treatment, patient diagnosis, and surgical procedures (i.e., approach, use of local autograft or ICBG or bone graft extenders, use of cage versus allograft dowel versus no device, single- versus multilevel design, primary versus revision surgery), we were not able to pool outcomes data. Study sponsorship was reported as follows: Medtronic (1 study), Medtronic and Norton Healthcare grants (1 study), no funding (2 studies), no direct funding but benefits may have been received (1 study), or funding not reported (3 studies). Additional details may be found in Table 19 and the surrounding text.

Outcomes:
Perioperative outcomes (Table 20):
Operative time: There were no differences in mean operative time between groups as reported by one study (N = 64). The strength of this evidence is *insufficient*.
Blood loss: Blood loss was lower in the rhBMP-2 group compared with the autograft group as reported by one study (N = 64) (mean blood loss of 1221 versus 1938 mL, respectively). The strength of this evidence is *insufficient*.
Length of hospital stay: There was no evidence on length of hospital stay.

Fusion (Table 21): Fusion outcomes were similar between rhBMP-2 and control groups at final follow-up (7 studies), while one study reported that outcomes were improved with rhBMP-2. The strength of this evidence is *low*. Specifically, the proportion of patients with successful fusion was similar between rhBMP-2 and autograft groups by final follow-up in five of six studies reporting. One study (N = 50) reported that more patients had successful fusion following PLF with rhBMP-2 versus ICBG (94% versus 77%), however the control group only had 11 patients. One study (N = 75) reported that fusion rates were higher following ALIF with rhBMP-2 compared with allograft chips (99% versus 82% at 24 months).

ODI (Table 22): There were no clinically meaningful differences in mean ODI score improvement at final follow-up (24 to 61 months) as reported by two studies (N = 64 – 75). The strength of this evidence is *insufficient*. The differences in mean score improvement ranged from 3 points (favoring rhBMP-2 over allograft chips) to 3 points (favoring autograft).

Pain (Table 22): There were no differences between groups in various reported pain outcomes (back pain VAS, leg pain VAS, unspecified pain VAS, Prolo Scale Pain Subscale, and the SRS-30 Pain Subscale) at final follow-up as reported by five studies. The strength of this evidence is *low*.

Function (Table 22): There were no differences between treatment groups in functional improvement as reported by two cohort studies. One study (N = 44) reported function using the Prolo Scale Functional Subscale, while the other (N = 64) utilized the SRS-30 Functional Subscale. The strength of this evidence is *insufficient*.

Patient satisfaction (Table 22): Patient satisfaction was similar between treatment groups as reported by two studies (N = 64 – 75). The strength of this evidence is *insufficient*.

Medication use (Table 22): Medication usage was similar in both treatment groups at a mean of 8 to 11 months as reported by one study (N = 44). The strength of this evidence is *insufficient*.

Mental health/self image (Table 22): Both mental health and self-image were similar between treatment groups at a mean of 40 to 61 months follow-up as reported by one study (N = 64). The strength of this evidence is *insufficient*.

Overall outcome (Table 22): There were no differences in the percent of patients between treatment groups who considered themselves to have "good" or "excellent" outcomes as reported by one study (N = 75). The strength of this evidence is *insufficient*.

Detailed study characteristics:

We identified eight cohort studies evaluated off-label use of rhBMP-2 in the lumbar spine. Detailed data abstraction tables can be found in Appendix F. The study characteristics are summarized in Table 19, and are organized in order of the surgical approach used.

Table 19. Off-label use of rhBMP-2 in the lumbar spine: cohort study overview

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Study type
Pradhan (2006)[30]	InFUSE (n = 9) (dose NR) ICBG (n = 27)	FRA	ALIF 1-level/pt primary	DDD, radiculitis	mean 26 mos. (23-29) mean 36 mos. (29-55) 100% (36/36)	III	None	Prospective cohort
Slosar (2007)[32]	InFUSE (n = 45) (3-9 mg/pt) Allograft chips (n = 30)	FRA; instrum.	ALIF mean 2.2 (1-3) levels/pt primary	DDD, grade ≤ 2 spondylo.	24 mos. 96% (72/75)	II	Medtronic	Prospective cohort
Taghavi (2010)[15]	InFUSE (n = 24) (12 mg/pt) BMAA (n = 18) Autograft (details NR) (n = 20)	local autograft extender; instrum.	PLIF mean 2.0 (1-7) levels/pt revision	Symptom-atic pseudo-arthrosis following previous PLIF for DDD	mean 27.9 mos. (24-NR)	III	None	Retrospective cohort
Crawford (2010)[27]	rhBMP-2 (n = 39) (dose NR) Autograft (ICBG, rib, or local) (n = 25)	instrum.	PLIF extension of prior fusion to sacrum	Scoliosis with degener-ation distal to prior long idiopathic scoliosis fusion site	3.3 ± 2.2 yrs. 5.1 ± 1.9 yrs. 94% (60/64)	III	No direct funding but 1+ authors will receive monetary benefits (source NR)	Retrospective cohort with historical control
Mumma-neni (2004)[13]	InFUSE (n = 25) (8.4 mg/pt)	ICBG (1/2 of group) or local autograft (1/2 of group), interbody cage, instrum.	TLIF mean 1.05 (1-2) levels/pt primary	DDD, grade ≤ 1 spondylo., radiculitis	mean 9 mos. (3-18) 91% (40/44)	III	NR	Retrospective cohort

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Study type
	ICBG (n = 19)	Interbody spacer, instrum.						
Singh (2006)[31]	InFUSE (n = 39) (12-36 mg/pt)	ICBG + local autograft	PLF mean NR (1-3) levels/pt primary	DDD, grade ≤ 2 spondylo., radiculitis	24 mos. 96% (48/50)	III	NR	Prospective case-control
	ICBG (n = 11)	local autograft						
Lee (2010)[29]	rhBMP-2 age ≥ 65 yrs (n = 34) (4.2-12 mg/pt) (~4.2mg/lvl)	Allograft; instrum.	PLF mean NR (1-3+ levels/pt) NR	DDD	38.3 ± 7.4 mos. (24-68)	III	NR	Retrospective cohort
	rhBMP-2 age < 65 yrs (n = 52) (4.2-12 mg/pt) (~4.2mg/lvl				39.2 ± 11.7 mos. (24-62)			
	ICBG age ≥ 65 yrs (n = 41)	instrum.			34.7 ± 8.7 mos. (24-58)			
					% f/u NR			

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Study type
Glassman (2007)[28]	InFUSE (n = 91) (12 mg/pt)	bone graft extenders, instrum.	PLF mean 1.41 (1-4) levels/pt primary 82% (75/91) or revision 18% (16/91)	DDD, spondylo., scoliosis, instabil., nonunion, stenosis, adj. segment degen.	mean 27 mos. (24-38) % f/u NR	III	Medtronic*, Norton Healthcare (grants)	Retrospective cohort (historical control)
	ICBG (n = 35)	instrum.	PLF 1-level/pt primary					

ACS: absorbable collagen sponge; ALIF: anterior lumbar interbody fusion; BMAA: bone marrow aspirates + allograft; CRM: compression resistant matrix (i.e., ceramic granules: mixture of hydroxyapatite + tricalcium phosphate, percentage varies by study); DDD: degenerative disc disease; FRA: femoral ring allograft; f/u: follow-up; InFUSE: rhBMP-2 applied to an absorbable collagen sponge (ACS) (Medtronic); NR: not reported; PLIF: posterior lumbar interbody fusion; PLF: posterolateral lumbar fusion; spondylo.: spondylolisthesis; TLIF: transforaminal lumbar interbody fusion

According to Carragee et al. (2011)[47], the authors had a financial relationship with Medtronic ranging from:
* $12,137,000-$16,116,000 (Dimar but not Glassman) plus an additional $1,500,000 for Burkus

Study overview

- Control groups:
  - ICBG (5 studies)
  - Allograft chips (1 study)
  - Autograft and/or bone marrow aspirates plus allograft (2 studies)

- Surgical approach:
  - ALIF (2 studies)
  - PLIF (2 studies)
  - TLIF (1 study)
  - PLF (3 studies)

- Primary diagnosis:
  - DDD, radiculitis, and for some studies spondylolisthesis, scoliosis, instability, adjacent segment degeneration (6 studies)
  - Symptomatic pseudarthrosis following previous PLIF for DDD (1 study)

- Scoliosis with degeneration distal to the site of a prior long idiopathic scoliosis fusion site (1 study)

- Length of follow-up: range from 9 to 39 months; two studies reported follow-up lengths that differed between treatment groups by 10 months or more.

- Funding source:
  - None (2 studies)
  - No direct funding but indirect benefits from an unreported source (1 study)
  - Medtronic only (1 study)
  - Medtronic and Norton Healthcare grant (1 study)
  - Not reported (3 studies)

- LoE grades:
  - LoE II: 1 study
  - LoE III: 7 studies
  - See Appendix YY for details on LoE ratings

Detailed outcomes

Perioperative outcomes (Table 20)

- Operative time (1 cohort study)
  There was no difference in the mean operative time between the rhBMP-2 and autograft groups (10.8 versus 11.3 hours, respectively) as reported by Crawford (2010)[27].

- Blood loss (1 cohort study)
  Crawford (2010)[27] reported that the average blood loss was statistically lower in the rhBMP-2 group compared with the autograft group (1221 versus 1938 mL, respectively; *P* = .007).

- Length of stay
  No studies reported on length of hospital stay.

Table 20. Effectiveness of off-label use of rhBMP-2 in the lumbar spine: perioperative outcomes

Outcome measure	Author	Results		p-value
Operative time		**Mean ± SD (hours)**		
	Crawford (2010)	**rhBMP-2 (n = 39)**	**Autograft (n = 25)**	
		10.8 ± 2.5	11.3 ± 3.0	NS
Blood loss		**Mean ± SD (mL)**		
	Crawford (2010)	**rhBMP-2 (n = 39)**	**Autograft (n = 25)**	
		1221 ± 903	1938 ± 1190	.007

Length of hospital stay		**Mean ± SD (days)**	
	Crawford (2010)	**rhBMP-2 (n = 39)**	**Autograft (n = 25)**
		NR	NR

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Radiographic outcomes (Table 21)

- Definitions of successful fusion varied by study, but in general encompassed the following criteria as evaluated on plain radiograph or CT:
    - Presence of bridging trabecular bone between the transverse processes, and
    - Absence of motion (definition varied), and
    - No implant loosening, and/or
    - Absence of radiolucent lines through the fusion mass or around more than half of the graft.

- Successful fusion (8 cohort studies)
    - *rhBMP-2 versus ICBG (5 studies)*
      In general, rates of successful fusion were similar in patients treated with rhBMP-2 compared with ICBG (5 studies)[13, 28-31]. Specifically, four studies reported similar fusion rates between the groups[13, 28-30] while one reported statistically better fusion outcomes following PLF with rhBMP-2 (n = 39) compared with ICBG (n = 11)[31].

    - *rhBMP-2 versus autograft (1 study)*
      Crawford et al. (2010) reported similar rates of fusion in scoliosis patients undergoing PLIF extension when treated with either rhBMP-2 (89% (32/36) at a mean of 3.3 ± 2.2 years) compared with autograft (79% (19/24) at a mean of 5.1 ± 1.9 years)[27]. The authors noted that the fusion grade was statistically better in patients treated with rhBMP-2 compared with autograft (1.7 ± 0.9 versus 2.3 ± 0.7, respectively; $P = .021$) (fusion graded on scale from 1 (definite fusion) to 4 (definite nonunion)).

    - *rhBMP-2 versus BMAA or autograft (1 study)*
      Rates of fusion in patients undergoing revision PLIF were statistically better at final follow-up (mean of 28 months) in those who received rhBMP-2 compared with those treated with bone marrow aspirate plus allograft but similar to those treated with autograft as reported by Taghavi et al. (2010)[15]. Patients treated with rhBMP-2 or autograft had a 100% (24/24 and 20/20, respectively) fusion rate compared with 78% (14/18) in those who received bone marrow aspirates plus allograft; the differences between the rhBMP-2/autograft and BMAA groups was statistically meaningful ($P = .01$).

- *rhBMP-2 versus allograft chips (1 study)*
  Slosar et al. (2007) reported that patients who underwent ALIF with rhBMP-2 had statistically higher fusion rates at 6, 12, and 24 months compared with those who received allograft chips[32]. Fusion rates at 24 months were 99% (44/45) versus 82% (25/30), respectively.

- Time to fusion (3 studies)
  - *rhBMP-2 versus ICBG (2 studies)*
    - Time to fusion was similar or shorter following surgery with rhBMP-2 compared with ICBG as reported by two retrospective cohort studies[13, 29].
    - Mummaneni et al. (2004) reported that patients undergoing primary TLIF with rhBMP-2 (n = 25) had shorter time to radiographically fusion compared with patients treated with ICBG (n = 19) (3.6 ± 2.0 versus 6.4 ± 2.4 months, respectively), though whether this result was statistically meaningful was not reported[13].
    - Lee et al. (2010) reported similar time to both noticed and solid fusion between treatment groups following PLF[29].

  - *rhBMP-2 versus BMAA or autograft (1 study)*
    Time to solid fusion was shorter in patients undergoing revision TLIF with rhBMP-2 compared with either BMAA or autograft as reported by Taghavi et al. (2010). Mean time to fusion was 7.2 ± 2.1 months in patients treated with rhBMP-2 compared with 9.8 ± 2.2 or 8.9 ± 2.0 in those who received BMAA or autograft, respectively (*P* = .002 and .03, respectively)[15].

Table 21. Effectiveness of off-label use of rhBMP-2 in the lumbar spine: radiographic outcomes

Outcome measure	Author	Results			p-value
Fusion success		**% (n)**			
(see text for definition)	Pradhan (2006)	**rhBMP-2 (n = 9)**	**ICBG (n = 27)**		
	24 mos.	44% (4)	63% (17)		NR
	Slosar (2007)	**rhBMP-2 (n = 45)**	**Allograft chips (n = 30)**		
	6 mos.	79% (36)	23% (7)		< .001
	12 mos.	96% (43)	73% (22)		< .001
	24 mos.	99% (44)	82% (25)		< .001
	Taghavi (2010)	**rhBMP-2 (n = 24)**	**Autograft (n = 20)**	**BMAA (n = 18)**	
	Mean 28 mos.	100% (24)	100% (20)	78% (14)	NS (vs. autograft) .01 (vs. BMAA)
	Crawford (2010)	**rhBMP-2 (n = 39)**	**Autograft (n = 25)**		
	Mean 40-61 mos.	89% (32)	79% (19)		NS

Mummaneni (2004)	**rhBMP-2 (n = 25)**	**ICBG (n = 19)**		
Mean 8-11 mos.	96% (24)	95% (18)		NR
Singh (2006)	**rhBMP-2 (n = 39)**	**ICBG (n = 11)**		
24 mos.	94% (68)	77% (17)		< .05
Lee (2010)	**rhBMP-2 (n = 34)***	**ICBG (n = 41)**		
Mean 34-39 mos.	82% (28)	78% (32)		NS
Glassman (2007)	**rhBMP-2 (n = 91 pts)**	**ICBG (n = 35)**		
24 mos.	96% (87)	86% (30)		NR
Time to fusion	**Time to successful fusion (months) (mean ± SD)**			
Taghavi (2010)	**rhBMP-2 (n = 24)**	**Autograft (n = 20)**	**BMAA (n = 18)**	
Solid fusion	7.2 ± 2.1	8.9 ± 2.0	9.8 ± 2.2	≤ .03
Mummaneni (2004)	**rhBMP-2 (n = 25)**	**ICBG (n = 19)**		
Radiographically demonstrated fusion	3.6 ± 2.0	6.4 ± 2.4		NR
Lee (2010)	**rhBMP-2 (n = 34)***	**ICBG (n = 41)**		
Noticed fusion	3.2 ± 0.8	3.4 ± 0.8		NS
Solid fusion	8.5 ± 2.5	9.6 ± 2.3		NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation
* Lee (2010): For simplicity, results reported here only for rhBMP patients age ≥ 65 years (n = 34) as the control group consisted of patients ≥ 65 years. Detailed results for both rhBMP-2 groups can be found in the Appendix data tables.

Pain (Table 22)

- ODI (2 cohort studies)
  Mean score improvement
  - *rhBMP-2 versus autograft (1 study)*
    In a study of scoliosis patients undergoing PLIF extension of prior fusion to sacrum, Crawford et al. (2010) reported similar improvement in ODI scores between rhBMP-2 (n = 39) and autograft (n = 25) groups (18.4 versus 22.3, respectively; P = NS)[27].

  - *rhBMP-2 versus allograft chips (1 study)*
    Slosar et al. (2007) reported that patients who underwent primary single- or multilevel ALIF had better improvement in their ODI scores when treated with rhBMP-2 (n = 45) compared with allograft chips (n = 30) at 6 months, but that this difference was not sustained at 12 or 24 months. The mean score improvement at 24 months was 33 in the rhBMP-2 group and 30 in the allograft group[32].

"Success": ODI success was not evaluated by any of the cohort studies.

- Back pain and leg pain (1 cohort study)
    - *rhBMP-2 versus autograft versus BMAA*
      Taghavi et al. (2010) reported that following revision PLIF, patient improvements in both back and leg pain were similar between all three treatment groups at 24 months[15].

- Pain (unspecified) (2 cohort studies)
    - *rhBMP-2 versus ICBG (1 study)*
      Lee et al. (2010) reported similar improvements in unspecified pain levels in the rhBMP-2 (n = 34) and ICBG (n = 41) treatment groups at 24 months following single- or multilevel PLF[29].

    - *rhBMP-2 versus allograft chips (1 study)*
      Slosar et al. (2007) found that following primary single- or multilevel ALIF , those treated with rhBMP-2 (n = 45) had greater improvement in their unspecified pain scores at 6 months versus those who received allograft chips (n = 30), but that this difference was not sustained at 12 or 24 months[32].

- Prolo Scale Pain Subscale (1 cohort study)
    - *rhBMP-2 versus ICBG (1 study)*
      Similarly, Mummaneni et al. (2004) found no difference in final pain scores between groups at a mean of 8 to 11 months following primary 1- or 2-level TLIF[13].

- SRS-30 Pain Subscale (1 cohort study)
    - *rhBMP-2 versus ICBG (1 study)*
      There was no difference in the mean pain subscale score improvement in scoliosis patients treated with rhBMP-2 versus autograft at a mean of 3.3 and 5.1 years, respectively, as reported by Crawford et al. (2010)[27].

Function (Table 22)

- Prolo Scale Functional Subscale (1 cohort study)
    - *rhBMP-2 versus ICBG (1 study)*
      Mummaneni et al. (2004) reported similar final functional subscale scores between treatment groups at a mean of 8 to 11 months following primary 1- or 2-level TLIF[13].

- SRS-30 Functional Subscale (1 cohort study)
    - *rhBMP-2 versus ICBG (1 study)*
      Crawford et al. (2010) found no difference between treatment groups in the mean functional subscale score improvement in scoliosis at 3.3 (rhBMP-2) and 5.1 (autograft) years[27].

Patient satisfaction and quality of life (Table 22)

- Patient satisfaction (2 cohort studies)
    - *rhBMP-2 versus autograft (1 study)*
      There was no difference in the final SRS-30 satisfaction subscale scores between groups as reported by scoliosis patients undergoing PLIF extension of prior fusion in Crawford et al. (2010) (mean follow-up ranged from 4 to 61 months)[27].

    - *rhBMP-2 versus allograft chips (1 study)*
      Slosar et al. (2007) reported similar patient satisfaction following primary single- or multilevel ALIF with rhBMP-2 or allograft chips at 24 months[32].

- Medication usage (1 cohort study)
    - *rhBMP-2 versus ICBG*
      Mummaneni et al. (2004) found no difference in Prolo Scale Medication use subscale scores reported at a mean of 8 to 11 months following primary 1- or 2-level TLIF with either rhBMP-2 or ICBG[13].

Social function and mental health outcomes (Table 22)

- *rhBMP-2 versus autograft (1 study)*
  In a study of scoliosis patients undergoing PLIF extension of prior fusion, Crawford et al. (2010) reported similar outcomes in self-image and mental health between rhBMP-2 and autograft treatment groups at 3.3-5.1 years follow-up as measured by the SRS-30 outcome measure[27].

Overall clinical outcomes (Table 22)

- Kirkaldy-Willis grading criteria (1 cohort study)
    - *rhBMP-2 versus ICBG*
      Similar percentages of patients aged 65 years or older who underwent PLF with rhBMP-2 or ICBG rated their clinical outcomes as "excellent" or "good" (as opposed to "fair" or "poor") at 2 years (85% (29/34) versus 73% (30/41), respectively; $P = .414$) as reported by Lee et al (2010)[29].

Work status (2 RCTs)

- Work status was not reported by any of the eight cohort studies evaluating off-label use of rhBMP-2 in the lumbar spine.

Neurological status

- Neurological status was not reported by any of the eight cohort studies evaluating off-label use of rhBMP-2 in the lumbar spine.

Table 22. Effectiveness of off-label use of rhBMP-2 in the lumbar spine: patient-reported outcomes

Outcome measure	Author	Results			p-value
ODI		**Mean score improvement (from baseline) (points)**			
Range: 0 – 100	Slosar (2007)	**rhBMP-2 (n = 45)**	**Allograft chips (n = 30)**		
	< 6 mos.	NR	NR		
	6 mos.	27	17		< .001
	12 mos.	30	26		NS
	24 mos.	33	30		NS
	Crawford (2010)	**rhBMP-2 (n = 39)**	**Autograft (n = 25)**		
	Mean 40-61 mos.	18.4	22.3		NS
Back pain		**Mean score improvement (from baseline) (points)**			
VAS	Taghavi (2010)	**rhBMP-2 (n = 24)**	**Autograft (n = 20)**	**BMAA (n = 18)**	
Range: 0 – 10	1.5 mos.	4.9	4.4	4.2	NS
	6 mos.	4.5	4.3	4.0	NS
	12 mos.	4.6	4.0	4.0	NS
	24 mos.	4.3	4.0	3.9	NS
Leg pain		**Mean score improvement (from baseline) (points)**			
VAS	Taghavi (2010)	**rhBMP-2 (n = 24)**	**Autograft (n = 20)**	**BMAA (n = 18)**	
Range: 0 – 10	1.5 mos.	5.0	4.7	4.3	NS
	6 mos.	4.5	4.3	4.0	NS
	12 mos.	4.5	4.2	4.1	NS
	24 mos.	4.3	4.1	4.0	NS
Unspecified pain		**Mean score improvement (from baseline) (points)**			
VAS	Slosar (2007)	**rhBMP-2 (n = 45)**	**Allograft chips (n = 30)**		
Range: 0 – 10	< 6 mos.	NR	NR		
	6 mos.	4.2	2.8		< .001
	12 mos.	4.7	4.4		NS
	24 mos.	4.8	4.3		NS
	Lee (2010)	**rhBMP-2 (n = 34)***	**ICBG (n = 41)**		
	< 6 mos.	NR	NR		
	6 mos.	5.0	4.9		NS
	12 mos.	4.4	4.5		NS
	24 mos.	3.7	3.9		NS
Prolo Scale		**Mean score at follow-up (points)**			
Pain subscale	Mummaneni (2004)	**rhBMP-2 (n = 25)**	**ICBG (n = 19)**		
Range: 1 – 5					

	Mean 8-11 mos.	3.8 ± 0.9	4.0 ± 0.7	NR
SRS-30		**Mean score improvement (from baseline) (points)**		
Pain subscale	Crawford (2010)	**rhBMP-2 (n = 39)**	**Autograft (n = 25)**	
Range: 0 – 25	Mean 40-61 mos.	1.0 ± 0.7	1.2 ± 0.9	NS
Prolo Scale		**Mean score at follow-up (points)**		
Functional subscale	Mummaneni (2004)	**rhBMP-2 (n = 25)**	**ICBG (n = 19)**	
Range: 1 – 5	Mean 8-11 mos.	3.8 ± 0.9	4.0 ± 0.7	
SRS-30		**Mean score improvement (from baseline) (points)**		
Function subscale	Crawford (2010)	**rhBMP-2 (n = 39)**	**Autograft (n = 25)**	
Range: 0 – 25	Mean 40-61 mos.	0.6 ± 0.5	0.9 ± 0.8	NS
SRS-30		**Mean score at follow-up (points)**		
Satisfaction	Crawford (2010)	**rhBMP-2 (n = 39)**	**Autograft (n = 25)**	
Range: 0 – 10	Mean 40-61 mos.	4.2 ± 0.9	4.0 ± 0.7	NS
Patient satisfaction		**% (n) satisfied**		
	Slosar (2007)	**rhBMP-2 (n = 45)**	**Allograft chips (n = 30)**	
	24 mos.	86% (39)	79% (24)	NR
Prolo Scale		**Mean score at follow-up (points)**		
Medication use subscale	Mummaneni (2004)	**rhBMP-2 (n = 25)**	**ICBG (n = 19)**	
Range: 1 – 5	Mean 8-11 mos.	3.8 ± 0.9	4.2 ± 0.8	NR
SRS-30		**Mean score improvement (from baseline) (points)**		
Self-image subscale	Crawford (2010)	**rhBMP-2 (n = 39)**	**Autograft (n = 25)**	
Range: 0 – 30	Mean 40-61 mos.	1.0 ± 0.9	0.8 ± 0.7	NS
SRS-30		**Mean score improvement (from baseline) (points)**		
Mental health subscale	Crawford (2010)	**rhBMP-2 (n = 39)**	**Autograft (n = 25)**	
Range: 0 – 25	Mean 40-61 mos.	0.3 ± 0.7	1.5 ± 1.0	NS
Kirkaldy-Willis		**% (n) with "good" or "excellent" outcomes**		
Overall clinical results	Lee (2010)	**rhBMP-2 (n = 34)***	**ICBG (n = 41)**	
	24 months	85% (29)	73% (30)	NS

4.2.4. rhBMP-7 off-label use: lumbar spine

EFFICACY

Summary

Studies: Five RCTs met our inclusion criteria, all of which were graded LoE IIb. Study size ranged from 20 to 293 patients. Patients with degenerative (or in one study, isthmic) spondylolisthesis up to grade 1 (or 2) who had not responded to six months of nonsurgical treatment underwent primary single- level posterior (four studies) or posterolateral (one study) lumbar fusion with either OP-1 (rhBMP-7) or iliac crest bone autograft (ICBG) (four studies) or autograft (1 study). RhBMP-7 was used at a dose of 7 mg per patient. The two studies by Vaccaro et al. were similar in design and length of follow-up; data were pooled from these studies when helpful. The remaining three studies were heterogeneous in design and patient characteristics and thus data from these RCTs were not pooled. The mean length of follow-up ranged from 12 to 54 months. Studies were sponsored as follows: Stryker Biotech (1 RCT); funding received but source not stated (2 RCTs); no direct funding but benefits may have been received (2 RCTs). Additional details are available in Table 23 and in the surrounding text.

Outcomes

Perioperative outcomes (Table 24):

Operative time: Operative time was shorter or similar for patients treated with OP-1 compared with ICBG (3 RCTs). The strength of this evidence is *low*. More specifically, one large RCT (N = 293) reported statistically lower operative time in the OP-1 group than in the ICBG group (2.4 versus 2.7 hours, respectively; *P* = .006). The two smaller RCTs (both with N = 36) reported no difference in mean operative times between groups.

Blood loss: Blood loss was lower or statistically similar for patients who received OP-1 versus ICBG as reported by two RCTs. The strength of this evidence was *low*. One study reported statistically lower blood loss in OP-1 versus ICBG patients (difference of 162 mL) (N = 293), while the other study reported statistically similar volumes of blood loss (difference of 49 mL) (N = 36).

Length of hospital stay: There was no difference in the mean length of hospital stay between treatment groups as reported by three RCTs (N = 36 – 293); though the large RCT did not report data, only that there was no difference between treatment groups (*P* = .529). The strength of this evidence is *high*. The mean length of hospital stay (in the two studies reporting the data) ranged from 3.9 to 10.5 days in the OP-1 group compared with 4.3 to 10.9 days in the ICBG group.

Fusion (Table 25): Overall, there were no differences in fusion success between treatment groups as reported by all five RCTs at 12-48 months follow-up (N = 20 – 293). The strength of this evidence is *high*.

ODI (Table 26): Both the percentage of patients with ODI "success" (defined as improvement from baseline scores by at least 20%) (as reported by two RCTs (N = 36 – 293)) and the mean ODI score improvement from baseline (as reported by three RCTs (N = 20 – 293)) were similar between groups at all reported follow-ups. The strength of this evidence is *high*. At 36 to 48 months, 69 to 74% of patients had ODI "success", compared with 57 to 77% of control patients. Regarding mean ODI score improvement, OP-1 and control treatment groups appeared to have clinically meaningful improvement at 12 to 36 month follow-ups, but the differences of 3 to 7 points between the treatment groups were not clinically meaningful.

Pain (Table 26): There were no differences in back or leg pain outcomes as reported by one study each. The strength of this evidence is *low*. Specifically, similar percentages of patients in both treatment groups had no back pain at 12 months follow-up as reported by one small RCT (N = 20). Another RCT reported no difference in the mean VAS score improvement for leg pain at 36 months follow-up between the OP-1 and ICBG treatment groups (N = 293) (3.2 versus 2.8 (on a 10-point scale), respectively).

SF-36: function (Table 26): The mean improvement in SF-36 physical component subscale scores was reported to be similar in both treatment groups at 36 months by one RCT (N = 293), however, no data were reported. The strength of this evidence is *low*.

Neurological success (Table 27): Neurological success was similar in both OP-1 and ICBG treatment groups at 36 months follow-up (or longer) as reported by one RCT (N = 293). The strength of this evidence is *low*.

Overall success (Table 27): The percentage of patients who achieved the composite measure of overall success was similar in both treatment groups as reported by two RCTs at 36 to 48 months follow-up (N = 36 – 293). The strength of this evidence is *low*. This composite measure required ODI success (improvement by 20%), lack of device-related serious adverse events, and radiographic fusion; the smaller RCT additionally required maintenance or improvement in neurologic fusion. Percentages of patients with overall success ranged from 47-62% (mean 48.4% (88/182) in the OP-1 group and from 33-47% (mean 45.9% (34/74)) in the ICBG group.

Detailed study characteristics

Five randomized controlled trials[10, 16-20, 104] were identified that evaluated off-label use of rhBMP-2 in the lumbar spine. Four studies compared OP-1 with ICBG; one compared OP-1 with autograft. Detailed data abstraction tables can be found in Appendix F. The study characteristics are summarized in Table 23.

Table 23. Off-label use of rhBMP-7 in the lumbar spine: RCT study overview

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Notes
Vaccaro (2004/ 2005/ 2008[16-18]	OP-1 Putty (n = 24) (7 mg/pt) ICBG (n = 12)		PLF 1 level/pt primary	Degen. grade ≤ 2 spondylo., neurogenic claudication	48 mos. Radio-graphic: 58% (21/36) Clinical: 72% (26/36)	IIb	Stryker Biotech	Pilot study for OP-1
Vaccaro, Lawrence (2008)[19]	OP-1 Putty (n = 207) (7 mg/pt) ICBG (n = 86)		PLF 1 level/pt primary	Degen. grade ≤ 2 spondylo., stenosis, neurogenic claudication	Mean 54 (44–66) mos. 60% (202/ 335)	IIb	No direct support; indirect benefit(s) for author(s) (source NR)	Pivotal study for OP-1
Johnsson (2002)[20]	OP-1/ACS (n = 10) (7 mg/pt) ICBG (n = 10)		PLF 1 level/pt primary	Degen. L5 spondylo., max. vertebral slip 50%	12 mos. 100% (20/20)	IIb	Funding received (source NR); indirect benefit(s) for author(s) (source NR)	
Kanayama (2006)[104]	OP-1 Putty (n = 10) (7 mg/pt) Local autograft/ CRM (n = 10)	instrum.	PLF 1 level/pt primary	Degen. grade ≤ 1 spondylo. with stenosis, neurogenic claudication	24 mos. 95% (19/20)	IIb	No direct support; indirect benefit(s) for author(s) (source NR)	
Delawi (2010)[10]	OP-1/local autograft (n = 18) (7 mg/pt) ICBG (n = 18)	instrum.	PLIF 1 level/pt primary	Grade ≤ 2 spondylo., with stenosis, neurogenic claudication	12 mos. 89% (32/36)	IIb	Funding received (source NR)	

CRM: compression resistant matrix (i.e., ceramic granules: mixture of hydroxyapatite + tricalcium phosphate, percentage varies by study); DDD; degenerative disc disease; f/u: follow-up; OP-1: osteogenic protein-1 (i.e., rhBMP-7 + collagen matrix + carboxymethyl cellulose sodium); NR: not reported; PLIF: posterior lumbar interbody fusion; spondylo.: spondylolisthesis

- Vaccaro (2004/2005/2008)[16-18] conducted a small pilot study in which N = 36 patients were randomized 2:1 to undergo primary single-level posterolateral lumbar fusion (PLF) with OP-1 Putty (n = 24) or ICBG (n = 12). All patients had grade I or II degenerative spondylolisthesis between L3 and L5 with symptoms of neurogenic claudication and were unresponsive to at least six months of nonoperative treatment. Complete follow-up was available for 58% (radiographic outcomes) and 72% (clinical outcomes) of patients at four years. The study was funded by Stryker Biotech. The study received a LoE grade of IIb. Patients were randomized using SAS software; no information was provided how allocation concealment was ensured. While there was no mention of following the intention to treat principle, data appear to have been handled in accordance with this method; the authors provided last-observation carried forward data separately. There were similar demographic characteristics and baseline scores between groups with one exception: 29% of patients in the OP-1 group had a positive straight leg tension sign at baseline compared with 0% of patients in the control group and this difference was not controlled for. All outcomes were either patient-reported or evaluated in a blinded manner.

- Vaccaro, Lawrence (2008)[19] enrolled 335 patients with single-level grade I or II degenerative spondylolisthesis (between L3-S1), spinal stenosis, neurogenic claudication, and who had not responded to nonoperative care. Patients underwent primary single-level PLF (n = 295/335) with either OP-1 Putty (n = 207) or ICBG (n = 86). Patients were followed for a mean of 54 (44 to 66) months; 60% (202/335) had complete follow-up data. The authors stated that no corporate or industry funds were received to support the work, though at least one author received or will receive benefits related to the study. This RCT was graded LoE IIb. Patients were randomized using computer software (SAS); both patients and surgeons became aware of their treatment allocation at the time of randomization, which may explain why up to 40 patients (≤ 20 in each group) withdrew prior to treatment. A modified intent-to-treat analysis was performed in which patients who had failed or died were excluded from further analysis. Both groups had similar baseline demographics and scores. All outcomes were either patient-reported or evaluated in a blinded manner.

- Johnsson (2002)[20] randomized 20 patients to undergo primary single-level PLF with either OP-1 (n = 10) or ICBG (n = 10). Patients meeting the inclusion criteria had L5 spondylolisthesis, a vertebral slip of no more than 50%, and back pain that was non-responsive to conservative measures for at least 6 months. All patients (100%) had complete follow-up at one year. While corporate or industry funds were not received to fund the work, one or more authors received or will receive benefits related to the study. This RCT received a LoE grade of IIb. Patients were randomized in blocks of six patients. Patients and the surgeon were blinded until the procedure began, but no information was provided as to how concealment was ensured. While there was no explicit statement that the intention to treat principle was followed, data appear to have been handled in accordance with this method. The study did not adequately control for possible confounding, as the demographics were poorly described. Further, the investigational group was comprised of 30% males compared with 50% males in the

control group; this difference was not controlled for. All outcomes were either patient-reported or evaluated in a blinded manner.

- Kanayama (2006)[104] enrolled 20 patients with L3-L4 or L4-L5 degenerative spondylolisthesis (grade ≤ 1) with spinal stenosis to receive primary single-level instrumented PLF supplemented with either OP-1 Putty (n = 10) or local autograft with ceramic bone substitute (n = 10). Results were presented for 95% of patients at 24 months postoperation. The authors stated that no corporate or industry funds were received to support the work, though at least one author received or will receive benefits related to the study. This RCT was graded LoE IIb. No information was provided regarding random sequence generation or allocation concealment. While there was no explicit statement that the intention to treat principle was followed, data appear to have been handled in accordance with this method. Confounding was not controlled for: patients in the OP-1 group were older that those in the autograft group (70 versus 59 years; $P < .05$), a difference which was not controlled for. All outcomes were either patient-reported or evaluated in a blinded manner.

- Delawi (2010)[10] randomized 36 patients in a 1:1 manner to undergo primary single-level PLIF with local autograft in addition to either OP-1 (n = 18) or ICBG (n = 18) for degenerative or isthmic spondylolisthesis (grade ≤ 2) with spinal stenosis and symptoms or neurogenic claudication or radiculopathy. Patients were followed for 12 months, and complete data were available for 89% of patients. Funding was received to support the work, though the funding source was not specified. This study received a LoE grade of IIb. Patients were randomized using a computer-generated randomization code produced according to the "random permuted block" by an independent researcher using SYSTAT. Surgeons were blinded to the treatment group until during the surgical procedure, however, there was no information regarding how allocation concealment was ensured. The intention to treat principle was not used: one patient in the autograft group received local autograft only (no ICBG) and the patient was excluded from analysis. While there were no statistically meaningful differences between treatment groups at baseline in demographic characteristics or ODI scores, the distribution of which spinal level fused was different between groups ($P = .01$) and was not controlled for. All outcomes were either patient-reported or evaluated in a blinded manner.

Detailed outcomes

Perioperative outcomes (Table 24)

- Operative time (3 RCTs)
  Operative time was shorter or similar for patients treated with OP-1 compared with ICBG as reported by three RCTs[10, 16-19]. One large RCT reported statistically lower operative time in the OP-1 group than in the ICBG group (2.4 versus 2.7 hours, respectively; $P = .006$)[19]. The two smaller RCTs reported no difference in mean operative times between groups[10, 16-18]. The average operative times for patients treated

with OP-1 ranged from 2.3 to 3.0 hours, while that for patients who received ICBG ranged from 2.6 to 3.0 hours.

- Blood loss (2 RCTs)
  Blood loss was lower or statistically similar for patients who received OP-1 versus ICBG as reported by two RCTs[10, 19].

- Length of stay (3 RCTs)
  There was no difference in the length of hospital stay between treatment groups as reported by three RCTs[10, 16-19]; though the large RCT did not report data, only that there was no difference between treatment groups ($P = .529$)[19].

Table 24. Efficacy of off-label use of rhBMP-7 in the lumbar spine: perioperative outcomes

Outcome measure	Author	Results		p-value
Operative time		**Mean ± SD (hours)**		
	Vaccaro (2004/2005/2008)	**OP-1 Putty (n = 24)**	**ICBG (n = 12)**	
		2.3 ± 0.7	2.6 ± 0.5	NS
	Vaccaro, Lawrence (2008)	**OP-1 Putty (n = 207)**	**ICBG (n = 86)**	
		2.4	2.7	.006
	Delawi (2010)	**OP-1/ local autograft (n = 18)**	**ICBG (n = 18)**	
		3.0 ± 1.2	3.0 ± 0.8	NS
Blood loss		**Mean ± SD (mL)**		
	Vaccaro, Lawrence (2008)	**OP-1 Putty (n = 207)**	**ICBG (n = 86)**	
		309	471	<.001
	Delawi (2010)	**OP-1/ local autograft (n = 18)**	**ICBG (n = 18)**	
		422 ± 265	373 ± 301	NS
Length of hospital stay		**Mean ± SD (days)**		
	Vaccaro (2004/2005/2008)	**OP-1 Putty (n = 24)**	**ICBG (n = 12)**	
		3.9 ± 1.7	4.3 ± 2.0	NS
	Vaccaro, Lawrence (2008)	**OP-1 Putty (n = 207)**	**ICBG (n = 86)**	
		NR	NR	NS (.529)
	Delawi (2010)	**OP-1/ local autograft (n = 18)**	**ICBG (n = 18)**	

	10.5 ± 4.9	10.9 ± 6.4	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Radiographic outcomes (Table 25)

- Definitions of successful fusion varied by study, but in general encompassed the following criteria as evaluated on plain radiograph or CT:
  - Presence of bridging trabecular bone between the transverse processes, and in most cases,
  - Absence of motion (≤ 3 mm of translation and < 5º angular motion on flexion-extension views.

- Successful fusion (5 RCTs):
  Overall, there were **no differences** in the rates of fusion success between treatment groups as reported by all five RCTs at 12-48 months follow-up[10, 16-20, 104]. Rates varied by definition of fusion used and are summarized in Table 25.

- **Time to fusion** was not reported by any of the five RCTs[10, 16-20, 104].

Table 25. Efficacy of off-label use of rhBMP-7 in the lumbar spine: radiographic outcomes

Outcome measure	Author	Results		p-value
Fusion success		**% (n)**		
	Vaccaro (2004/2005/2008)	**OP-1 Putty (n = 16/24)**	**ICBG (n = 6/12)**	
	Solid fusion, 48 mos.	69% (11)	50% (3)	NS
	Bridging bone, 48 mos.	81% (13)	50% (3)	NS
	Vaccaro, Lawrence (2008)	**OP-1 Putty (n = 143/207)**	**ICBG (n = 53/86)**	
	Bridging bone, 36+ mos.	75% (107)	77% (41)	NS
	≤ 5° angulation 36+ mos.	69% (99)	68% (36)	NS
	≤ 3 mm translation 36+ mos.	76% (109)	75% (40)	NS
	Johnsson (2002)	**OP-1/ACS (n = 10)**	**ICBG (n = 10)**	
	Bridging bone, 12 mos.	60% (6)	80% (8)	NS
	Kanayama (2006)	**OP-1 Putty (n = 9/10)**	**Local autograft (n = 10)**	
	Radiographic fusion 15 mos.	78% (7)	90% (9)	NS

Surgical evidence of solid fusion 15 mos.	57% (4/7)	78% (7/9)	NS
Delawi (2010)	**OP-1/ local autograft (n = 16/18)**	**ICBG (n = 15/18)**	
Definite fusion 12 mos.	63% (10)	67% (10)	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Pain (Table 26)

- ODI (4 RCTs)
  Mean score improvement in ODI scores were similar between groups at 12 months (2 RCTs)[10, 104] and at a minimum of 36 months (1 RCT)[19]. Vaccaro, Lawrence, et al. (2008) reported that at a minimum of three years, the percentage of mean score improvement from baseline in those who received OP-1 was 52% (86/166) compared with 54% (37/68) in the patients treated with ICBG[19]. The two smaller RCTs reported mean ODI score improvement at 12 months that ranged from 17 to 27 points in the OP-1 group and was 24 points in both studies in the autograft or ICBG group (1 study each)[10, 104].

  "Success": There was no difference between groups in the percentage of patients with ODI "success" (defined as improvement from baseline by at least 20% at 36 to 48 months as reported by two RCTs. Rates of ODI success[16-19] ranged from 69-74% (mean 70.0% (129/185) in the OP-1 group and from 57-77% (mean 74.7% (56/75)) in the ICBG group.

- Back pain (1 RCT)
  There were no differences in the percentages of patients with no back pain (versus minor back pain versus major back pain) between the OP-1 (n = 10) and ICBG (n = 10) treatment groups at 12 months follow-up as reported by one small RCT[20].

- Leg pain (1 RCT)
  Mean score improvement Vaccaro, Lawrence et al. (2008) reported no difference in the mean VAS score improvement for leg pain at 36 months follow-up between the OP-1 (n = 166) and ICBG (n = 68) groups[19].

Function (Table 26)

- SF-36 physical component subscale (1 RCT)
  Vaccaro, Lawrence et al. (2008) reported no difference in the mean SF-36 score improvement at 36 months follow-up between the OP-1 (n = 166) and ICBG (n = 68) groups[19], however no data were reported.

Patient satisfaction, social function, mental health outcomes, and work status were not reported by any of the five RCTs.

Neurological success (Table 27) (1 RCT)

Vaccaro, Lawrence et al. (2008) reported similar rates of neurologic success between OP-1 and ICBG treatment groups at 36 months follow-up (or longer) (84% (139/166) versus 80% (54/68), respectively; $P = 54$)[19]. Neurological success was a composite outcome defined as maintenance or improvement in muscle strength, reflexes, sensation, and straight leg raise.

Overall success (Table 27) (1 RCT)

The two RCTs by Vaccaro et al. (2004/2005/2008; 2008)[16-19] both reported the primary efficacy endpoint to be "overall success". Both groups defined this composite measure as requiring ODI success (improvement by 20%), lack of device-related serious adverse events, and radiographic fusion; Vaccaro, Lawrence et al. (2008) additionally required maintenance or improvement in neurologic fusion. Rates of overall success[16-19] ranged from 47-62% (mean 48.4% (88/182) in the OP-1 group and from 33-47% (mean 45.9% (34/74)) in the ICBG group.

Table 26. Efficacy of off-label use of rhBMP-7 in the lumbar spine: patient-reported outcomes

Outcome measure	Author	Results		p-value
ODI		**Mean score improvement (from baseline) (% of points)**		
Range: 0 – 100	Vaccaro, Lawrence (2008)	**OP-1 Putty (n =166/ 207)**	**ICBG (n = 68/86)**	
	< 36 mos.	NR	NR	
	36+ mos.	52%	54%	NS
		Mean score improvement (from baseline) (points)		
	Kanayama (2006)	**OP-1 Putty (n = 9/10)**	**Local autograft (n = 10)**	
	1.5 mos.	NR	NR	
	3 mos.	15	17	NS
	6 mos.	23	31	NS
	9 mos.	16	24	NS
	12 mos.	17	24	NS
	Delawi (2010)	**OP-1/ local autograft (n = 18)**	**ICBG (n = 18)**	
	1.5 mos.	11	6	NS
	3 mos.	27	18	NS
	6 mos.	24	23	NS
	12 mos.	27	24	NS
ODI "success"		**% (n)**		
(≥ 20%) improvement from baseline	Vaccaro (2004/2005/2008)	**OP-1 Putty (n = 19/24)**	**ICBG (n = 7/12)**	

score)				
	< 48 mos.	NR	NR	
	48 mos.	74% (14/19)	57% (4/7)	NS
	Vaccaro, Lawrence (2008)	**OP-1 Putty (n =166/ 207)**	**ICBG (n = 68/86)**	
	36+ mos.	69% (115/166)	77% (52/68)	NS
Back pain		**% (n)**		
Subjective evaluation	Johnsson (2002)	**OP-1/ACS (n = 10)**	**ICBG (n = 10)**	
	Back pain < 12 mos.	NR	NR	
	% with no back pain 12 mos.	40% (4)	50% (5)	NR
	% with minor back pain w/o med. 12 mos.	40% (4)	20% (2)	
	% with major back pain with med. 12 mos.	20% (2)	30% (3)	
Leg pain scores		**Mean score improvement (from baseline) (points)**		
VAS Range: 0 – 10	Vaccaro, Lawrence (2008)	**OP-1 Putty (n =166/ 207)**	**ICBG (n = 68/86)**	
	< 36 mos.	NR	NR	
	36+ mos.	~3.2	~2.8	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Table 27. Efficacy of off-label use of rhBMP-7 in the lumbar spine: physician-reported outcomes

Outcome measure	Author	Results		p-value
Neurological success		**Mean score improvement (from baseline) (% of points)**		
See text for definition	Vaccaro, Lawrence (2008)	**OP-1 Putty (n =166/ 207)**	**ICBG (n = 68/86)**	
	36+ mos. Neuro success composite comprising muscle strength, reflexes, sensation, and straight leg raise	84% (139/166)	80% (54/68)	NS
Overall success		**% (n)**		
See text for definition	Vaccaro (2004/2005/2008)	**OP-1 Putty (n = 16/24)**	**ICBG (n = 6/12)**	
	48 mos.	62% (10/16)	33% (2/6)	NS

Vaccaro, Lawrence (2008)	**OP-1 Putty (n =166/ 207)**	**ICBG (n = 68/86)**	
36+ mos.	47% (78/166)	47% (32/68)	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

EFFECTIVENESS

No studies were identified that evaluated the effectiveness off-label use of rhBMP-7 (OP-1).

4.2.5. BMP (unspecified) use: lumbar spine

EFFICACY

No studies were identified.

EFFECTIVENESS

Summary

One retrospective case control study was identified that utilized data from the longitudinal health insurance claims MarketScan Commercial Claims and Encounters database (Thomson Reuters, Inc.). This study, conducted by Cahill et al. (2011)[138], evaluated perioperative outcomes in n = 2,372 patients who underwent single-level interbody, posterolateral, or circumferential lumbar fusion using BMP (type not specified) and had at least one year follow-up; these patients were matched to n = 2,372 patients who underwent fusion without BMP. Patients undergoing fusion for spinal cancer, infection, or trauma were excluded. This study received an LoE grade of III. Detailed data abstraction tables can be found in Appendix F. A summary of the study characteristics may be found in Table 28.

Table 28. On- or off-label use of rhBMP-2 or rhBMP-7 in the lumbar spine: cohort study overview

	Treatment	**Device(s)**	**Surgical details**	**Primary diagnosis**	**Length f/u** **% f/u**	**LoE**	**Sponsorship**	**Study type**
Cahill (2011)[138]	BMP (n = 2372) (dose NR) No BMP (n = 2372)	± instrum.	Interbody, circumferential, or postero-lateral 1 level/pt primary or revision	NR (excluded cancer, infection, and trauma)	≥ 12 mos. % f/u NR	III	Harvard Catalyst; Harvard University	Retrospective case control database study

DDD: degenerative disc disease; f/u: follow-up; InFUSE: rhBMP-2 applied to an absorbable collagen sponge (ACS) (Medtronic); NR: not reported
* According to Carragee et al. (2011)[47], the authors had a financial relationship with Medtronic ranging from $22,732,000 - $23,192,000 plus an additional $1,500,000 for Burkus.

Detailed outcomes

Perioperative outcomes (Table 29)

- Length of stay (1 RCT)
  The median length of hospital stay was identical (3 days) in both treatment groups following single-level lumbar fusion[138].

Table 29. Effectiveness of on- or off-label use of rhBMP-2 or rhBMP-7 in the lumbar spine: perioperative outcomes

Outcome measure	Author	Results		p-value
Length of hospital stay		**Median ± SD (days)**		
	Cahill (2011)	**BMP (n = 2372)**	**no BMP (n = 2372)**	
		3	3	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

No other effectiveness results were reported.

4.2.6. rhBMP-2 off-label use: cervical spine

EFFICACY

Summary

Studies:

One RCT was identified that met our inclusion criteria (LoE IIb). Thirty-three patients with cervical disc disease with radiculopathy and/or myelopathy underwent one- or two-level plated anterior cervical discectomy with fusion (ACDF) with either rhBMP-2 (n = 18) or ICBG (n = 15). BMP was used at a dose of 0.6 to 1.2 mg per patient. Patients were followed for 24 months. The study reported no direct funding, but benefits may have been received. Additional details may be found in Table 30 and the surrounding text.

Outcomes

Perioperative outcomes (Table 31): All perioperative outcomes (operative time, blood loss, and length of stay) were similar in both treatment groups. The strength of this evidence is *low*.

Fusion (Table 32): Fusion was identical in both treatment groups at all time points, with 100% of patients achieving fusion success at 6, 12, and 24 months post-ACDF[5]. The strength of this evidence is *low*.

NDI (Table 33): The mean score improvement in the NDI was higher in patients who received rhBMP-2 compared with ICBG at 24 months post-ACDF (53 versus 37 points, respectively). The strength of this evidence is *low*. This score difference of 16 points is likely clinically meaningful in neck pain patients, although we did not find an accepted definition of MCID for NDI in spine fusion patients. There were no differences in NDI score improvements between groups at 6 or 12 months.

Pain (Table 33): Neck pain outcomes were clinically similar in both treatment groups at all reported follow-ups (1.5 to 24 months). At 24 months, 100% of patients in both treatment groups were considered to have neck pain "success" (defined as an improvement in VAS score by at least 3 points from baseline). The strength of this evidence was low. Arm pain was clinically improved in rhBMP-2 patients compared with those who received ICBG at 24 months (VAS scores of 14 versus 8 (on a 20-point scale). The strength of this evidence is *low*.

Function: SF-36 (Table 33): There were no differences in the SF-36 physical component subscale scores between treatment groups at 6, 12, or 24 months following ACDF with rhBMP-2 or ICBG. The strength of this evidence is *low*.

Patient satisfaction (Table 33): Patient satisfaction was similar in both groups at 24 months (92-93%). The strength of this evidence is low.

Mental health: SF-36 (Table 33): Mean score improvement in the SF-36 physical component subscale scores were similar between treatment groups at 6, 12, or 24 months following ACDF with rhBMP-2 or ICBG. The strength of this evidence is *low*.

Neurological status (Table 34): The percentage of patients who achieved neurological success was similar for both treatment groups at 6, 12, and 24 months. The strength of this evidence is *low*. By 24 months, 100% of patients in both groups were considered to have neurological success, which was defined as maintenance or improvement in both motor and sensory function.

Detailed study characteristics

One randomized controlled trial[5] was identified that evaluated off-label use of rhBMP-2 in the cervical spine; the study compared rhBMP-2 with ICBG. Detailed data abstraction tables can be found in Appendix F. The study characteristics are summarized in Table 30.

Table 30. Off-label use of rhBMP-2 in the cervical spine: RCT study overview

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Notes
Baskin (2003)[5]	InFUSE (n = 18) (0.6–1.2 mg/pt) ICBG (n = 15)	Allograft ring; cervical plate	ACDF 1-2 level/pt primary	Degen. cervical disease with radiculo-pathy and/or myelopathy	24 mos. Radio-graphic: 61% (20/33) Clinical: 72% (26/36)	IIb	Funding received (source NR); indirect benefit(s) for author(s) (source NR)	Pilot study

ACDF: anterior cervical discectomy and interbody fusion; DDD; degenerative disc disease; f/u: follow-up; NR: not reported

- Baskin (2003)[5] conducted a small pilot study in which N = 33 patients were randomized 1:1 to undergo primary 1- or 2-level ACDF with InFUSE (n = 18) or ICBG (n = 15). Included patients had one- or two-level cervical disease with radiculopathy and/or myelopathy. Two year follow-up data were available for 89% of patients. The authors received funds to support the work, but did not state the source of the funding. Similarly, one or more of the authors stood to benefit from the study, but the source was not reported. The study received a LoE grade of IIb. No information was provided regarding how randomization was achieved. Neither the surgeon or patients were blinded to the treatment allocation following randomization. While there was no mention of following the intention to treat principle, data appear to have been handled in accordance with this method. There was inadequate controlling for confounding differences between groups: 28% of the BMP group used tobacco compared with 47% of the control patients at

baseline, a difference which was not controlled for. All outcomes were either patient-reported or evaluated in a blinded manner.

Detailed outcomes

Perioperative outcomes (Table 31)

- Operative time (1 RCT)
  Mean operative times were identical in both treatment groups (1.8 hours)[5].

- Blood loss (1 RCT)
  There was no difference in mean blood loss in patients undergoing ACDF with either rhBMP-2 or ICBG (91 versus 123 mL, respectively; *P* = NS)[5].

- Length of stay (1 RCT)
  The mean length of hospital stay was similar following ACDF with rhBMP-2 versus ICBG (1.4 versus 1.1 days, respectively; *P* = NS)[5].

Table 31. Efficacy of off-label use of rhBMP-2 in the cervical spine: perioperative outcomes

Outcome measure	Author	Results		p-value
Operative time		**Mean ± SD (hours)**		
	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
		1.8	1.8	NS
Blood loss		**Mean ± SD (mL)**		
	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
		91	123	NS
Length of hospital stay		**Mean ± SD (days)**		
	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
		1.4	1.1	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Radiographic outcomes (Table 32)

- Successful fusion was defined as:
  - On plain radiograph:
    - Presence of bridging trabecular bone, and
    - Absence of motion (< 4° difference in angular motion between flexion and extension), and
    - Absence of radiolucent lines > 2mm thick around more than half of the graft.
  - CT:
    - Presence of bridging trabecular bone.

- Successful fusion (1 RCT)
  Rates of successful fusion were identical in both treatment groups at all time points, with 100% of patients achieving fusion success at 6, 12, and 24 months post-ACDF[5].

- Time to fusion was not reported.

Table 32. Efficacy of off-label use of rhBMP-2 in the cervical spine: radiographic outcomes

Outcome measure	Author	Results		p-value
Fusion success		**% (n)**		
	Baskin (2003)	**rhBMP-2 (n = 10-15/18)**	**ICBG (n = 10-13/15)**	
	6 mos.	100% (15)	100% (13/13)	NS
	12 mos.	100% (14/14)	100% (12/12)	
	24 mos.	100% (10/10)	100% (10/10)	

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Pain (Table 33)

- NDI (1 RCT)
  The *mean score improvement* in NDI was statistically better in patients who received rhBMP-2 compared with ICBG at 24 months post-ACDF (53 versus 37 points, respectively; $P < .03$) as reported by Baskin et al. (2003)[5]. Based on the results for Key Question 1, this score difference of 16 points is likely clinically meaningful in neck pain patients, although we did not find an accepted definition of MCID for NDI in spine fusion patients. There were no differences in NDI score improvements between groups at 6 or 12 months.

  NDI success rates were not evaluated.

- Neck pain (1 RCT)
  Mean score improvement: The mean score improvement in neck pain VAS pain intensity and frequency scores was similar between treatment groups at all follow-ups[5]. By 24 months, the rhBMP-2 group had a mean score improvement of 13 points compared with 9 points in the ICBG group (range of possible scores: 0–20 points).

  "Success": All patients in both groups had neck pain "success", which was defined as an improvement in neck pain VAS scores by 3 points or more from baseline[5].

- Arm pain (1 RCT)
  The *mean score improvement* in arm pain VAS was greater in patients treated with rhBMP-2 versus ICBG at 24 months (14 versus 8 points, respectively; $P < .03$)[5]. The difference in pain scores is clinically meaningful. There were no differences in NDI score improvements between groups at 6 or 12 months. Arm pain frequency and intensity was evaluated on a scale of 0 to 20.

  Arm pain success was not evaluated.

Function (Table 33)

- SF-36 physical component subscale (1 RCT)
  Mean score improvement: There were no differences in the SF-36 physical component subscale scores between treatment groups at 6, 12, or 24 months following ACDF with rhBMP-2 or ICBG[5].

  "Success": Similarly, the percentage of patients with "successful" SF-36 physical component subscale outcomes at 24 months was similar between treatment groups (92% and 100% for rhBMP-2 and ICBG, respectively; *P* = NS). Success was defined as maintenance or improvement in scores from baseline.

Patient satisfaction (Table 33) (1 RCT)
Patient satisfaction rates were similar in both groups at 24 months (92-93%)[5]. This rate includes patients who were satisfied with their treatment outcome, believed they were helped as much as expected, and would have the same surgery again for the same condition.

Social function and mental health outcomes (Table 33)

- SF-36 mental component subscale (1 RCT)
  Mean score improvement in the SF-36 physical component subscale scores were similar between treatment groups at 6, 12, or 24 months following ACDF with rhBMP-2 or ICBG[5].

  "Success": A similar percentage of patients in each treatment group was considered to have "successful" SF-36 mental component subscale outcomes at 24 months (92% and 72% for rhBMP-2 and ICBG, respectively; *P* = NS)[5]. Success was defined as maintenance or improvement in scores from baseline.

Work status was not reported.

Neurological success (Table 34) (1 RCT)
Rates of neurological success were similar for both treatment groups at 6, 12, and 24 months[5]. By 24 months, 100% of patients in both groups were considered to have neurological success, which was defined as maintenance or improvement in both motor and sensory function.

Table 33. Efficacy of off-label use of rhBMP-2 in the cervical spine: patient-reported outcomes

Outcome measure	Author	Results		p-value
NDI		**Mean score improvement (from baseline) (points)**		
Range: 0 – 100	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
	1.5 mos.	37	33	NS

	3 mos.	39	34	NS
	6 mos.	48	39	NS
	12 mos.	46	41	NS
	24 mos.	53	37	< .03

Neck pain		**Mean score improvement (from baseline) (points)**		
Range: 0 – 20	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
	1.5 mos.	11	7	NS
	3 mos.	11	8	NS
	6 mos.	11	10	NS
	12 mos.	12	9	NS
	24 mos.	13	9	NS

Neck pain "success"		**% (n)**		
(≥ 3-point) improvement from baseline score)	Baskin (2003)	**rhBMP-2 (n = 14/18)**	**ICBG (n = 12/15)**	
	< 24 mos.	NR	NR	
	24 mos.	100% (14)	100% (12)	NS

Arm pain scores		**Mean score improvement (from baseline) (points)**		
Range: 0 – 20	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
	1.5 mos.	14	9	NS
	3 mos.	14	8	NS
	6 mos.	15	10	NS
	12 mos.	14	10	NS
	24 mos.	14	8	< .03

SF-36 Physical component subscale		**Mean score improvement (from baseline) (points)**		
Range: 0 – 100	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
	1.5 mos.	9	7	NS
	3 mos.	13	12	NS
	6 mos.	14	14	NS
	12 mos.	14	16	NS
	24 mos.	17	16	NS

SF-36 Physical component subscale "success"		**% (n)**		
Success = maintenance or improvement	Baskin (2003)	**rhBMP-2 (n = 14/18)**	**ICBG (n = 12/15)**	
	< 24 mos.	NR	NR	
	24 mos.	92% (13)	100% (11)	NS

in scores from baseline				
SF-36 Mental component subscale		**Mean score improvement (from baseline) (points)**		
Range: 0 – 100	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
	6 mos.	22	12	NS
	12 mos.	22	8	NS
	24 mos.	22	7	NS
SF-36 Mental component subscale "success"		**% (n)**		
Success = maintenance or improvement in scores from baseline	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
	24 mos.	92% (17)	75% (11)	NS
Patient satisfaction		**% satisfied (n)**		
See text for definition	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
	24 mos.	92% (17)	93% (14)	NS

Table 34. Efficacy of off-label use of rhBMP-2 in the cervical spine: physician-reported outcomes

Outcome measure	Author	Results		p-value
Neurological "success"		**% (n)**		
See text for definition	Baskin (2003)	**rhBMP-2 (n = 18)**	**ICBG (n = 15)**	
	6 mos.	88% (16)	100% (15)	NS
	12 mos.	100% (18)	93% (14)	NS
	24 mos.	100% (18)	100% (15)	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

EFFECTIVENESS

Summary

Studies:

Five cohort studies met our inclusion criteria, including one prospective cohort study, three retrospective cohort studies, and one retrospective case-control database study. All

were graded LoE III. Study size ranged from 58-775 patients. Two of the cohort studies included patients with DDD; another included patients with DDD, herniated nucleus pulposus, or stenosis. A fourth study treated patients for stenosis, spondylosis, or nonunion from a previous fusion. The fifth study did not report patient diagnoses. Patients underwent primary or revision single- or multi-level anterior (two studies), posterior (two studies) cervical fusion with rhBMP-2 or iliac crest bone autograft (ICBG) (two studies), allograft and demineralized bone matrix (one study), a combination of autograft and/or allograft materials (one study). One study did not report surgical approach or the details of the control treatment (referred to as "non-BMP"). BMP was used a dose that ranged from 0.9 to 12 mg per patient (when reported). Due to heterogeneity in control treatments and surgical procedures (i.e., approach, use of local autograft or ICBG or allograft, single- versus multilevel design, primary versus revision surgery), we were not able to pool outcomes data. Patients were followed for 1 to 36 months. Studies were sponsored as follows: no funding received (1 study); funding received but source not stated (1 study); no direct funding but benefits may have been received (2 studies), and funding not reported (1 study). Additional details are available in Table 35 and in the surrounding text.

Outcomes:

Perioperative outcomes (Table 36):

Operative time: There were no differences in the mean operative time between groups as reported by two studies (N = 66 – 77). The strength of this evidence is *insufficient*.

Blood loss: Mean blood loss was similar in both treatment groups as reported by three studies. The strength of this evidence is *low*.

Length of hospital stay: Length of hospital stay was similar between treatment groups in four (N = 58 – 204) of five studies reporting, while one large study (N = 775) reported a longer postoperative stay for those patients treated with rhBMP-2 (7.2 days) than those who did not receive rhBMP-2 (4.3 days, n = 156). The strength of this evidence is *insufficient*.

Fusion (Table 37): Use of rhBMP-2 was associated with similar or higher rates of fusion as reported by two cohort studies. The strength of this evidence is *insufficient*. One study (N = 58) reported no difference in fusion rates between the rhBMP-2 and allograft groups at up to 24 months follow-up (100% versus 96%, respectively), while the other study (N = 204) reported higher rates of fusion in patients who received rhBMP-2 (100%) versus autograft and/or allograft (87.6%).

ODI (Table 38): ODI outcomes were clinically similar in both treatment groups at all follow-ups as reported by two studies (N = 58 – 66). The strength of this evidence is *insufficient.* At final follow-up (24 to 36 months), the difference in mean score improvement between treatment groups ranged from 1 to 9 points.

Pain (Table 38): Neck pain outcomes were similar in two studies and worse in one study following fusion with rhBMP-2 compared with control. The strength of this evidence is *insufficient*. Specifically, two studies reported differences in neck pain VAS scores between treatment groups that ranged from 0 to 2 points at 24 to 36 months follow-up (N = 58 – 66); these differences are not clinically meaningful. One study reported that more patients in the rhBMP-2 group had persistent neck pain at a mean of 24 months compared to those in the control group (48% versus 23%) (N = 204). Arm pain improvement was similar in both treatment groups at all follow-ups out to 24 to 36 months as reported by two studies (N = 58 – 66). The strength of this evidence is *insufficient*.

There were no differences in Nurick or ASIA scores (1 study), improvement in neurological deficits (1 study), use of narcotic medications (1 study), or patient-reported success (1 study).

Detailed study characteristics:

We identified five cohort studies evaluating off-label use of rhBMP-2 in the cervical spine. Detailed data abstraction tables can be found in Appendix F. The study characteristics are summarized in Table 35, and are organized in order of the surgical approach used.

Table 35. Off-label use of rhBMP-2 in the cervical spine: cohort study overview

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Study type
Buttermann (2008)[11]	InFUSE/ allograft ring (n = 30) (0.9-3.7 mg/pt)	plate	ACDF 1-3 level/pt primary	DDD, HNP, or stenosis	24-36 mos. 100% (66/66)	III	None	Prospective cohort
	ICBG (n = 36)	plate (n = 26) OR no plate (n = 10)						
Vaidya, Carp (2007)[34]	InFUSE (n = 22) (1-3 mg/pt)	Interbody fusion cages; instrum.	ACDF 1- or multi-level/pt primary	DDD	≤24 mos. 79% (46/58)	III	NR	Retrospective cohort
	Allograft/ DBM (n = 36)							

	Treatment	Device(s)	Surgical details	Primary diagnosis	Length f/u % f/u	LoE	Sponsorship	Study type
Crawford (2009)[33]	InFUSE/ local autograft/ (± allograft and/or ceramics) (n = 41) (4.2-12 mg/pt) ICBG (n = 36)	instrum.	posterior 1- or multi-level/pt primary or revision	Stenosis, spondylo-lysis, nonunion	≤ 3 mos. 100% (77/77)	III	No direct support; indirect benefit(s) for author(s) (source NR)	Retrospective cohort
Xu (2011)[35]	rhBMP-2/ ± DBM, local autograft, allograft, and/or hydroxyapatite crystals (n = 48) (dose NR) Non-BMP: Some/all: DBM, local autograft, allograft, and/or hydroxyapatite crystals (n = 156)		posterior 1- or multi-level/pt primary	DDD	24 ± 10 mos. 83% (169/ 204)	III	No direct support; indirect benefit(s) for author(s) (source NR)	Retrospective cohort
Yaremchuk (2010)[36]	BMP (n = 260) (dose NR) Non-BMP (n = 515)		Approach NR levels/pt NR	NR	1 mo. % f/u NR	III	Funding received (source NR)	Retrospective case control (database study)

DBM: demineralized bone matrix; DDD; degenerative disc disease; f/u: follow-up; HNP: herniated nucleus pulposus; NR: not reported; spondylo.: spondylolisthesis

Study overview

- Control groups:
  - ICBG (2 studies)
  - Allograft and demineralized bone matrix (DBM) (1 study)
  - Non-BMP: some or all of the following: DBM, local autograft, allograft, and/or hydroxyapatite crystals (1 study)
  - Non-BMP (details not reported) (1 study)

- Surgical approach:
    - ACDF (2 studies)
    - Posterior (2 studies)
    - Not reported (1 study)

- Primary diagnosis:
    - DDD (2 studies)
    - DDD, herniated nucleus pulposus, or stenosis (1 study)
    - Stenosis, spondylolysis, or nonunion from previous fusion (1 study)
    - Not reported (1 study)

- Length of follow-up: range from 1 to 36 months; one study did not report length of follow-up.

- Funding source:
    - None (1 study)
    - No direct funding but indirect benefits from an unreported source (2 studies)
    - Funding received, source not reported (1 study)
    - Not reported (1 study)

- LoE grades:
    - LoE III: 5 studies
    - See Appendix YY for details on LoE ratings

Detailed outcomes

Perioperative outcomes (Table 36)

- Operative time (2 studies)
    - *rhBMP-2 versus ICBG (2 studies)*
      There was no difference in the mean operative time between groups as reported by two studies; one using the ACDF and the other the posterior approach[11, 33].

- Blood loss (3 studies)
    - *rhBMP-2 versus ICBG (2 studies)*
      Patients receiving either rhBMP-2 versus ICBG had similar mean perioperative blood loss as reported by two studies[11, 33].
    - *rhBMP-2 versus autograft and/or allograft (1 study)*
      There was no difference in the average perioperative blood loss between treatment groups as reported by one study ($P = .45$)[35].

- Length of stay (5 studies)
    - *rhBMP-2 versus ICBG (2 studies)*

No differences were reported in mean length of hospital stay in three studies comparing rhBMP-2 with ICBG[11, 33]. The mean length of stay ranged from 1.3 to 4.2 days and 1.2 to 3.5 days, respectively.

- *rhBMP-2 versus autograft and/or allograft (2 studies)*
  There was no difference in the average length of stay between treatment groups as reported by two studies[34, 35]. Mean length of stay was 2.9 to 6.1 days in the rhBMP-2 groups and 2.3 to 7.4 days in the control groups.
- *rhBMP-2 versus no BMP (1 study)*
  Mean (postoperative) length of stay was statistically longer in patients treated with rhBMP-2 (7.2 days) than those who did not receive rhBMP-2 (4.3 days) (P = .001) as reported by one study[36].

Table 36. Effectiveness of off-label use of rhBMP-2 in the cervical spine: perioperative outcomes

Outcome measure	Author	Results		p-value
Operative time		**Mean ± SD (hours)**		
	Buttermann (2008)	**rhBMP-2 (n = 30)**	**ICBG (n = 36)**	
		1.9 ± 0.4	1.9 ± 0.4	NS
	Crawford (2009)	**rhBMP-2 (n = 41)**	**ICBG (n = 36)**	
		2.8 ± 1.0	2.7 ± 0.9	NS
Blood loss		**Mean ± SD (mL)**		
	Buttermann (2008)	**rhBMP-2 (n = 30)**	**ICBG (n = 36)**	
		65 ± 51	65 ± 84	NS
	Crawford (2009)	**rhBMP-2 (n = 41)**	**ICBG (n = 36)**	
		275 ± 224	337 ± 317	NS
	Xu (2011)	**rhBMP-2 (n = 48)**	**Autograft and/or allograft (n = 156)**	
		500	300	NS
Length of hospital stay		**Mean ± SD (days)**		
	Buttermann (2008)	**rhBMP-2 (n = 30)**	**ICBG (n = 36)**	
		1.3 ± 0.5	1.2 ± 0.4	NS
	Vaidya, Carp (2007)	**rhBMP-2 (n = 22)**	**Allograft (n = 36)**	
		2.9	2.3	NR
	Crawford (2009)	**rhBMP-2 (n = 41)**	**ICBG (n = 36)**	
		4.2 ± 2.6	3.5 ± 1.2	NS

Xu (2011)	**rhBMP-2 (n = 48)**	**Autograft and/or allograft (n = 156)**	
	6.1 ± 4.7	7.4 ± 6.9	NS
Yaremchuk (2010)	**BMP (n = 260)**	**No BMP (n = 515)**	
Total LOS:	8.4 ± 7.3	5.5 ± 4.5	NR
LOS after surgery:	7.2 ± 11.1	4.3 ± 5.2	.001

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Radiographic outcomes (Table 37)

- Vaidya, Carp et al. (2007) assessed bone formation as no new bone, visible new bone, possible fusion, and probable fusion (for the rhBMP-2 group), and for the allograft group, assessed fusion at the graft endplate junction as united, possibly united, and probably united[34].
- Xu et al. (2011) did not provide any details as to how radiographic fusion was assessed[35].

- Successful fusion (2 studies):
    - *rhBMP-2 versus autograft and/or allograft (2 studies)*
      Use of rhBMP-2 was associated with similar or higher rates of fusion as reported by two cohort studies[34, 35]. Vaidya, Carp et al. (2007) reported no difference in fusion rates between the rhBMP-2 (n = 22) and allograft groups (n = 36) (100% versus 96%, respectively)[34], while Xu et al. (2011) reported statistically higher rates of fusion in patients who received rhBMP-2 (100%, n = 48) versus autograft and/or allograft (87.6%, n = 156) ($P = .01$)[35].

- Time to fusion was not evaluated by any of the included cohort studies.

Table 37. Effectiveness of off-label use of rhBMP-2 in the cervical spine: radiographic outcomes

Outcome measure	Author	Results		p-value
Fusion success		**% (n)**		
	Vaidya, Carp (2007)	**rhBMP-2 (n = 22)**	**Allograft (n = 36)**	
	≤ 24 mos.	100% (22)	96% (23)	NR
	Xu (2011)	**rhBMP-2 (n = 48)**	**Autograft and/or allograft (n = 156)**	

> 6 mos.	100% (48)	87.6% (106)	.01

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Pain (Table 38)

- ODI (2 cohort studies)
  Mean score improvement
  - *rhBMP-2 versus ICBG (1 study)*
    Buttermann et al. (2008) reported similar improvements in ODI scores at all follow-ups in patients treated with rhBMP-2 (n = 30) versus ICBG (n = 36)[11]. The mean score improvement was 30 and 31 points, respectively.

  - *rhBMP-2 versus allograft (1 study)*
    Vaidya, Carp et al. (2007) similarly found no difference in the mean improvement in ODI scores at 24 months in the rhBMP-2 (n = 22) and allograft (n = 36) groups (24 versus 33, respectively; *P* = NS)[34].

  "Success": ODI success was not evaluated by any of the cohort studies.

- Neck pain (3 cohort studies)
  Mean score improvement
  - *rhBMP-2 versus ICBG (1 study)*
    Buttermann et al. (2008) reported identical improvements in neck pain VAS (scale, 0-10) scores at two to three years postoperation in patients treated with rhBMP-2 (n = 30) versus ICBG (n = 36)[11]. The mean score improvement at two to three years follow-up was five points in both treatment groups.

  - *rhBMP-2 versus allograft (1 study)*
    There were no differences in the mean improvements in neck pain CAS scores at any reported follow-up as reported by Vaidya, Carp et al. (2007). The mean improvement in 24 month VAS scores was 4 points in the rhBMP-2 group (n = 22) and 6 points in the allograft group (n = 36) group[34].

  Patients with neck pain
  - *rhBMP-2 versus autograft and/or allograft (1 study)*
    Xu et al. (2011) reported that statistically more patients in the rhBMP-2 group (n = 48) were experiencing recurrent neck pain at the final follow-up visit (mean, 24 ± 10 months) compared to those in the control group (n = 156) (48% versus 23%, respectively; *P* = .003)[35].

- Arm pain (2 cohort studies)
  Mean score improvement
  - *rhBMP-2 versus ICBG (1 study)*
    Patients in both treatment groups had similar improvements in neck pain VAS (scale, 0-10) scores at two to three years follow-up in patients[11]. The

mean score improvement at two to three years follow-up was 5.5 and 4.8 points in the rhBMP-2 and ICBG treatment groups, respectively.

- *rhBMP-2 versus allograft (1 study)*
  Patients in the rhBMP-2 group had similar improvements in VAS arm pain scores at all follow-ups as those in the allograft group as reported by Vaidya, Carp et al. (2007)[34].

Function (Table 38)

- Nurick score (1 cohort study)
  - *rhBMP-2 versus autograft and/or allograft (1 study)*
    There was no difference in mean Nurick score improvement in the rhBMP-2 and control groups at final follow-up (mean, 24 ± 10 months) as reported by Xu et al. (2011)[35].

- ASIA muscle grading score (1 cohort study)
  - *rhBMP-2 versus autograft and/or allograft (1 study)*
    Xu et al. (2011) similarly reported that patients in both treatment groups had similar improvements in their ASIA muscle grading score at a mean of 24 months follow-up[35].

Patient satisfaction and quality of life (Table 38)

- Patient-reported success (1 cohort study)
  - *rhBMP-2 versus autograft (1 study)*
    A similar percentage of patients in both treatment groups considered themselves to have successful outcomes as reported between one and two years (90% and 94% for rhBMP-2 and ICBG, respectively) as reported by Buttermann et al. (2008)[11].

- Medication usage (1 cohort study)
  - *rhBMP-2 versus autograft and/or allograft (1 study)*
    Xu et al. (2011) reported that a similar percentage of patients in both treatment groups used narcotic medications at all follow-ups[35]. However, the use of narcotics increased with time: at 7 to 12 months, 22-23% of patients used narcotic medications, compared with 43-55% of patients at 25-36 months.

Social function and mental health outcomes

- Social function and mental health outcomes was were reported by any of the five cohort studies evaluating off-label use of rhBMP-2 in the cervical spine.

Work status

- Work status was not reported by any of the five cohort studies evaluating off-label use of rhBMP-2 in the cervical spine.

Neurological status (Table 39)

- *rhBMP-2 versus autograft (1 study)*
  All patients in both treatment groups had resolution of neurological deficits (weakness and/or altered sensation) by 12 months follow-up as reported by Buttermann et al. (2008)[11].

Table 38. Effectiveness of off-label use of rhBMP-2 in the cervical spine: patient-reported outcomes

Outcome measure	Author	Results		p-value
ODI		**Mean score improvement (from baseline) (points)**		
Range: 0 – 100	Buttermann (2008)	**rhBMP-2 (n = 30)**	**ICBG (n = 36)**	
	< 7 mos.	NR	NR	
	7-12 mos.	~14	~11	NS
	13-24 mos.	~25	~17	NS
	25-36 mos.	~30	~31	NS
	Vaidya, Carp (2007)	**rhBMP-2 (n = 22)**	**Allograft (n = 36)**	
	0.5 mos.	3.6	2	NS
	1.5 mos.	6	6	NS
	3 mos.	8	10	NS
	6 mos.	8	21	NS
	12 mos.	14	28	NS
	24 mos.	24	33	NS
Neck pain scores		**Mean score improvement (from baseline) (points)**		
Range: 0 – 10	Buttermann (2008)	**rhBMP-2 (n = 30)**	**ICBG (n = 36)**	
	< 7 mos.	NR	NR	
	7-12 mos.	~4	~4	NS
	13-24 mos.	~4.5	~4	NS
	25-36 mos.	~5	~5	NS
	Vaidya, Carp (2007)	**rhBMP-2 (n = 22)**	**Allograft (n = 36)**	
	0.5 mos.	2	4	NS
	1.5 mos.	2	4	NS
	3 mos.	2	4	NS
	6 mos.	2	4	NS
	12 mos.	3	5	NS
	24 mos.	4	6	NS
Patients with neck pain		**% (n)**		
	Xu (2011)	**rhBMP-2 (n = 48)**	**Autograft and/or allograft (n = 156)**	
	24 ± 10 mos.	48% (19)	23% (31)	.003
Arm pain scores		**Mean score improvement (from baseline) (points)**		
Range: 0 – 10	Buttermann (2008)	**rhBMP-2 (n = 30)**	**ICBG (n = 36)**	
	< 7 mos.	NR	NR	
	7-12 mos.	~3.3	~3.9	NS
	13-24 mos.	~4.2	~3.8	NS

	25-36 mos.	~5.5	~4.8	NS
	Vaidya, Carp (2007)	**rhBMP-2 (n = 22)**	**Allograft (n = 36)**	
	0.5 mos.	1	3	NS
	1.5 mos.	1	4	NS
	3 mos.	2	3	NS
	6 mos.	2	5	NS
	12 mos.	3	5	NS
	24 mos.	4	5	NS
Nurick score		**Mean score improvement (from baseline) (points)**		
Range: 0-5	Xu (2011)	**rhBMP-2 (n = 48)**	**Autograft and/or allograft (n = 156)**	
	24 ± 10 mos.	1.07 ± 0.36	1.17 ± 0.13	NS
Narcotic medication use		**% decrease from baseline (n)**		
	Buttermann (2008)	**rhBMP-2 (n = 30)**	**ICBG (n = 36)**	
	7-12 mos.	23%	22%	NS
	13-24 mos.	30%	42%	NS
	25-36 mos.	43%	55%	NS
Patient-reported success		**% successful (n)**		
	Buttermann (2008)	**rhBMP-2 (n = 30)**	**ICBG (n = 36)**	
	13-24 mos.	90% (27)	94% (24)	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

Table 39. Effectiveness of off-label use of rhBMP-2 in the cervical spine: physician-reported outcomes

Outcome measure	Author	Results		p-value
ASIA score		**Mean score improvement (from baseline) (points)**		
Range: 0-5	Xu (2011)	**rhBMP-2 (n = 48)**	**Autograft and/or allograft (n = 156)**	
	24 ± 10 mos.	0.37 ± 0.12	0.51 ± 0.03	NS
Neurologic deficits		**Resolution of symptoms (from baseline) % (n)**		
	Buttermann (2008)	**rhBMP-2 (n = 30)**	**ICBG (n = 36)**	
	12 mos.	100%	100%	NS

NR: not reported; NS: not statistically significant ($P \geq .05$); SD: standard deviation

4.2.7. rhBMP-7 off-label use: cervical spine

No studies were identified that evaluated on-label use of rhBMP-7 (OP-1).

4.3. Key Question 3: Safety

What is the evidence of the safety of on- or off-label use of rhBMP-2 or rhBMP-7 for spinal fusion compared with spinal fusion using ICBG or alternative bone graft substitutes? Including consideration of:

- Short- and long term adverse events and complications type and frequency (pain, donor site morbidity, resorption/osteolysis, heterotopic bone formation, graft subsidence, graft migration, dysphagia or respiratory difficulties, elevated antibody responses to BMPs or collagen, wound complications (infection, hematoma, seroma, or dehiscence), local or systemic toxicity, mispositioned graft, neurological complications, retrograde ejaculation, urogenital complications, allergic reactions, mortality, other major morbidity).
- Revision/re-operation rates

4.3.1. Overgrowth and uncontrolled bone formation

Summary:
On-label use: The strength of evidence for uncontrolled bone formation is *insufficient* with respect to on-label use of rhBMP-2 compared with autologous bone graft. There were no on-label comparative studies reporting on this outcome.

Off-label use: The risk of uncontrolled bone formation varied widely among three RCTs and four cohort studies assessing off-label use of rhBMP. While the majority of studies reported no cases of uncontrolled bone formation in either the rhBMP or control groups, one RCT identified an incidence of 75% in the rhBMP group compared with 13% in the control group two years after treatment. Some of the differences among studies may be a result of whether the studies assessed uncontrolled bone formation using standard radiography or computerized tomography (CT). Due to the high variability in the results of this outcome, the strength of evidence for these estimates is *low*.

Data are summarized in Table 40.

BMP off-label use: lumbar spine

- RCTs
  - *rhBMP-2 versus ICBG (2 studies)*

The mean rates per study of uncontrolled bone formation at 24 months varied from 0-75% in the rhBMP-2 group compared with 0-13% of patients in the ICBG group as reported by two RCTs[8, 9]. The larger RCT, performed by Dimar et al. (2009), reported no cases of heterotopic ossification in surrounding soft tissue in patients who underwent single level primary posterolateral lumbar fusion (PLF) with either rhBMP-2 AMPLIFY matrix (40 mg/patient) (n = 239) or ICBG (n = 224)[8]. In contrast, a small RCT by Haid (2004) reported new bone formation outside the disc space and into the spinal canal or neuroforamina in 75% (24/34) of rhBMP-2 (4.2-8.4 mg/pt) patients compared with 12% (4/33) ICBG patients following posterior lumbar interbody fusion (PLIF) using thin-cut 1 mm CT scans and plain radiographs[9]. This difference was statistically meaningful ($P < .0001$). The bone formation was not associated with increased leg pain. Both studies were funded by Medtronic.

- *rhBMP-7 versus ICBG (1 study)*
  Vaccaro et al. (2004/2005/2008) reported no instances of heterotopic ossification. This small pilot study treated patients with primary single-level posterolateral lumbar fusion (PLF) with OP-1 Putty (7 mg/pt) (n = 16) or ICBG (n = 6) and was funded by Stryker Biotech[16-18].

- Cohort studies
  - *rhBMP-2 versus control (4 studies)*
    Risks of heterotopic or ectopic bone formation were low in the rhBMP-2 (3.7% (7/191)) and control groups (1% (1/99)) as reported by four cohort studies between 8 and 48 months[14, 27, 32, 55]. Control groups included ICBG, local autograft, autograft (ICBG, rib, or local), and allograft chips. Two patients had associated radiculopathy, one of which underwent reoperation.

    Joseph et al. (2007) conducted a small prospective cohort study designed to examine heterotopic bone ossification following 1- or 2- level minimal access PLIF or TLIF (transforaminal lumbar interbody fusion) with rhBMP-2 (4.2 mg/level) (n = 23; 24 levels) or local autograft (n = 10; 12 levels)[55]. Radiographs were obtained at a mean of 7.9 (range, 6 to 16) months. Risks of heterotopic bone formation were 21% (5/24) in the rhBMP-2 group and 8% (1/12) in the control group ($P = .64$). There were no adverse clinical outcomes associated with heterotopic bone formation. Rihn et al (2009) reported ectopic bone formation in 2% (2/86) of rhBMP-2 patients compared with 0% (0/33) of ICBG patients at a mean of 28 months following primary or revision single-level TLIF[14]. The ectopic bone formation was in the neuroformamen, and both patients had postoperative radiculitis. One patient underwent reoperation. Both Crawford (2010) and Slosar (2007) reported no cases of heterotopic or ectopic bone formation[27, 32].
- Case series
  - *rhBMP-2 (4 case series)*

Extradiscal bone formation following rhBMP-2 use occurred in a mean of 1.8% of 169 patients (range, 0 to 7% of patients) at 7 to 36 months follow-up as reported by four case series[139-142].

- *rhBMP-7 (2 case series)*
  Following rhBMP-7 use, there were no cases of extradiscal bone formation at 24 months in 86 patients as reported by two case series[16, 143, 144].

- Case reports
  - *rhBMP-2*
    There were nine cases of heterotopic bone formation as reported by five case reports[62, 145-148].
  - *rhBMP-7*
    One case report was identified in which an ectopic bone mass was removed at ten months postoperation[149].

rhBMP off-label use: cervical spine

- RCTs
  - *rhBMP-2 versus ICBG (1 study)*
    Baskin et al. (2003) conducted a small RCT in which 33 patients were randomized to undergo primary 1- or 2-level ACDF with InFUSE (0.6-1.2 mg/pt) (n = 18) or ICBG (n = 15)[5]. Bone formation anterior to adjacent segments was identified at 12 months in 11% (2/18) of BMP-2 patients and in 7% (1/15) of ICBG patients.
- Case series
  - *rhBMP-2*
    The mean risk of extradiscal bone formation following rhBMP-2 use in the cervical spine was 7.8% (28/236) of patients (range, 2-68%) as reported by two case series[150-152].
  - *rhBMP-7*
    One case series reported that 7% of 14 patients had extradiscal bone formation following rhBMP-7 use in the cervical spine[143].

Table 40. Extradiscal, ectopic, or heterotopic bone formation

BMP off-label use: lumbar spine	**Studies**	**Mean length follow-up**	**Patients (n)**	**Mean % (n)***	**Range of means (%)**
RCTs					
rhBMP-2	2[8, 9]	24 mos.	271	9.9%	0-75%
Control			256	1.6%	0-13%
rhBMP-7	1[16-18]	24 mos.	16	0%	n/a
Control			6	0%	n/a
Cohort studies					

rhBMP-2	4[14, 27, 32, 55]	8-48 mos.	191	3.7%	0-21%
Control			99	1%	0-8%
Case series					
rhBMP-2	4[139-142]	7-36 mos.	169	1.8%	0-7%
rhBMP-7	2[16, 143, 144]	24 mos.	86	0%	0%

BMP off-label use: cervical spine					
	Studies	**Mean length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	1[5]	24 mos.	18	11%	n/a
Control			15	7%	n/a
Case series					
rhBMP-2	2[150-152]	12-17 mos.	236	7.8%	2-68%
rhBMP-7	Furlan 2007 1[143]	24 mos.	14	7%	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients

4.3.2. Osteoclast activity

Summary:

On-label use: The occurrence of resorption, osteolysis, or graft subsidence/migration/loosening occurred infrequently in both treatment groups in the FDA pilot and pivotal RCTs for InFUSE, 1.3% in the rhBMP group and 0.0% in the control groups. The strength of evidence for these estimates is *low*.

Off-label use: Three RCTs consistently reported similar risks of subsidence or migration between the rhBMP-2 and control groups with risks ≤ 6% in each group. One cohort study in patients with a variety of indications used spinal levels as the unit of measure. That study reported a high subsidence risk of 62% of the levels in the rhBMP-2 and 10% of the levels in the control group. The strength of evidence that the risk of resorption, osteolysis or graft subsidence is similar between groups in off-label use is *moderate*.

Data are summarized in Table 41.

rhBMP-2 on-label use: lumbar spine

- RCTs (2 RCTs)
  Two RCTs reported that risks of graft migration, rotation, or subsidence were low in both treatment groups (1.3% (2/154) versus 0% (0/139) for rhBMP-2 versus ICBG, respectively). There were two cases of implant displacement in the rhBMP-2 group, and both required implant removal: one at five days postoperation due to vertebral bone fracture, and the other case at four months[6].

- Case series (1 case series)
  One case series of 277 patients reported no incidences of implant displacement or loosening or subsidence at 6 years follow-up[153]. At two years follow-up, 1.8% (4/222) of patients had implant displacement, one of which required surgery; 3.2% of patients had subsidence (7/222), four of which required surgery.

rhBMP-2 off-label use: lumbar spine

- RCTs (3 RCTs)
  Three RCTs reported low risks of resorption, osteolysis, or graft subsidence/ migration/ loosening at 24 months follow-up: 0.8% versus 1.6% of patients in the rhBMP-2 versus ICBG groups, respectively[8, 9, 12].

  Burkus et al. (2005) reported no cases of graft migration or extrusion in 79 rhBMP-2 and 52 ICBG patients[12]. Dimar et al. (2009) reported a total of four cases of implant displacement and/or loosening: one in the rhBMP-2 group and three in the ICBG group[8]. It was not reported whether there were any clinical sequelae for these patients. Haid et al. (2004) reported graft subsidence occurred in two patients (6%) in each treatment group[9].

- Cohort studies (3 studies)
  Three cohort studies reported low risks of resorption, osteolysis, or graft subsidence/ migration/ loosening at 24 months follow-up: 3.8% versus 0% of patients in the rhBMP-2 versus control groups, respectively [14, 32, 154].

  Rihn et al. (2009) reported vertebral osteolysis in 6% (5/86) patients following rhBMP-2 compared with 0% of 33 ICBG patients following primary or revision single-level TLIF[14]. All cases were reported between one and five months and the patients reported with increased low back pain. Two patients received revision anterior/posterior debridement and reconstruction and long-term intravenous antibiotics after being diagnosed with osteomyelitis, one patient was diagnosed with a nonunion but refused reoperation, and two patients had fusion and resolution of back pain by one year follow-up. Vaidya, Weir et al. (2007) reported early subsidence (>10%) following ALIF or TLIF in 62% of levels treated with rhBMP-2/allograft compared with 10% of those treated with demineralized bone matrix/allograft; the subsidence was detected at 12 months[154]. Mean subsidence in the rhBMP-2 group was 24-27% (range, 13-42%) compared with 12-15% (range, 11-15%) in the control group. The authors found "significant end plate erosion in each rhBMP-2 case" but not any control patient. Slosar et al. (2007) reported no cases of osteolysis or fragmentation of the graft at 24 months[32].

- Case series (6 case series)
  Six case series reported resorption, osteolysis, or graft subsidence/ migration/ loosening occurring in 0% to 50% of patients (mean, 16.1-25.2%) between 4 and 24 months follow-up[56, 57, 59, 140, 155, 156]. One case series reported this in 22% of 50 treated levels[58].

- Case reports

We identified seven cases of osteolysis in two case reports[157, 158].

rhBMP-2 off-label use: cervical spine

- Cohort studies (1 study)
  Vaidya, Weir et al. (2007) reported early lucency/subsidence at 12 months following ACDF in 62% of 18 cervical levels treated with rhBMP-2/allograft compared with 0% of 22 levels treated with demineralized bone matrix/allograft (*P* = .18). The mean subsidence was 53% (range, 40-58%) in the rhBMP-2 group.

- Case series (4 case series)
  Resorption, osteolysis, or graft subsidence/ migration/ loosening was reported in 1-2% of patients by one case series[159](implant dislodgement and/or graft resorption) and in a mean of 23.9% of levels (range, 44-100%) by three case series[57, 58, 151] (including moderate to severe endplate resorption in 57% of levels[151], transient end plate resorption in 100% of levels that resolved between 3 and 6 months in most patients[57, 58], and subsidence/narrowing of disc space in 50% of patients[57]).

Table 41. Resorption or osteolysis; subsidence, migration, or loosening of graft

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
RCTs					
rhBMP-2	2[6, 21]	24 mos.	154	1.3%	0-1.4%
Control			139	0%	0%
Case series					
rhBMP-2	1[153]	60 mos.	277	0%‡	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	3[1, 8, 9, 12]	24 mos.	397	0.8%	0-6%
Control			309	1.6%	0-6%
Cohort studies					
rhBMP-2	3[14, 32, 154]	24-28 mos.	131 (2 studies)	3.8%	0-6% pts (2 studies)
			37 levels (1 study)	62% levels	
Control			66 (2 studies)	0% pts	0% pts (2 studies)
			41 levels (1 study)	10% levels	
Case series					

rhBMP-2	7[56-59, 140, 155, 156]	4-24 mos.	254 (6 studies) 50 levels (1 study)	16.1-25.2% 22% levels	0-50% pts (6 studies)

BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	1[154]	24 mos.	18 levels	33% levels	n/a
Control			22 levels	0% levels	n/a
Case series					
rhBMP-2	4[57, 58, 151, 159]	12-24 mos.; NR by one study	151 (1 study) 117 levels (3 studies)	1-2% 23.9% levels	 44-100% levels (3 studies)

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† FDA SSED for InFUSE: implant displacement/loosening: 1.7% (5/288) vs. 0.7% (1/139); subsidence: 2.4% (7/288) vs. 1.4% (2/139) for rhBMP-2 vs. ICBG, respectively.
‡ Burkus (2009): data reported for 6 years. At 2 years, risks of implant displacement/loosening were 1.8% and of subsidence were 3.2%.

4.3.3. Local safety

Wound infections

Summary:

On-label use: The strength of evidence for risk of superficial wound infection is *insufficient* with respect to on-label use of rhBMP-2 compared with autologous bone graft. There was only one very small pilot study (n=14) reporting on this outcome.

Off-label use: The risk of superficial wound infections (including superficial infection, dehiscence, edema, and superficial hematoma or seroma) was low (<10%) and similar between treatment groups as reported by two RCTs and five cohort studies. The strength of evidence for these results is *moderate*.

All data are summarized in Table 42.

In addition, one large database study by the Scoliosis Research Society (SRS) focusing on complications in the intraoperative and immediate postoperative periods reported similar risks of superficial wound infection in patients receiving fusion with or without rhBMP for degenerative spinal diseases of the posterior cervical, thoracic and lumbar spines together (1.2% and 1.1%, respectively)[160]. The risk for superficial infection with anterior cervical fusion was slightly higher in the fusion group with versus without rhBMP, 0.9% vs. 0.2%

(risk difference, 0.7%; P = 0.007). It is noted that the data reported for this database were submitted by candidate members to the society who were required to report their operative spine cases, and from full active members who were encouraged to report their cases. Whether the experience of the spine surgeons in this study represents the experience of the general spine surgeon is unknown. No information on the SRS registry/database was reported with respect to completeness and quality control.

Table 42. Superficial wound complications (superficial infection, dehiscence, edema, and superficial hematoma or seroma)

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
RCTs					
rhBMP-2	1[21]	Perioperative	11	9% (1/11)	n/a
Control			3	0% (0/3)	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	2[25, 26]	Perioperative	75	3%	2-4%
Control			73	7%	5-8%
Cohort studies					
rhBMP-2	3[27, 29, 32]	Perioperative – 3 mos.	115	0.9%	0-3%
Control			95	2%	0-3%
BMP (unspecified)	2[161, 162]	Perioperative - ≤ 48 mos	15,675	2.05%	2.01-2.4%
Control			37,954	2.22%	2.2-2.22%
Case series					
rhBMP-2	5[56, 163-167]	3-29 mos. (NR by one study)	1457	2.3-3.8%†	0-20%
rhBMP-7	1[143]	24 mos.	30	7%‡	n/a
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	1[33]	≤ 3 mos.	41	5%	n/a
Control			36	3%	n/a
Cohort studies					
rhBMP-2	1[35]	24 mos.	48	2%	n/a
Control			156	5.1%	n/a
BMP (unspecified)	1[161]	Perioperative	2777	1.5%	n/a
Control			27,159	0.8%	n/a
Case series					
rhBMP-2	3[163, 168, 169]	6-40 mos. (NR by 1 study)	165	6.1-12.7%†	0-20%
rhBMP-7	1[143]	24 mos.	30	7%‡	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† Carreon 2008 reported data for cervical, thoracic, and lumbar patients after the first and second surgeries: after the first surgery, wound drainage or hematoma (not requiring surgical intervention) occurred in 9% (9/96) of patients; after the second surgery, this complication occurred in 11% (11/96) of patients.
‡ Furlan 2007: data reported for lumbar and cervical fusions

Infection, seroma, or hematoma (type unspecified)

Summary: Many studies reported infection, seroma, or hematoma but did not specify whether these were superficial or deep infections.

On-label use: The risk of infection, seroma or hematoma not specified as superficial or deep in the FDA trial for InFUSE was similar for the rhBMP and control groups, 12.2% and 11.5%, respectively. The strength of evidence for these estimates is *low*.

Off-label use: The risk of infection, seroma or hematoma not specified as superficial or deep was reported in four RCTs and two cohort studies of off-label use of rhBMP. While the risks varied from study to study (from 0% to 20% depending on the study), they were similar between the rhBMP and control groups. The strength of evidence for these results is *moderate*.

Data are reported in Table 43.

Table 43. Infection, seroma, or hematoma (type unspecified)

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
Cohort studies					
rhBMP-2	1 (FDA SSED)[22]	24 mos.	288	12.2% (5/288)	
Control			139	11.5% (16/139)	
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	2[1, 8, 12]	24 mos. 24 mos.	318	12.3%	0-16.3%
Control			276	16.3%	0-20.1%
rhBMP-7	2[10, 16-18]	Perioperative - 24 mos.	42	7-8†	4-17%
Control			28	7%	6-8%
Cohort studies					
rhBMP-2	1[27]	3 mos.	36	0%	n/a
Control			24	0%	n/a
rhBMP-7	1[3] (FDA SSBP)	NR	228	7%	n/a
Control			98	2%	n/a
Case series					
rhBMP-2	3[67, 141, 170]	24-29 mos. (NR by 1 study)	240	3.0-4.2%	0-5%
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	2[34, 35]	24 mos.	70	9-10%‡	0513%
Control			180	9.4-11.1%	0-13%
Case series					
rhBMP-2	3[150, 152, 159, 171]	17 - ≥ 24 mos. (NR by one study)	478	3.6-4.0%	0-10%

n/a: not applicable; NR: not reported

* Mean calculated as: number of patients with adverse event ÷ total number of patients

† Delawi (2010) reported surgical infection in 6% (1/18) versus 6% (1/16) and hematoma in 11% (2/18) versus 0% (0/16) of patients for the investigational versus control groups, respectively.

‡ Xu (2011) reported infection in 11% (5/48) versus 11% (17/156) and hematoma in 2% (1/48) versus 2% (3/156) for the investigational versus control groups, respectively.

Deep infection or epidural hematoma and/or surgical evacuation

Summary:

On-label use: The strength of evidence for deep infection or epidural hematoma is *insufficient* with respect to on-label use of rhBMP-2 compared with autologous bone graft. There were no on-label comparative studies reporting on this outcome.

Off-label use: There were no differences in the risk of this outcome between rhBMP-2 and control groups with respect to deep infections; the risks across one RCT and four cohort studies were ≤10% in each group. There were no reports of long-term sequelae resulting from deep infection. We did not identify any studies evaluating rhBMP-7 use that reported on deep infection or surgical evacuation. The strength of evidence for these results is *low*.

All data are summarized in Tables 44 and 45.

Detailed results

BMP off-label use: lumbar spine

- RCTs:
  - *rhBMP-2 versus control (1 RCT)*
    Boden et al. (2002) reported two cases of hematoma in the rhBMP-2 group (10%) compared with none in the ICBG group[7]. One rhBMP-2 patient developed an epidural hematoma that was evacuated at postoperative day 5, the patient had residual numbness in both legs. The other patient had a superficial hematoma that required evacuation four days following surgery and had no negative sequelae. Glassman et al. (2008) reported that one patient (2%) in the rhBMP-2 group required surgery for wound infection compared with two patients (4%) in the ICBG group[7, 26]. No other details were reported.

- Cohort studies:
  - *rhBMP-2 versus allograft bone chips (1 study)*
    Slosar et al. reported one case of a posterior deep wound infection in the rhBMP-2 group (2%) that required irrigation, debridement, delayed closure and intravenous antibiotics; no other details were provided. No deep infections were reported in the control group[32].
  - *rhBMP-2 versus autograft (1 study)*
    Crawford et al. reported one case of deep infection in each treatment group following extension of a prior idiopathic scoliosis fusion to the sacrum (3% versus 4% for rhBMP-2 versus autograft, respectively)[27]. No details of these cases were provided.
  - *rhBMP-2 versus ICBG (1 study)*
    Rihn et al. reported the following infections for patients treated with rhBMP-2 (n = 86) versus ICBG (n = 33), respectively: lumbar infection (3.5% versus

6.1%), lumbar hematoma (1% versus 3%), and lumbar seroma (1% versus 0%). Surgery was necessary in 6% (5/86) of patients in the rhBMP-2 group (lumbar hematoma (n = 1), lumbar seroma (n = 1)) compared with 15% (5/33) of those in the control group (lumbar hematoma (n = 1), lumbar wound infection (n = 2), and ICBG donor site infection (n = 1)[14].

- Case series
  - *rhBMP-2 (4 case series)*
    The mean pooled incidence of deep infection or hematoma was less than 3% of 1407 patients as reported in four case series evaluating off-label use of rhBMP-2 in the lumbar spine[163-167].

In addition, one large registry/database study by the Scoliosis Research Society (SRS) focusing on complications in the intraoperative and immediate postoperative periods reported similar risks of deep wound infection in patients receiving fusion with or without rhBMP for degenerative spinal diseases of the posterior cervical, thoracic and lumbar spines together (1.2% and 1.3%, respectively)[160]. The risk for superficial infection with anterior cervical fusion was slightly higher in the fusion group with versus without rhBMP, 1.2% vs. 0.2% (risk difference, 1.0%; P < 0.001). It is noted that the data reported for this database were submitted by candidate members to the society who were required to report their operative spine cases, and from full active members who were encouraged to report their cases. Whether the experience of the spine surgeons in this study represents the experience of the general spine surgeon is unknown. No information on the SRS registry/database was reported with respect to completeness and quality control.

BMP off-label use: cervical spine

- Cohort studies:
  - *rhBMP-2 versus control (1 cohort study)*
    In a retrospective cohort study, Crawford et al. (2009) reported that 10% (4/41) of rhBMP-2 patients developed deep infection compared with 3% (1/36) of ICBG patients following ACDF (*P* = .118). The ICBG patient developed deep infection at the graft harvest site. All patients were treated with irrigation, debridement, and intravenous antibiotics.

    Vaidya, Carp et al. (2007) also reported one wound exploration in the rhBMP-2 group (1/22) for suspected infection early in the postoperative period; no evidence of infection was found[34].

- Case series
  - *rhBMP-2 (1 case series)*
    One case series reported a deep infection risk of 4-6% of 96 patients following rhBMP-2 use in the cervical spine; these patients required multiple debridements[163].

No studies were identified that reported on the incidence of deep infection following rhBMP-7 use in the spine.

Table 44. Deep infection and epidural hematoma

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP off-label use: lumbar spine					
RCTs					
rhBMP-2	1[7]	17 mos	20	10%	n/a
Control			5	0%	n/a
Cohort studies					
rhBMP-2	3[14, 27, 32]	Perioperative – 54 mos.	167	3.0-4.2%†	2-6%
Control			154	1.9-2.6%†	0-9%
Case series					
rhBMP-2	4[163-167]	3-29 mos. (NR by 1 study)	1407	2.27-2.63‡	2.0-6%
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	1[33]	≤ 3 mos.	41	10%	n/a
Control			36	3%§	n/a
Case series					
rhBMP-2	1[163]	NR	96	4-6%‡	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† Rihn 2009 reported lumbar infection in 4% (3/86) versus 6% (2/33), lumbar hematoma in 1% (1/86) versus 3% (1/33), and lumbar seroma in 1% (1/86) versus 0% (0/33) for the investigational versus control groups, respectively.
‡ Carreon 2008 reported data for cervical, thoracic, and lumbar patients after the first and second surgeries: after the first surgery, wound drainage or hematoma (not requiring surgical intervention) occurred in 9% (9/96) of patients; after the second surgery, this complication occurred in 11% (11/96) of patients.
§ One ICBG patient had a deep infection at the graft harvest site.

Table 45. Surgery due to infection (includes graft site infection)

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP off-label use: lumbar spine					
RCTs					
rhBMP-2	2[7, 26]	17 mos	70	4%	2-10%
Control			57	4%	0-4%
Cohort studies					
rhBMP-2	2[14, 32]	Perioperative – 54 mos.	131	4.6%	2-6%
Control			66	8%†	0-15%
Case series					
rhBMP-2	1[163]	NR	96	4-6%‡	n/a
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	2[33, 34]	Perioperative – 24 mos.	63	8%	5-10%
Control			60	2%	0-3%
Case series					
rhBMP-2	1[163]	NR	96	4-6%‡	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† Rihn 2009, Crawford 2009: one patient in the control group underwent surgery due to graft donor site infection.
‡ Carreon 2008 reported data for cervical, thoracic, and lumbar patients after the first and second surgeries: after the first surgery, wound drainage or hematoma (not requiring surgical intervention) occurred in 9% (9/96) of patients; after the second surgery, this complication occurred in 11% (11/96) of patients.

Dysphagia

Summary:

On-label use: The strength of evidence for dysphagia is *insufficient* with respect to on-label use of rhBMP-2 compared with autologous bone graft. One cohort study (FDA summary on InFUSE) reported "respiratory" complications with 1.7% having this complication among those in the rhBMP-2 compared with 8.6% in the control group.

Off-label use: One RCT reported "respiratory" complications in approximately 5 to 6% of patients in both treatment groups (rhBMP-2 and controls). When rhBMP-7 was used for off-label indications in the lumbar spine, two RCTs and the FDA SSPB for rhBMP-7 reported low risks of "respiratory, thoracic, and mediastinal" complications in both treatment groups. One large retrospective database study reported that a similarly low percentage of rhBMP patients experienced dysphagia or hoarseness as the control group following primary or revision fusion in the lumbar spine. The strength of evidence that the risk of dysphagia and respiratory difficulties in off-label use in the lumbar spine is similar between rhBMP and control groups is *moderate.*

In studies that evaluated rhBMP-2 use in cervical spine fusion, risks of dysphagia were consistently higher in the rhBMP group in four cohort studies, pooled risk of 34.9% in rhBMP-2 patients compared with 9.2% in the control patients. Two database studies reported statistically higher risks of dysphagia in patients who underwent cervical fusion with versus without BMP; one of these studies found that the difference in dysphagia risks was statistically meaningful in those who underwent anterior but not posterior cervical fusion. The strength of evidence that the risk of dysphagia is higher with the use of rhBMP versus control in the cervical spine is *moderate*.

All data are summarized in Table 46.

Detailed results

rhBMP-2 on-label use: lumbar spine

- FDA SSED for InFUSE:
  The SSED for InFUSE reported "respiratory" complications in 1.7% of rhBMP-2 compared with 2.9% of ICBG patients[22]. These events did not occur in the perioperative period. No other details were reported.

BMP off-label use: lumbar spine

- RCTs:
  - *rhBMP-2 versus ICBG (1 RCT)*
    The FDA report for AMPLIFY that corresponds to the Dimar RCT reported "respiratory" complications in a similar percentage of patients in the rhBMP-2 and ICBG treatment groups (6.7% versus 5.4%, respectively; *P* = NS) at 24 months[46].

  - *rhBMP-7 versus ICBG (2 RCTs)*
    Risks of respiratory complications were low in both treatment groups as reported by two RCTs[10, 16-18]. Vaccaro et al. (2004/2005/2008) reported no cases of "respiratory, thoracid, and mediastinal" complications at up to four years follow-up[16-18]. Delawi et al. (2010) reported one case of "respiratory" complications at 12 months follow-up in the rhBMP-7 group (6%)[10].

- Cohort studies:
  - *rhBMP-7 versus ICBG (FDA SSPB for OP-1)*
    The SSPB for OP-1 reported "respiratory, thoracid, and mediastinal" complications in 7% of rhBMP-7 compared with 4% of ICBG patients[3]. No other details were reported.

  - *BMP (type unspecified) versus control (1 study)*
    In a retrospective database study of patients in the Nationwide Implant Sample database who underwent a primary or revision fusion in 2006, Cahill et al. (2009) reported that dysphagia or hoarseness occurred in 0.25% (36/13,972) BMP patients compared with 0.21% (49/22,835) of control patients[161]. This result was not statistically meaningful, with an unadjusted odds ratio of 1.20 (95% CI, 0.78, 1.84).

- Case series:
    - *rhBMP-2*
      One case series reported that 0.29% (3/1037) of patients experienced respiratory failure within 3 months following posterolateral fusion with rhBMP-2[164, 165]. No other details were reported.

BMP off-label use: cervical spine

- Cohort studies:
    - *rhBMP-2 versus control (4 studies)*
      The mean risk of dysphagia or neck swelling was considerably higher in the rhBMP-2 groups compared with the control groups as reported by four cohort studies (34.9% (59/169) versus 9.2% (35/381), respectively). Of note, there was a large range of dysphagia or neck swelling risks between studies: for rhBMP-2, the mean risk per study ranged from 11% to 91%, and for the control group it ranged from 3.6% to 75% per study. This large range precluded doing meta-analysis on these data. Three studies reported higher risks of dysphagia or neck swelling following cervical fusion with rhBMP-2 versus control treatment[11, 34, 64]. One study reported no difference between treatment groups[35].

      Buttermann et al. (2008) reported dysphagia in 50% (15/30) of rhBMP-2 patients compared with 14% (5/36) of control patients[11]. Further, the authors reported that the symptoms were more severe in the rhBMP-2 patients than in those in the control group. In the rhBMP-2 group, symptoms occurred at a mean of 4 ± 3 days following surgery and lasted for 21 ± 16 days; duration of symptoms was not reported for the control group. In the rhBMP-2 group, three patients were readmitted to the intensive care unit where they received intravenous steroids; no surgery was required. Evaluation by a doctor for neck swelling/dysphagia occurred in 23% (7/30) versus 8% (3/36) of rhBMP-2 versus ICBG patients; a phone call to a nurse because of these symptoms was made by 33% of patients in the rhBMP-2 group (10/30) compared with 11% (4/36) of those in the control group. Patients underwent primary single- or multilevel ACDF with 0.9-3.7 mg per patient of rhBMP-2 or with ICBG and were followed for 24 to 36 months. Neck swelling and dysphagia in the rhBMP-2 group occurred more frequently in patients who received a 2-level ACDF (10/16) compared with those who received a 1-level (2/4) or 3-level ACDF (3/10).

      Smucker et al. (2006) retrospectively reported any perioperative swelling complications in patients who underwent single- or multilevel instrumented ACDF with rhBMP-2 (dose not reported) (with or without additional allograft or autograft) or control (allograft or autograft)[64]. Patients were followed for six weeks. Perioperative swelling occurred in 28% (19/69) of rhBMP-2 patients compared with 4% (6/165) of patients in the control (allograft or autograft) group ($P < .0001$). After adjusting for potentially

confounding differences between the treatment groups (prior anterior surgery, smoking status, number of levels fused, three or more levels fused, inclusion or proximity to C4-C5, use of a plate; and use of allograft, autograft, or PEEK cage), the authors reported that rhBMP-2 patients were 10.1 times more likely to have swelling complications compared with those in the control group (adjusted odds ratio of 10.1 (95% CI, 3.8, 26.6)). The onset of swelling could be determined in 11 patients, and in these patients the swelling began at a mean of 4.2 (range, 2 to 7) days. Of these patients, a delay in discharge occurred in 13% (n = 9) of the rhBMP-2 patients and 3.0% (n = 5) of the control patients for one or more of the following reasons: visible neck swelling (3% versus 0% of patients, respectively), severe dysphagia (7% versus 1.2%), reintubation (3% versus 0%), PEG placement (1% versus 1%), tracheostomy (1% versus 0.6%), and/or delay in extubation (0% versus 0.6%). Surgical exploration and drainage was required in 4% (n = 3) of rhBMP-2 patients compared with no control patients; the procedures occurred at 4, 5, and 7 days following the initial surgery. None of these patients had experienced acutely compromised breathing. The swelling in these patients was diffuse in the soft tissue. The patients recovered fully without any other related complications. Readmission for treatment of swelling was necessary in 3% (n = 2) of rhBMP-2 patients compared with no control patients, as was outpatient ENT consult. Four percent of rhBMP-2 patients (n = 3) returned to the clinic or ER prematurely, compared with 0.6% (n= 1) of control patients.

Vaidya, Carp et al. (2007) conducted a retrospective cohort study of patients who underwent single- or multilevel primary instrumented ACDF with rhBMP-2 (n = 22) or allograft and demineralized bone matrix (n = 24)[154]. More patients in the rhBMP-2 group had dysphagia at both two and six weeks following surgery compared with the control group (2 weeks: 85% versus 39% of patients, respectively; $P = .00092$) (6 weeks: 65% versus 22%, respectively; $P = .0189$). When stratifying patients according to whether they underwent single- versus two- or three level fusion, differences between treatment groups were statistically meaningful only for patients who underwent two- or three-level fusion at six (but not two or twelve) weeks (6 weeks, multilevel: 92% versus 40%; $P = .023$). The incidence of dysphagia symptoms at two years was similar between treatment groups (20% versus 22% for rhBMP-2 and control). The percentage of patients experiencing hoarseness was similar between treatment groups (60% versus 62% for rhBMP-2 and control, respectively).

Xu et al. (2011) reported no difference in dysphagia risks between patients who received rhBMP-2 and local autograft versus demineralized bone matrix and allograft (6% (3/48) versus 4% (6/156); $P = .48$). Patients underwent primary single- or multilevel posterior cervical fusion (mean of 5.9 ± 1.9 levels per patient).

- *BMP (type unspecified) versus control (2 studies)*
  Cahill et al. (2009) conducted a retrospective study of patients in the Nationwide Implant Sample database who underwent a primary or revision anterior or posterior cervical fusion in 2006. For the patients who underwent ACDF, the authors reported that dysphagia or hoarseness occurred in statistically more patients in the BMP compared with the control group (BMP: 4.35% (100/2299) versus control: 0.21% (2.45% (608/24,768))[161]. This result held true when using either an unadjusted or an adjusted odds ratio (unadjusted OR: 1.80 (95% CI, 1.45, 2.24) (adjusted OR: 1.67 (95% CI 1.30, 2.05). In contrast, patients who underwent posterior cervical fusion had similar risks of dysphagia or hoarseness irrespective of whether they did or did not receive BMP (BMP: 2.1% (10/478) versus control: 1.63% (39/2392)), with an unadjusted odds ratio of 1.28 (95% CI, 0.63, 2.59).

  Yaremchuk et al. (2010) retrospectively reviewed a U.S. hospital claims database. Patients who underwent cervical fusion during a five-year period (2004-2009) with (n = 260) or without (n = 2387) BMP were followed for one month. Dysphagia risks were nearly nine times higher in patients in the BMP versus the control group (6.9% (18/260) versus (3.3% (17/515); *P* = .001), with an adjusted odds ratio of 8.94 (95% CI, 3.63, 21.99)[36]. In addition, respiratory failure rates were over three times higher in the BMP group compared with the risks group (13.1% versus 4.7%; *P* = .001; adjusted OR = 3.35 (95% CI, 1.88, 5.97)). Patients who received BMP versus no BMP also had a higher risk of the following complications within the first postoperative month: tracheotomies (3.1% versus 0.6%; *P* = .024; adjusted OR = 4.87 (95% CI, 1.23-19.23)), unplanned intubations after surgery (6.2% versus 1.6%; *P* = .008; adjusted OR = 3.91 (95% CI, 1.60, 9.54)); readmissions (8.8% versus 5.0%; *P* = 0.40; adjusted OR = 1.96 (1.03, 3.70)); and dypsnea (20.4% versus 8.0%; *P* = .001; adjusted OR = 2.43 (95% CI, 1.53, 3.88)).

- Case series:
  - *rhBMP-2 (7 case series)*
    Seven case series with a total of 508 patients reported neck swelling or dysphagia in a mean of 12.8-15.6% of patients who received rhBMP-2 during cervical spinal fusion (range, 0-100%)[57, 150-152, 159, 168, 169, 172]. Follow-up ranged from 1 to a mean of 40 months.

  - *rhBMP-7 (1 case series)*
    One case series reported that 0.8% (1/131) patients who received rhBMP-7 during cervical fusion developed dysphagia[173]. Patients were followed for one month.

- Case reports:

- *rhBMP-2 (1 case report)*
  We identified one case report[174] in which a patient developed severe neck swelling following ACDF with rhBMP-2; the patient was irrigated and remained on a ventilator for 24 hours. The patient returned home three days postoperatively.

Table 46. Dysphagia, respiratory difficulties and/or neck swelling

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
Cohort studies					
rhBMP-2	1[22] (FDA SSED)	24 mos.	288	1.7%	n/a
Control			139	8.6%	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	1[46]	24 mos.	239	6.7%	n/a
Control			224	5.4%	n/a
rhBMP-7	2[10, 16-18]	12-48) mos.	18 24	6% pts 1 event	n/a
Control			16 12	0% pts 0 events	n/a
Cohort studies					
rhBMP-7	1[3] (FDA SSPB)	NR	228	7%	n/a
Control			98	4%	n/a
BMP (unspecified)	1[161]	Perioperative	13,972	0.25%	n/a
Control			22,835	0.21%	n/a
Case series					
rhBMP-2	1[164, 165]	3 mos.	1037	0.29%	n/a
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	4[11, 34, 35, 64]	0-24 mos.	169	34.9%	11-91%
Control			381	9.2%	3.6-75%
BMP (unspecified)	2[36, 161]	Perioperative-1 mo.	2537	5.04%	3.96-6.9%
Control			27,674	3.06%	0.6-3.11%
Case series					
rhBMP-2	7[57, 150-152, 159, 168, 169, 172]	1-40 mos. (NR by one study)	506	12.8-15.6%	0-100%
rhBMP-7	1[173]	1 mos.	131	0.8%	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients

4.3.4. Neurologic events

Retrograde ejaculation

Summary:

On-label use: The strength of evidence for retrograde ejaculation is *low* regarding on-label use of rhBMP-2 compared with autologous bone graft. The evidence base for on-label use consists of one cohort study (FDA summary on InFUSE) that reported a higher risk of retrograde ejaculation in the rhBMP-2 group, 7.9% versus 1.4%.

Off-label use: One retrospective cohort study gathered data from patients who underwent 1- or 2-level ALIF spanning L5/S1 and identified a 12-fold increase of retrograde ejaculation in those receiving rhBMP-2, 7.2% compared with 0.6%. The strength of evidence in off-label use regarding retrograde ejaculation is *low.*

All data are summarized in Table 47.

Detailed results

rhBMP-2 on-label use: lumbar spine

- FDA SSED for InFUSE:
  The SSED for InFUSE reported "respiratory" complications in 1.7% of rhBMP-2 compared with 2.9% of ICBG patients[22]. These events did not occur in the perioperative period. No other details were reported.

rhBMP-2 on-label use: lumbar spine

- FDA SSED for InFUSE:

  By 24 months, the total reported incidence of retrograde ejaculation (RE) in males was 7.9% (11/140) in the rhBMP-2 group (12 events) compared with 1.4% (1/70) in the ICBG group (1 event), broken down as follows:
  - Perioperative period: 0 versus 0 events
  - Postoperative period (1 day- 4 weeks): 4 versus 1 events
  - 4-9 weeks: 5 versus 1 events
  - 9 weeks – 5 months: 0 versus 0 events
  - 9-19 months: 2 versus 0 events
  - 19- < 30 months: 0 versus 0 events

rhBMP-2 off-label use: lumbar spine

- Cohort studies:
  - Carragee et al. (2011) conducted a retrospective analysis of prospectively gathered data on the incidence of (RE) following one- or two- level anterior lumbar interbody fusion. Patients with degenerative spondlyolisthesis, low-grade isthmic spondylolisthesis, recurrent disc herniation, or presumed discogenic pain involving the L5/S1 level were treated with one- or two-level ALIF via an open retroperitoneal approach with a femoral ring allograft or titanium mesh cage filled with rhBMP-2 (4.2 mg/patient) (n = 69) or control

(osteophytes or ICBG) (n = 174). Instrumentation was used at the discretion of the surgeon. Patients were followed for twelve months. The overall incidence of RE was statistically higher in the rhBMP-2 group compared with the control group (7% (5/69) versus 0.6% (1/174); *P* = .0025). Patients who underwent 1-level fusion (L5/S1) had similar results (7% (3/45) versus 0% (0/110); *P* = .0233), but those who underwent 2-level fusion (L4/L5 and L5/S1) had statistically similar risks of RE between treatment groups (8% (2/24) versus 2% (1/64) for rhBMP-2 versus control; *P* = .179). At one year postoperation, 40% (2/5) of affected patients in the rhBMP-2 group and 100% (1/1) of those in the control group reported resolution of RE symptoms. The authors noted that the two oldest patients (of the six affected) did not have resolution of symptoms. There was no association of RE with patient diagnosis.

No case series or case reports in which patients received rhBMP-2 or rhBMP-7 for spinal fusion reported on retrograde ejaculation.

Table 47. Retrograde ejaculation

BMP on-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	1[22] (FDA SSED)	≤ 24 mos.	140	7.9%	n/a
Control			70	1.4%	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	1[60]	12 mos.	69	7.2%	n/a
Control			174	0.6%	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients

Ileus/ bowel obstruction

Summary:

On-label use: The strength of evidence for ileus/bowel obstruction is *insufficient* with respect to on-label use of rhBMP-2 compared with autologous bone graft. There was only one very small pilot study (n=11) reporting on this outcome.

Off-label use: The strength of evidence for ileus/bowel obstruction is *insufficient* with respect to off-label use of rhBMP-2 compared with autologous bone graft. There was only one retrospective cohort study that reported low risks of ileus in both the rhBMP-2 (1%) and ICBG (3%) treatment groups following primary or revision single-level TLIF.

Table 48. Ileus/ bowel obstruction

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
RCTs					
rhBMP-2	1[21]	Perioperative	11	9%	n/a
Control			3	33%	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	1[14]	mean 27 mos.	86	1%	n/a
Control			33	3%	n/a
Case series					
rhBMP-2	3[56, 164, 165, 167]	3-29 mos.	1291	2.6%	2-2.9%

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients

Urinary retention

Summary:

On-label use: The strength of evidence for urinary retention is *insufficient* with respect to on-label use of rhBMP-2 compared with autologous bone graft. There was only one very small pilot study (n=11) reporting on this outcome.

Off-label use: There were no off-label comparative studies reporting on this outcome.

All data are summarized in Table 49.

Table 49. Urinary retention

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
RCTs					
rhBMP-2	1[21]	Perioperative	11	0%	n/a
Control			3	33%	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Case series					
rhBMP-2	1[56]	12 mos.	50	2%	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients

Episodes of radiculitis

Summary:

On-label

Only one nonrandomized comparative study was found, the FDA SSED for InFUSE, which reported similar risks of radiculitis following ALIF with rhBMP-2 compared with ICBG (23% vs. 22%, respectively). The strength of evidence is *insufficient* regarding these estimates.

Off-label

Two comparative studies evaluating rhBMP-2 in the lumbar spine, one RCT and one cohort study, reported similar low risks of radiculitis in the BMP compared with the control groups (0-2%). For rhBMP-7 use in the lumbar spine, only one RCT was found which reported a lower risk in the BMP compared with the control group (6% vs. 13%). The strength of evidence is *low* regarding these estimates.

All data are summarized in Table 50.

Table 50. Radiculitis

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
Cohort studies					
rhBMP-2	1[22] FDA SSED†	24 mos.	288	22.6% (65/288)*	n/a
Control			139	21.6% (30/139)*	
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	1[26]	24 mos.	50	0% (0/50)	n/a
Control			52	2% (1/52)	n/a
rhBMP-7	1[10]	12 mos.	18	6% (1/18)	n/a
Control			16	13% (2/16)	n/a
Cohort studies					
rhBMP-2	1[27]	3 mos.	36	0% (0/36)	n/a
Control			24	0% (0/24)	n/a
Case series					
rhBMP-2	3[164, 165, 167, 175]	3-30 mos. (NR by one study)	1276	1.33%	0.68-11%
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Case series					
rhBMP-7	1[173]	3 days	131	0.8%	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† FDA SSED for InFUSE: Back and/or leg pain events

Dural injury or CSF leak

Summary:

On-label

Only one nonrandomized comparative study was found, the FDA SSED for InFUSE, which reported similar risks of dural injury or durotomy following ALIF with rhBMP-2 compared with ICBG (0% vs. 0.7%, respectively). The strength of evidence is *insufficient* regarding these estimates.

Off-label

Evidence from three RCTs and seven cohort studies evaluating the use of rhBMP-2 or rhBMP-7 in the lumbar or cervical spine shows similar risks of dural injury or durotomy in the BMP groups compared with the control groups; risks ranges from 2.4%–11% irrespective of treatment group. The strength of the evidence is *high* regarding these between-group comparisons.

All data are summarized in Table 51.

Table 51. Dural injury (or CSF leak where noted†)

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
Cohort studies					
rhBMP-2	1[22] (FDA SSED)	24 mos.	288	0%	n/a
Control			139	0.7%	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	3[8, 9, 25]	Perioperative-24 mos.	298	6.0%	4-9%
Control			278	7.6%	5-8%
Cohort studies					
rhBMP-2	4[14, 15, 29, 32]	Perioperative-40 mos.	189	3.7%	2-5%
Control			124	2.4%	0-7%
rhBMP-2	1[13]†	Perioperative	21	10%†	n/a
Control			19	11%†	n/a
rhBMP-7	1[10]	0 mos.	18	6%	n/a
Control			16	6%	n/a
Case series					
rhBMP-2	3[67, 164, 165, 170]	3-29 mos. (NR by 1 study)	1203	5.2%	2-5.6%
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	1[35]	24 mos	48	0%	n/a
Control			156	2.6%	n/a

CSF: cerebrospinal fluid; n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† CSF leak

Neurological (unspecified or other) adverse events

Summary:

On-label

Only one nonrandomized comparative study was found, the FDA SSED for InFUSE, which reported similar risks of neurological adverse events following ALIF with rhBMP-2

compared with ICBG (12.5% vs. 15.1%, respectively). The strength of evidence is *insufficient* regarding these estimates.

Off-label

Evidence from four RCTs and four cohort studies (to include the FDA SSPB for OP-1) evaluating the use of rhBMP-2 or rhBMP-7 in the lumbar spine shows similar risks of neurological adverse events in the BMP groups compared with the control groups; risks ranged from 4.0%–26.0% irrespective of treatment group. The strength of the evidence is *high* regarding these between-group comparisons.

All data are summarized in Table 52.

In addition, one small cohort study[13] evaluating off-label use of rhBMP-2 in the lumbar spine (Mummaneni 2004) reported similar risks of L-5 paresis between the rhBMP-2 and control groups (5% (1/21) versus 5% (1/19), respectively) (paresis data not included in Table 52).

Table 52. Neurological (unspecified/other) adverse events

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
Cohort studies					
rhBMP-2	1[22] (FDA SSED)	24 mos.	288	12.5%	n/a
Control			139	15.1%	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	3[8, 9, 26]	Perioperative-24 mos.	323	26.0%	0-41%
Control			309	24.3%	2-42%
rhBMP-7	1[16-18]	48 mos.	24	4%	n/a
Control			12	0%	n/a
Cohort studies					
rhBMP-2	3[13, 27, 29]	Perioperative-9 mos.	91	3-4%†,‡	0-6%
Control			84	4-5%†,‡	0-5%
rhBMP-7	1[3] (FDA SSPB)	NR	228	11%	n/a
Control			98	10%	n/a
Case series					
rhBMP-2	2[164, 165, 167]	3 mos mean 29 mos.	1241	0.64%	0.10-3.4%
rhBMP-7	1[176]	mean 5 mos.	9	11%	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† Mummaneni (2004) reported worsening of preoperative partial foot drop in 0/21 rhBMP-2 patients compared with 1/19 control patients and weakness of ankle dorsiflexion in 1/21 rhBMP-2 patients and 0/19 control patients.
‡ Crawford (2010): reported nerve root deficits, all of which were resolved after reoperation.

4.3.5. Antibody responses to BMP

Summary:

On-label

One RCT was found which reported similar low risks of elevated anti-BMP antibodies following the use of rhBMP-2 in the lumbar spine compared with controls (0.7% vs. 0.8%, respectively). The strength of the evidence is *low* regarding these estimates.

Off-label

Four RCTs reported similar low risks of elevated anti-BMP antibodies following the use of rhBMP-2 in the lumbar (3 RCTs) and cervical spine (1 RCT) compared with controls (0–0.7%).

One RCT which evaluated rhBMP-7 use in the lumbar spine reported a higher proportion of patients with elevated anti-BMP antibodies in the BMP group compared with the control group (93.7% vs. 21%). The strength of evidence is *high* regarding the results for rhBMP-2 and *low* for rhBMP-7.

None of the studies reported any negative consequences to elevated or positive antibody responses.

All data are summarized in Tables 53.

Table 53. Elevated anti-BMP antibodies

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
RCTs					
rhBMP-2	1[6]	24 mos.	143	0.7%	n/a
Control			134	0.8%	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	3[1, 7, 9, 12]	17-24 mos.	134	0.7%	0-5%
Control			89	0%	0%
rhBMP-7	1[2, 19]	24 mos.	208	93.7%†	n/a
Control			87	21%†	n/a
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	1[5]	24 mos.	18	0%	n/a
Control			15	0%	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† Vaccaro, Lawrence (2008)/Hwang (2010) reported elevated anti-rhBMP-7 antibodies at any time up to 24 months in 93.7% and 20.9% of rhBMP-7 and control patients, respectively; in addition, 25.6% and 1.2% of patients were positive for anti-rhBMP-7 neutralizing antibodies.

4.3.6. Antibody responses to collagen

Summary:

On-label

One RCT was found which reported similar low risks of anti-bovine collagen or elevated anti-human collagen antibodies following the use of rhBMP-2 in the lumbar spine compared with controls (0.7% vs. 0.8%, respectively). The strength of the evidence is *low* regarding these estimates.

Off-label

Pooled estimates from two RCTs showed similar risks of anti-bovine collagen or elevated anti-human collagen antibodies following the use of rhBMP-2 in the lumbar spine compared with controls (9% vs. 11%, respectively). The strength of the evidence is *low* regarding these estimates.

None of the studies reported any negative consequences to elevated or positive antibody responses.

Data are summarized in Table 54.

Table 54. Anti- bovine collagen or elevated anti- human collagen antibodies

BMP on-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	1[6]	24 mos.	143	0.7%	n/a
Control			134	0.8%	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	2[1, 9, 12]	24 mos.	113	9%	9%
Control			82	11%	8-15%

n/a: not applicable; NR: not reported

* Mean calculated as: number of patients with adverse event ÷ total number of patients

4.3.7. Other complications

Back and leg pain

Summary:

We considered back and leg pain in the efficacy and effectiveness section. However, some have suggested that early back and leg pain morbidity among patients receiving rhBMP may be the result of a pro-inflammatory effect of the protein[47] and, therefore, could be considered

as an adverse event. We found no evidence of increased back or leg pain in patients receiving on-label use of rhBMP-2 at early follow-up periods of ≤6 months, Table 55. Of the six RCTs reporting off-label use, only Burkus et al[1, 12] reported a statistically worse back and leg pain in the rhBMP-2 group compared with control at the 6 month follow-up, but the mean difference in pain between groups was less than the MCID of 20 mm.

Table 55. Randomized controlled trials evaluating back or leg pain at early follow-up periods of ≤6 months.

Outcome measure	Author	Results		p-value
ON-LABEL USE				
ODI Range: 0 – 100		Mean score improvement from baseline (points)		
	Boden (2000)	rhBMP-2 (n = 11)	ICBG (n = 3)	
	3 mos.	9	35	NR
	6 mos.	12	-18	NR
	Burkus (2002)	rhBMP-2 (n = 143)	ICBG (n = 136)	
	1.5 mos.	12	14	NS
	3 mos.	20	14	NS
	6 mos.	25	21	NS
ODI "success" (≥ 15% improvement from baseline score)		% (n)		
	Boden (2000)	rhBMP-2 (n = 11)	ICBG (n = 3)	
	3 mos.	55% (6)	0% (0)	NR
	6 mos.	64% (7)	67% (2)	NR
Back pain NRS Range: 0 – 20		Mean score improvement from baseline (points)		
	Burkus (2002)	rhBMP-2 (n = 143)	ICBG (n = 136)	
	1.5 mos.	6.5	7.3	NS
	3 mos.	7.1	7.1	NS
	6 mos.	7.2	7.2	NS
Back pain "success" (> 3 point improvement from baseline score)		% (n)		
	Burkus (2002)	rhBMP-2 (n = 143)	ICBG (n = 136)	
	1.5 mos.	77% (110)	76% (103)	NS
	3 mos.	74% (106)	78% (106)	NS
	6 mos.	78% (112)	72% (98)	NS
Leg pain NRS Range: 0 – 20		Mean score improvement from baseline (points)		
	Burkus (2002)	rhBMP-2 (n = 143)	ICBG (n = 136)	
	1.5 mos.	5.0	4.1	NS
	3 mos.	5.7	5.7	NS
	6 mos.	6.2	6.2	NS
OFF-LABEL				

USE				
ODI		Mean score improvement from baseline (points)		
Range: 0 – 100	Boden (2002)	rhBMP-2 (n = 22)*	ICBG (n = 5)	
	1.5 mos.	~11	~10	NS
	3 mos.	~20	~15	NS
	6 mos.	~23	~17	NS
	Glassman (2008)	rhBMP-2 (n = 25)	ICBG (n = 21)	
	3 mos.	14	13	NS
	6 mos.	18	17	NS
	Burkus (2005/06)	rhBMP-2 (n = 79)	ICBG (n = 52)	
	6 mos.	32	26	.031
ODI "success"		% (n)		
(≥ 15% (or > 20%†) improvement from baseline)	Boden (2002)	rhBMP-2 (n = 22)*	ICBG (n = 5)	
	1.5 mos.	63% (14)	80% (4)	NS
	3 mos.	84% (18)	60% (3)	NS
	6 mos.	86% (19)	80% (4)	NS
Back pain		Mean score improvement from baseline (points)		
Range: 0 – 20	Boden (2002)	rhBMP-2 (n = 22)*	ICBG (n = 5)	
	1.5 mos.	~7	~7	NS
	3 mos.	~8.5	~5	NS
	6 mos.	~8	~4	NS
	Glassman (2008)	rhBMP-2 (n = 25)	ICBG (n = 21)	
	1.5 mos.	4.3	4.0	NS
	6 mos.	4.1	4.0	NS
	Burkus (2005/06)	rhBMP-2 (n = 79)	ICBG (n = 52)	
	6 mos.	9.2	7.7	.006
Leg pain		Mean score improvement from baseline (points)		
Range: 0 – 20	Boden (2002)	rhBMP-2 (n = 22)*	ICBG (n = 5)	
	1.5 mos.	~5.5	~7	NS
	3 mos.	~6.5	~3	NS
	6 mos.	~4	~3	NS
	Glassman (2008)	rhBMP-2 (n = 25)	ICBG (n = 21)	
	1.5 mos.	4.6	4.1	NS
	6 mos.	4.4	4.2	NS
	Burkus (2005/06)	rhBMP-2 (n = 79)	ICBG (n = 52)	
	6 mos.	7.7	7.3	.043
ODI		Mean score improvement from baseline (points)		
Range: 0–100	Kanayama (2006)	OP-1 Putty (n = 9/10)	Local autograft (n=10)	

	3 mos.	15	17	NS
	6 mos.	23	31	NS
		OP-1/local autograft (n=18)	ICBG (n = 18)	
	Delawi (2010)			
	1.5 mos.	11	6	NS
	3 mos.	27	18	NS
	6 mos.	24	23	NS
NDI		Mean score improvement from baseline (points)		
Range: 0–100	Baskin (2003)	rhBMP-2 (n = 18)	ICBG (n = 15)	
	1.5 mos.	37	33	NS
	3 mos.	39	34	NS
	6 mos.	48	39	NS
Neck pain		Mean score improvement from baseline (points)		
Range: 0 – 20	Baskin (2003)	rhBMP-2 (n = 18)	ICBG (n = 15)	
	1.5 mos.	11	7	NS
	3 mos.	11	8	NS
	6 mos.	11	10	NS
Arm pain		Mean score improvement from baseline (points)		
Range: 0 – 20	Baskin (2003)	rhBMP-2 (n = 18)	ICBG (n = 15)	
	1.5 mos.	14	9	NS
	3 mos.	14	8	NS
	6 mos.	15	10	NS

* Boden (2002): results pooled for both rhBMP-2 groups (i.e., with (n = 11) or without (n = 11) instrumentation).

Cancer

Summary:

On-label

Only one RCT was found which reported no difference in the risk of cancer at 24 months following ALIF with rhBMP-2 (0.7%) compared with ICBG (0.7%). The strength of evidence is *low* regarding these estimates.

Off-label

Three RCTs and one cohort studies were identified which generally reported higher cancer risks at 1, 2, 4 and 5 years following the use of rhBMP-2 or rhBMP-7 in the lumbar spine. One RCT of a higher dose (40 mg) of rhBMP-2 reported higher cancer risks following PLF with rhBMP-2 compared with controls at 24 months (3.8% vs. 0.9%) and at 60 months (6.3% vs. 2.2%). Similarly, higher incidences of cancer were reported following PLF with rhBMP-7 compared with controls in two RCTs, one with 12 months (5.6% vs. 0%) and one with 48 months (12.5% vs. 8.3%) follow-up. One cohort study, a retrospective chart review, reported higher risks of cancer following various surgical approaches with rhBMP-2 (16.7%) compared with control (7.6%). The strength of evidence is *moderate* regarding these between-group comparisons.

Data are summarized in Table 56.

Detailed results

rhBMP-2 on-label use: lumbar spine

- RCTs (1 RCT):
  The pivotal RCT of the FDA SSED for InFUSE conducted by Burkus (2002)[6] reported a cancer risk of 0.7% for both the rhBMP-2 group (n = 143) and the ICBG group (n = 136)[22]. There was one case of pancreatic cancer and one case of breast cancer, respectively.

rhBMP-2 off-label use: lumbar spine

- RCTs (1 RCT):
  The FDA executive summary for AMPLIFY reported outcomes from one RCT, which corresponds to Dimar (2009)[8]. The FDA executive summary reported higher risks of cancer in the rhBMP-2 group (n = 239) compared with the ICBG group (n = 224) at 24 months (3.8% vs. 0.9%) and 60 months (6.3% vs. 2.2%)[46]. Nine cases of cancer were documented in the rhBMP-2 group (laryngeal, lung, ovarian, pancreatic, prostate, 2 basal cell, 2 squamous cell) and two cases in the ICBG group (lymphoma and colon) at 24 months; by 60 months, six additional cases of cancer were documented in the rhBMP-2 group (lymphoma, prostate, stomach, thyroid, ocular, leukemia) for a cumulative total of 15 cases and three additional cases in the ICBG group (breast, thyroid, squamous cell) for a cumulative total of 5 cases. No other details were provided.

- Cohort studies (2 studies):
  Latzman et al. (2010) reported a higher incidence of cancer in patients who had undergone lumbar spinal fusion with allograft or autograft and either with (n = 101) versus without rhBMP-2 (n = 24) (17% versus 8%, respectively; however the difference was not statistically significant (*P* = .12)). Four cases of cancer in the rhBMP-2 included lung adenocarcinoma (n = 1, presented at 1 month), pancreatic adenocarcinoma (n = 1, presented at 13 months), prostate adenocarcinoma (n = 1, presented at 21 months), and rectal adenocarcinoma (n = 1, presented at 24/8 months following the first/second surgeries); the authors noted that the lung cancer patient had a nodule on his chest radiographs 10 months prior to the cancer diagnosis. Eight cases of cancer in the control group included lung adenocarcinoma (n = 1, presenting at 3 months), basal cell carcinoma (n = 3, presenting at 22, 24, and 63 months), colon adenocarcinoma (n = 1, presenting at 25 months), prostate adenocarcinoma (n = 2, presenting at 12 and 32 months), and bladder carcinoma (n = 1, presentation unknown).

  Mines et al. (2011) conducted a retrospective database cohort study with the purpose of assessing the risk of pancreatic cancer following rhBMP-2 use in the lumbar spine in Medicare patients[177]. There was no difference in the incidence of pancreatic cancer in those who received rhBMP-2 (0.052% of 15,460 patients) compared with those who did not (0.106% of 78,194 patients), with a hazard ratio of 0.70 (95% CI: 0.34, 1.45).

- Case reports
  Steib et al. (2010) published a case report of a patient who developed a large posterior neurofibromatosis type I (NF1) tumor on his back five months following revision surgery with rhBMP-2[178]. The patient died several months later from septic shock. The authors noted that the patient's mother also died from a NF1 tumor on the radial nerve, and his three brothers had symptoms (but no tumors) associated with the disease.

rhBMP-7 off-label use: lumbar spine

- RCTs (2 RCTs):
  Two small RCTs reported a higher risk of cancer in rhBMP-7 patients compared with ICBG patients[10, 16-18]. Vaccaro et al. (2008) reported "neoplasms, benign, malignant, and unspecified" in three rhBMP-2 patients (13%) and one ICBG patient (8%) at 48 months; no other details were reported[18]. One rhBMP-7 patient (5.6%) in the RCT by Delawi et al. (2010) was diagnosed with a grade IV glioblastoma 11 months following surgery; the patient withdrew from the study and no other information was available[10].

Table 56. Cancer

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
RCTs					
rhBMP-2	1[22]†	24 mos.	143	0.7%	n/a
Control			136	0.7%	n/a
BMP off-label use: lumbar spine					

	Studies	Length follow-up	Patients (n)	Mean (%)	Range of means (%)
RCTs					
rhBMP-2	1[46]‡	24 mos.	239	3.8%	n/a
Control			224	0.9%	n/a
rhBMP-2	1[46]‡	60 mos.	239	6.3%	n/a
Control			224	2.2%	n/a
rhBMP-7	1[10]	48 mos.	12 mos.	18	5.6%
Control				16	0%
rhBMP-7	1[16-18]	48 mos.	24	12.5%	n/a
Control			12	8.3%	n/a
Cohort studies					
rhBMP-2	1[179]	18 mos.	24	16.7%	n/a
Control			105	7.6%	n/a
rhBMP-2	1[177]§	47 mos.	15,640	0.052%	n/a
Control			78,194	0.106%	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients.
† Data from the pivotal RCT in the FDA SSED for InFUSE, which corresponds to Burkus et al. (2002). Results from the pilot RCT (N = 14) and the prospective case series (N = 134) submitted to FDA were not included in this analysis.
‡ Dimar et al. is the corresponding RCT for the FDA data summary. Dimar reports only 8 cases of cancer in the rhBMP group, while the SSED reports 9 at 24 month follow-up.
§ Mines (2011): database study looking at incidence of pancreatic cancer.

Cardio/vascular

Summary:

On-label

One RCT and one nonrandomized comparative study, the FDA SEED for InFUSE, reported similar risks of cardio/vascular events following rhBMP-2 use in the lumbar spine compared with controls (4.2%–10.1 vs. 2.2%–12.2%, respectively). The strength of evidence is *low* regarding these between-group comparisons.

Off-label

In general, results from four RCTs and three cohort studies (to include the FDA SSPB for OP-1) show similar risks of cardio/vascular events following rhBMP-2 or rhBMP-7 use in the lumbar spine compared with controls (3.9%–22.2% vs. 2%–24.1%, respectively). The strength of evidence is *high* regarding these between-group comparisons.

All data are summarized in Table 57.

Detailed results

rhBMP-2 on-label use: lumbar spine

- RCTs (1 RCT):
  Burkus et al. (2002) reported similar risks of vascular events between the rhBMP-2 and ICBG treatment groups (4.2% versus 2.2%, respectively)[6]. The cases in the rhBMP-2 group comprised of laceration of the iliac vein, while details of the three cases in the control group were not reported. These risks do not include deep vein thrombosis, as these were reported and accounted for separately.

- FDA SSED for InFUSE reported "cardio/vascular" complications in 5.2% of patients in the rhBMP-2 group (18 events in 15 patients) compared with 8.6% of the ICBG patients (14 events in 12 patients). Vascular intraoperative complications occurred in 4.9% of the BMP patients (15 events in 14 patients) and in 3.6% of control patients (5 events in 5 patients)[22]. No further details were reported.

rhBMP-2 off-label use: lumbar spine

- RCTs (2 RCTs):
  The risk of cardiovascular adverse events was similar between treatment groups 18.3% for rhBMP-2 versus 22.1% for ICBG) as reported by two RCTs[8, 26]. Glassman et al. reported perioperative cardiac complications in 2% of rhBMP-2 patients (1/50) and 13% of IBCG patients (7/52)[26], while the FDA report for AMPLIFY[46] (which corresponds to the Dimar RCT[8]) reported cardiovascular events in 22.2% (53/239) of rhBMP-2 patients and in 24.1% (54/224) of ICBG patients (P = NS) through 24 months follow-up[8]. No further details were provided.

- Cohort studies (1 study):
  Lee et al. (2010) reported perioperative cardiac events in 5.9% (2/34) of rhBMP-2 patients compared with 9.8% of patients in the ICBG group[29]. All patients were age 65 years or older; no other details were reported.

- Case series (2 studies):
  Anderson et al. (2011) reported one case (2%) of tachycardia with transient hypotension and trace pericardial infusion; the patient was medically managed[56]. In three months follow-up, Glassman et al. (2010/2011) reported two cases of myocardial infarction (0.19%), one case of cardiac ischemia (0.10%), and six cases of arrhythmia (0.58%)[164, 165].

rhBMP-7 off-label use: lumbar spine

- RCTs (2 RCTs):
  Two RCTs reported cardio/vascular events in similar percentages of patients in the rhBMP-2 and ICBG treatment groups. Vaccaro et al. (2004/2005/2008) documented two cases of cardiac complications in the rhBMP-7 group (8% of patients) compared with

none in the ICBG group, with 48 months follow-up[16-18]. In addition, 17% of patients in each group had cardiovascular complications. Delawi reported cardiovascular events in 6% of patients in each treatment group[10].

- FDA SSPB for OP-1 reported similar risks of cardio/vascular events between treatment groups. Specifically, cardiac disorders were documented in 4% of rhBMP-7 patients (9/228) and in 1% (1/98) of ICBG patients; vascular disorders occurred in 8% of rhBMP-7 patients (17/228) compared with 10% (10/98) of control patients.

BMP (unspecified) use in the lumbar spine

- Cohort studies (1 study):
  Deyo et al. (2011) reported no difference in cardiac, pulmonary, or stroke complications between patients treated with BMP versus no BMP (5.1% versus 5.7%, *P* = .285) in a large database study.

Table 57. Cardio/vascular

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
RCTs					
rhBMP-2	1[6]	Perioperative	143	4.2% (6/143)	n/a
Control			136	2.2% (3/136)	n/a
Cohort studies					
rhBMP-2	1[22] (FDA SSED)	24 mos.	288	4.9-10.1%†	n/a
Control			129	3.6-12.2%†	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	1[26]	Perioperative	50	2% (1/50)	n/a
Control			52	13% (7/52)	n/a
rhBMP-2	1[46]	24 mos.	239	22.2% (53/239)	n/a
Control			224	24.1% (54/224)	n/a
rhBMP-7	2[10, 16-18]	12-48 mos.	42	7-17%‡	6-25%
Control			28	7-11%‡	6-17%
Cohort studies					
rhBMP-2	1[29]	Perioperative	34	6%	n/a
Control			41	10%	n/a
rhBMP-7	1[3] (FDA SSPB)	NR	228	3.9-11.4%§	n/a
Control			98	1-2%§	
BMP (unspecified)	1[162]	Min. 48 mos.	1703	5.1%	

Control			15,119	5.7%	
Case series					
rhBMP-2	2[56, 164, 165]	3-12 mos.	1087	0.68-0.92%	0.87-2%

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† FDA SSED for InFuse reported cardio/vascular adverse events in 5.2% (15/288) versus 8.6% (12/139), and intraoperative vascular events in 4.9% (14/288) versus 3.6% (5/139) of rhBMP-2 versus ICBG patients, respectively.
‡ Vaccaro 2004/2005/2008 reported cumulative "cardiac" adverse events at 48 months in 2/24 rhBMP-7 patients versus 0/12 control patients, and "cardiovascular" events at 12 months in 4/24 rhBMP-7 patients versus 2/12 control patients.
§ FDA SSPB for OP-1 reported cardiac disorders in 9/228 versus 1/98 patients in the OP-1 versus control groups, respectively; and vascular disorders in 17/228 and 1/98 patients in the OP-1 versus control groups.

Deep vein thrombosis

Summary:

On-label

There was no difference in the incidence of DVT in patients treated with rhBMP-2 compared with control as reported by one RCT: 0% versus 1.5%, respectively. The strength of evidence is *low* regarding these estimates.

Off-label

According to three comparative studies, the risks of DVT in patients treated with rhBMP-2 in the lumbar (1 RCT, 1 cohort) and cervical spine (1 cohort) were similar compared with controls: 0%–9% versus 1.9%–12%, respectively. The strength of evidence is *low* regarding these between-group comparisons.

All data are summarized in Table 58.

Table 58. Deep vein thrombosis

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
RCTs					
rhBMP-2	1[6]	0 mos.	143	0%†	n/a
Control			136	1.5%†	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	1[9]	0 mos.	34	0%	n/a
Control			33	3%	n/a
Cohort studies					
rhBMP-2	1[29]	0 mos.	34	9%	n/a
Control			41	12%	n/a
BMP off-label use: cervical spine					

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
Cohort studies					
rhBMP-2	1[35]	24 mos.	48	0%	n/a
Control			156	1.9%	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† Burkus 2002: Deep vein thrombosis included in the tally of cardio/vascular events, reported above.

Death

Summary:

On-label

No difference in the incidence of death between patients treated with rhBMP-2 in the lumbar spine compared with control was reported by one RCT at 24 months (0% vs. 0.7%, respectively). The strength of evidence is *low* regarding these estimates.

Off-label

In the lumbar spine, similar risks of death following the use of rhBMP-2 or rhBMP-7 (1.6%–5.3%) compared with controls (1.7%–6.0%) were reported by four RCTs and two cohort studies at 24 and 36 months, respectively. Following cervical fusion, one retrospective cohort study reported a statistically higher risk of death up to 90 days post-operative in patients treated with (4.2%; 11/260) versus without (1.7%; 9/515) BMP (type unspecified); *P* = .047. The causes of death were not reported, and the significance of this result should be interpreted with some caution as no demographic or surgical details were provided and there is thus an absence of controlling for possible confounding between treatment groups. The strength of evidence is *high* regarding these estimates in the lumbar spine and *insufficient* in the cervical spine.

All data are summarized in Table 59.

Detailed results

rhBMP-2 on-label use: lumbar spine

- RCTs (1 RCT):
  One patient enrolled in the pivotal trial evaluating rhBMP-2 use in the lumbar spine died due to cardiovascular disease between 5 and 9 months following spinal fusion[6]. The death was not attributed to rhBMP-2 use. The same patient was reported in the FDA SSED for InFUSE[22].

rhBMP-2 off-label use: lumbar spine

- RCTs (3 RCTs):
  Data from three RCTs examining rhBMP-2 off-label use in the lumbar spine compared with ICBG suggested there were no differences in the incidence of patient death between treatment groups, with pooled risks of 1.6% (five deaths) and 1.7% (5 deaths), respectively[8, 25, 26]. In addition, one perioperative death was reported by Glassman et al.

(2008), however the authors provided no details or the treatment group the patient was in[26]. Dawson et al. (2009) reported one rhBMP-2 patient death; neither the timing nor the cause of death were reported[25]. Dimar et al. (2009) reported three deaths in the rhBMP-2 group (1.3%) compared with four in the ICBG group (1.8%) (*P* = .717); no information was provided other that the causes of death were "unrelated to surgery"[8]. Finally, Glassman et al. (2008) reported one death in each treatment group during follow-up[26].

- Cohort studies (2 studies)
  Mines et al. (2011) retrospectively evaluated complications following lumbar fusion with or without rhBMP-2 in a Medicare database study. With a median follow-up of 17 months, 3.1% of patients in the rhBMP-2 group died (479/15,460) compared with 5.1% of patients in the control group (2988/78,194)[177]. No other information was provided. Crawford et al. (2010) reported on death in the control group prior to the 24 month follow-up; no other details were reported[27].

- Case reports
  Steib et al. (2010) published a case report of a patient who developed a large posterior neurofibromatosis type I (NF1) tumor on his back five months following revision surgery with rhBMP-2[178]. The patient died several months later from septic shock.

rhBMP-7 off-label use: lumbar spine

- RCTs (1 RCT):
  In two reports of one RCT, Vaccaro, Lawrence et al. (2008) and Hwang et al. (2010) reported no difference in the incidence of death between those who received lumbar fusion with rhBMP-7 or autograft (5.3% (11/208) versus 6% (5/87)[2, 19]. Time of death in the rhBMP-7 group was soon after surgery (n = 1), between 6 weeks and 3 months (n = 1), between 3 and 6 months (n = 3), between 12 and 24 months (n = 4), and between 24 and 36 months (n = 2). In the autograft group, time of death was between 6 and 12 months (n = 2), between 12 and 24 months (n = 2), and between 24 and 36 months (n = 1). No other information was reported.

BMP (unspecified) off-label use: cervical spine

- Cohort studies (1 study)
  Yaremchuk et al. (2010) conducted a retrospective cohort study in which patients received cervical spinal fusion with BMP (n = 260) or without BMP (n = 515). In the 0-days following surgery, there were statistically more deaths in the BMP group compared with the control group (4.2% (11/260) versus 1.7% (9/515), respectively; *P* = 0.47). The hazard ratio estimates that patients who received BMP were 2.44 times more likely to die than those who did not receive BMP (HR = 2.44 (95% CI, 1.01, 5.89)[36]. The causes of death were not reported. Note that this study (like all but one cohort study included in this HTA) received a LoE grade of III: no demographic information or surgical information was provided, making it difficult to know whether the results are due to inherent differences in the treatment or patient characteristics between treatment groups.

Thus, since there was no controlling for potential confounding, the significance of this result should be viewed with some caution.

Table 59. Death (any)

	Studies	Length follow-up	Patients (n)	Mean (%) (n)*	Range of means (%)
BMP on-label use: lumbar spine					
RCTs					
rhBMP-2	1[6]	24 mos	123	0%	n/a
Control			136	0.7% (1)†	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%) (n)***	**Range of means (%)**
RCTs					
rhBMP-2	3[8, 25, 26]	24 mos.	314	1.6% (5)	1.3-4%
Control			297	1.7% (5)	0-2%
rhBMP-7	1[2, 19]	36 mos.	208	5.3% (11)	n/a
Control			87	6% (5)	n/a
Cohort studies					
rhBMP-2	2[27, 177]	3-17 mos.	15,496	3.09% (479)	0-3.1%
Control			78,218	3.82% (2990)	4-5.1%
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range (%)**
Cohort studies					
BMP (unspecified)	1[36]	3 mos.	260	4.2% (11)	n/a
Control			515	1.7% (9)	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† One patient enrolled in the Burkus 2002 RCT died; the same patient was reported in the FDA SSED for InFUSE. The patient died between 5 and 9 months postoperation due to cardiovascular disease.

Infrequently reported adverse events

Sepsis

rhBMP-2 off-label use: lumbar spine

Glassman et al. (2008) reported one case of line-related sepsis in a rhBMP-2 patient (2%, or 1/50); no patients in the ICBG group had this complication[26]. No other details were reported. Moshel et al. (2008) published a case study in which the patient was suspected to have sepsis, which was diagnosed following the development of transient supraventricular tachycardia[180]. No evidence of bacterial infection was found in blood, urine, or sputum cultures. The patient was treated with prophylactic antibiotics.

Local or systemic toxicity

There were no cases of local or systemic toxicity as reported for off-label use of rhBMP-2 in the lumbar spine by one case series (N = 70)[166]; off-label use of rhBMP-7 in the lumbar spine by one RCT (N = 36)[16-18] and two case series (N = 28)[143, 144, 181]; and off-label use of rhBMP-7 in the cervical spine by one case series (N = 14)[143].

Paresis

rhBMP-2 off-label use: lumbar spine

One small cohort study of 40 patients reported identical risks of perioperative L5 paresis in the rhBMP-2 and ICBG groups (5%); no other information was reported[13].

Pulmonary embolism

rhBMP-2 off-label use

Xu et al. (2011) reported no cases of PE in the rhBMP-2 group compared with two cases in the control group (1.3%) following cervical fusion[35]. Glassman et al. (2010/2011) reported one case of pulmonary embolism (0.10%) following lumbar fusion with rhBMP-2[164, 165].

Renal failure

rhBMP-2 off-label use: lumbar spine

Latzman et al. (2010) reported no cases of progressive renal failure but found that more patients in the rhBMP-2 versus the autograft/allograft group experienced transient renal insufficiency (13% (3/24) versus 0% (0/105); $P = .006$)[179]. Renal insufficiency was defined as blood urea nitrogen levels over 30 mg/dL and creatine levels over 1.5 mg/dL. All patients had their values return to normal by two months postoperation. Of note, two of the three patients with renal insufficiency were diagnosed with malignancies within eight months of surgery: one patient had a nodule on his lung for at least ten months prior to spinal fusion and was diagnosed with adenocarcinoma one month after surgery, the other patient was diagnosed with rectal adenocarcinoma eight months following a second surgery with rhBMP-2. One case series reported acute renal failure in 0.19% of patients (2/1037) following rhBMP-2 use in posterolateral fusion[164, 165].

4.3.8. Secondary surgical procedures

Revision: *surgery that modified or adjusted the original implant*

Summary:

On-label

One nonrandomized comparative study (an integrated analysis) reported no difference in the risks of revision, defined as surgery that modified or adjusted the original implant, between patients treated with rhBMP-2 in the lumbar spine compared with controls (0.4% vs. 2.0%, respectively). The strength of evidence is *insufficient* regarding these estimates.

Off-label

In general, risks of revision following lumbar spinal fusion were similar between rhBMP and control groups as reported by seven RCTs (five rhBMP-2 and two rhBMP-7) over a range of 17 to 48 months follow-up: 6.0% versus 6.2%, respectively (pooled results). Overall risks were slightly higher with rhBMP-7 use (9.5% vs. 11%) compared with rhBMP2 use (3.8% vs. 4.8%). Results from three cohorts, two evaluating rhBMP-2 in the lumbar spine and one in the cervical spine, indicate lower risks of revision in the BMP groups compared with the control groups: 3% versus 10% (lumbar) and 0% versus 4% (cervical), respectively. The strength of evidence is *high* regarding these between-group comparisons.

All data are summarized in Table 60.

Detailed results

rhBMP-2 on-label use: lumbar spine

- Integrated analysis (1 study):
  In the integrated analysis of the Burkus 2002 RCT[6] as well as some data from one case series[24] and unpublished trials, Burkus et al. (2003) reported similar risks of revision in rhBMP-2 (0.4% (1/277)) and ICBG (2.0% (8/402)) patients following anterior lumbar fusion[23]. No details were reported.

- Case series (1 study):
  Burkus et al. (2009) reported revision in 0.4% of patients from a case series with six years follow-up (1/277); no details were reported[153].

rhBMP-2 off-label use: lumbar spine

- RCTs:
  There was no difference in the pooled risk of revision between rhBMP-2 and ICBG groups as reported by five RCTs with a mean follow-up ranging from 17 to 24 months[7-9, 25, 26]. In the rhBMP-2 group, 3.8% of all patients (14/368) underwent revision; the mean incidence reported in each study ranged from 1.7-10% of patients. In the ICBG group, 4.8% of all patients (16/335) underwent revision; the mean incidence ranged from 0-13% of patients. Because of heterogeneity in surgical approaches, product(s) used, and patient demographics (see Table "off label use of rhBMP-2 in lumbar spine RCT overview), we did not perform a meta-analysis.

  Details of the timing and causes of revision are as follows:

  rhBMP-2:
  - 1 day: malpositioned screws (n = 1)[25]
  - 8 months: low back pain (n = 1)[7]
  - 12 months: cause not reported (n = 1)[7]
  - Timing not reported:
    - Nonunion (n = 1)[26]
    - Adjacent level degeneration (n = 3)[26]
  - No details reported for seven revision patients reported in two studies[8, 9]

  ICBG:
  - 12-24 months: nonunion (n = 2)[25]

- Timing not reported:
  - Nonunion (n = 5)[26]
  - Adjacent level degeneration (n = 1)[26]
  - Repositioning of pedicle screw (n = 1)[26]
- No details reported for seven revision patients reported in two studies[8, 9]

- Cohort studies:
  Revision risks were similar in the rhBMP-2 and control groups as reported by two retrospective cohort studies with two different types of control treatments[15, 29]:
  - *rhBMP-2 versus ICBG (1 study)*
    Lee et al. (2010) reported similar risks of revision in both treatment groups in a small retrospective cohort study with a mean of 38 months follow-up (17% (1/6) versus 22% (2/9), respectively)[29]. No further details were reported.

  - *rhBMP-2 versus allograft/bone marrow aspirates (1 study)*
    Taghavi et al. (2010) reported slightly lower levels of revision in the rhBMP-2 group compared with the control group in a retrospective cohort study evaluating revision posterolateral fusion with mean of 28 months follow-up.

rhBMP-7 off-label use: lumbar spine

- RCTs (2 RCTs):
  Pooled revision risks were similar in the rhBMP-7 and ICBG treatment groups as reported by two RCTs with 24 to 48 months follow-up (9.5% (22/232) versus 11% (11/99), respectively)[2, 16-19].

  Details of the timing and causes of revision are as follows:

  rhBMP-7:
  - 0-36 months: details not reported (n = 21); between 36 and 48 months an additional 3 patients underwent revision (i.e., 3/144 patients versus the 208 available for the 36 month follow-up; this information is not included in the table below due to different denominators)[2, 19]
  - 24-48 months: revision decompression (n = 1)[16-18]

  ICBG:
  - 0-36 months: details not reported (n = 11); between 36 and 48 months an additional 3 patients underwent revision (i.e., 3/58 patients versus the 87 available for the 36 month follow-up; this information is not included in the table below due to different denominators)[2, 19]

- Case series (1 study):
  One small case series reported a revision risk of 8% (1/12) in the 24 months following rhBMP-7 use in the lumbar spine[144, 181]. Revision was performed for nonunion; the timing was not reported.

rhBMP-2 off-label use: cervical spine

- Cohort studies (1 study):

Vaidya, Carp et al. (2007) reported similar risks of revision in the rhBMP-2 and allograft/demineralized bone matrix groups following ACDF (0% (0/22) versus 4% (1/24), respectively)[34]. The one revision was performed at 12 months for nonunion.

- Case series (1 study)
  Shen et al. reported a revision risk of 6% (8/127) in patients undergoing single- or multilevel fusion with rhBMP-2[171]. All revisions were performed for nonunion within 24 months follow-up.

Table 60. Revision: surgery that modified or adjusted the original implant

BMP on-label use: lumbar spine	Studies	Length follow-up	Patients (n)	Mean (%) (n)*	Range of means (%)
Cohort studies					
rhBMP-2	1[23] (integrated analysis)	24 mos.	277	0.4% (1)	n/a
Control			402	2.0% (8)	n/a
Case series					
rhBMP-2	1[153]	72 mos.	277	0.4% (1)	n/a
BMP off-label use: lumbar spine	**Studies**	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	5[7-9, 25, 26]	17-24 mos.	368	3.8% (14)	1.7-10%
Control			335	4.8% (16)	0-13%
rhBMP-7	2[2, 16-19]	24-48 mos.	232	9.5% (22)	4-10.1%
Control			99	11% (11)	0-13%
Cohort studies					
rhBMP-2	2[15, 29]	28-36 mos.	30	3% (1)	0-17%
Control			51	10% (5)	7-22%
Case series					
rhBMP-7	1[144, 181]	24 mos.	12	8% (1/12)	n/a
BMP off-label use: cervical spine	**Studies**	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	1[34]	24 mos.	22	0% (0)	n/a
Control			24	4% (1)	n/a
Case series					
rhBMP-2	1[171]	24 mos.	127	6% (8)	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients

Hardware removal: *removal of one or more components of the original implant (including replacement with a different implant)*

Summary:

On-label

The incidence of hardware removal, defined as removal of one or more components of the original implant (including replacement with a different implant), was similar between rhBMP-2 and control groups as reported by one RCT and one cohort study (integrated analysis with partial overlap of data with the RCT) at 24 months: 1.4% versus 0% and 1.4% vs. 1.7%, respectively. The strength of evidence is *low* regarding these estimates.

Off-label

Risks of hardware removal were slightly less in patients receiving rhBMP-2 in the lumbar spine compared with controls across four RCTs with 24 months follow-up (2.8% vs. 7.2%) and similar as reported by two cohort studies with 3 to 28 months follow-up (8.0% for both groups). The strength of evidence is *moderate* regarding these between-group comparisons.

All data are summarized in Table 61.

Detailed results

rhBMP-2 on-label use: lumbar spine

- RCTs (1 RCT):
  Burkus et al. (2002) reported similar risks of hardware removal between groups (1.4% (2/143) versus 0% (0/136) for rhBMP-2 versus ICBG, respectively)[6].

  Details of the timing and causes of revision are as follows:
  - rhBMP-2:
    - 5 days: vertebral bone fracture and displacement (n = 1)
    - 4 months: implant displacement and possible failed fusion (n = 1)

- Integrated analysis (1 study):
  In the integrated analysis of the Burkus 2002 RCT[6] (reported above) as well as some data from one case series[24] and unpublished trials, Burkus et al. (2003) reported similar risks of hardware removal in rhBMP-2 (1.4% (4/277)) and ICBG (1.7% (7/402)) patients following anterior lumbar fusion[23]. No details were reported.

- Case series (1 study):
  Burkus et al. (2009) reported hardware revision in 0.4% of patients from a case series with six years follow-up (1/277); no details were reported[153].

rhBMP-2 off-label use: lumbar spine

- RCTs:
  Fewer patients in the rhBMP-2 groups underwent hardware removal compared with the ICBG groups as reported by four RCTs with a mean follow-up of 24 months in all four studies (2.8% (11/393) versus 7.2% (25/349), respectively)[1, 8, 12, 25, 26]. Because of heterogeneity in surgical approaches, product(s) used, and patient demographics (see

Table "off label use of rhBMP-2 in lumbar spine RCT overview), we did not perform a meta-analysis.

Details of the timing and causes of revision are as follows:

rhBMP-2:

- 6 months: details not reported (n = 1)[25]
- ≤ 24 months: nonelective, no other details reported (n = 10)[8]

ICBG:

- Postoperative: disc material removal (n = 1)[1, 12]
- ≤ 24 months for the following reasons:
  - Nonelective, no other details reported (n = 23)[8]
  - Radiculitis/weakness (n = 1)[26]

- Cohort studies:
  Risks of hardware removal were identical (8%) in the rhBMP-2 and control groups as reported by two retrospective cohort studies with two different types of control treatments[15, 27]:
  - *rhBMP-2 versus autograft (1 study)*
    Crawford et al. (2010) reported hardware removal in 8% of rhBMP-2 and control patients at up to three months follow-up[27].
  - *rhBMP-2 versus allograft/bone marrow aspirates (1 study)*
    Taghavi et al. (2010) reported hardware removal in 8% of rhBMP-2 and control patients at a mean of 28 months follow-up[15].

Details of the timing and causes of revision are as follows:

rhBMP-2:

- Timing not reported, for the following reasons:
  - Iliac screw removed (n = 3)[27]
  - Persistent irritation (n = 2)[15]

ICBG:

- Timing not reported, for the following reasons:
  - Iliac screw removed (n = 2)[27]
  - Persistent irritation (n = 3)[15]

Table 61. Hardware removal: removal of one or more components of the original implant (including replacement with a different implant)

BMP on-label use: lumbar spine	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
RCTs					
rhBMP-2	1[6]	24 mos.	143	1.4% (2)	n/a
Control			136	0% (0)	n/a

Cohort studies					
rhBMP-2	1[23] (integrated analysis)	24 mos.	277	1.4% (4)	n/a
Control			402	1.7% (7)	n/a
Case series					
rhBMP-2	1[153]	72 mos.	277	0.4% (1)	n/a

BMP off-label use: lumbar spine

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
RCTs					
rhBMP-2	4[1, 8, 12, 25, 26]	24 mos.	393	2.8% (11)	0-4.2%
Control			349	7.2% (25)	0-10.3%
Cohort studies					
rhBMP-2	2[15, 27]	3-28 mos.	60	8% (5)	8%
Control			62	8% (5)	8%

n/a: not applicable; NR: not reported

* Mean calculated as: number of patients with adverse event ÷ total number of patients

Supplemental fixation: *surgery to provide additional stabilization to the index site*

Summary:

On-label

Patients treated with rhBMP-2 in the lumbar spine had lower risks of supplemental fixation, defined as surgery to provide additional stabilization to the index site, compared with patients who received ICBG according to data from one small pilot RCT and one pivotal RCT (5.2% vs. 10.8%, respectively). The strength of evidence is *low* regarding these between-group comparisons.

Off-label

The incidence of supplemental fixation following rhBMP-2 or rhBMP-7 in the lumbar spine compared with controls varied across a total of eight studies, depending on the protein evaluated. Two RCTs with 24 months follow-up and four cohort studies with follow-up periods ranging from 3 to 36 months reported lower mean risks of supplemental fixation in the rhBMP-2 groups compared with the controls: 2.5% versus 6.2% and 6.7% versus 9.5%, respectively. Conversely, two small RCTs found a higher mean risk among those treated with rhBMP-7 (10%) than with control (0%) at 24 months. Only one nonrandomized comparative study was found that reported incidence of supplemental fixation following rhBMP-2 use in the cervical spine and showed a lower risk compared with controls at 30 months (0% versus 3.0%). The strength of evidence is *moderate* regarding the risk of supplemental fixation with off-label use of rhBMP-2 in the lumbar spine, *low* for the estimates of the risk for rhBMP-7 use in the lumbar spine and *insufficient* for the estimates of the risk for rhBMP-2 use in the cervical spine.

All data are summarized in Table 62.

Detailed results

rhBMP-2 on-label use: lumbar spine

- RCTs (2 RCTs):
  Pooled data from two RCTs suggest that risks of supplemental fixation are lower following lumbar anterior fusion with rhBMP-2 compared with ICBG (5.2% (8/154) versus 10.8% (15/139), respectively)[6, 21].

  Details of the timing and causes of supplemental fixation are as follows:

  rhBMP-2:
  - ≤ 24 months:
    - Nonunion (n = 7)[6]
    - Radiculitis (decompression also performed) (n = 1)[6]

  ICBG:
  - 18 months: nonunion (n = 1)[21]
  - ≤ 24 months:
    - Nonunion (n = 12)[6]
    - Radiculitis (n = 2)[6]

- Integrated analysis/FDA SSED for InFUSE (2 reports with overlapping data):
  Pooled data from the integrated analysis and FDA SSED for InFUSE, which overlap with each other and the RCTs reported above suggest similar risks of supplemental fixation in rhBMP-2 (6.1% (17/277)) and ICBG (7.0% (28/402)) patients following anterior lumbar fusion[22, 23]. Burkus et al. did not report any details, while the patients reported in the FDA report were described as having nonunion and likely overlap with the patients reported in the RCTs above.

- Case series (1 study):
  Burkus et al. (2009) reported supplemental fixation in 8.3% of patients from a case series with six years follow-up (23/277); no details were reported[153].

rhBMP-2 off-label use: lumbar spine

- RCTs (2 RCTs):
  Fewer patients in the rhBMP-2 groups underwent supplemental fixation compared with the ICBG groups as reported by two RCTs with a mean follow-up of 24 months in both studies (2.5% (8/318) versus 6.2% (17/276), respectively)[1, 8, 12]. No details were reported in either study.

- Cohort studies (3 studies)
  Pooled risks of supplemental fixation were lower (6%) in the rhBMP-2 groups versus the control groups (17%) as reported by three cohort studies with three different types of control treatments[27, 30, 32]:
    - *rhBMP-2 versus ICBG (1 study)*
      Pradhan et al. reported similar risks of hardware removal in both treatment groups (33% (3/9) versus 26% (7/27), respectively) at up to a mean of 34 months follow-up[30].
    - *rhBMP-2 versus autograft (1 study)*
      Crawford et al. (2010) reported hardware removal in 6% (2/36) of rhBMP-2 and 13% (3/24) control patients at up to three months follow-up[27].
    - *rhBMP-2 versus allograft chips (1 study)*
      Slosar et al. (2007) reported lower risks of hardware removal in the rhBMP-2 group (0% (0/45)) compared with the control group (13% (4/30)) at up to 24 months follow-up[32].

  Details of the timing and causes of revision are as follows:
  rhBMP-2:
    - Timing not reported, for the following reasons:
      - Salvage posterior fusion for nonunion (n = 3)[30]
      - Surgery for nonunion (n = 2)[27]

  Control:
    - Timing not reported, for the following reasons:
      - Salvage posterior fusion for nonunion (n = 7)[30]

- Salvage posterolateral fusion for nonunion (n = 4 plus 1 pending at time manuscript was published)[32]
- Surgery for nonunion (n = 3)[27]

rhBMP-7 off-label use: lumbar spine

- RCTs (2 RCTs)
  Data pooled from two RCTs suggest that the risks of supplemental fixation are higher or similar in patients treated with rhBMP-7 (10% (3/34)) compared with ICBG (0% (0/22))[16-18, 20].

  Details of the timing and causes of revision are as follows:

  rhBMP-2:
  - ≤ 12 months: instrumented fusion[20]
  - 24-48 months: decompression and fusion (non-revision) (n = 1)[16-18]

rhBMP-2 off-label use: cervical spine

- Cohort studies (1 study)
  Buttermann et al. (2008) reported a lower risk of supplemental following rhBMP-2 (0%) versus ICBG (3%)[11]. A posterior fusion was performed on one control patient for pseudarthrosis.

Table 62. Supplemental fixation: surgery to provide additional stabilization to the index site

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
RCTs					
rhBMP-2	2[6, 21]	24 mos.	154	5.2% (8)	0-5.6%
Control			139	10.8% (15)	10.3-33%
Cohort studies					
rhBMP-2	1[23]†	24 mos.	277	6.1% (17)†	n/a
Control			402	7.0% (28)†	n/a
Case series					
rhBMP-2	1[153]	72 mos.	277	8.3% (23)	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	2[1, 8, 12]	24 mos.	318	2.5% (8)	2.5-3%
Control			276	6.2% (17)	4.0-15%
rhBMP-7	2[16-18, 20]	12-48 mos.	34	10% (3)	4-20%
Control			22	0% (0)	0%

Cohort studies					
rhBMP-2	3[27, 30, 32]	3-34 mos.	90	6% (5)	0-33%
Control			81	17% (14)	13-26%

BMP off-label use: cervical spine

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
Cohort studies					
rhBMP-2	1[11]	30 mos.	30	0% (0)	n/a
Control			36	3% (1)	n/a

n/a: not applicable; NR: not reported
* Mean calculated as: number of patients with adverse event ÷ total number of patients
† This study is comprised of the population from Burkus 2002 (pivotal RCT for the FDA SSED for InFUSE), one published case series of rhBMP-2 by Kleeman 2001, and one unpublished case-series of ICBG. This study is largely a repeat of the Burkus 2002 RCT due to considerable overlap in patient populations.

Reoperation: *additional procedure at the index level besides a revision, hardware removal, or supplemental fixation*

Summary:

On-label

One nonrandomized comparative study, an integrated analysis, reported a lower risk of reoperation, defined as an additional procedure performed at the index level besides a revision, hardware removal, or supplemental fixation, following rhBMP-2 use in the lumbar spine compared with control (2.9% vs. 8.0%, respectively). The strength of evidence is *insufficient* regarding these estimates.

Off-label

Three RCTs, two investigating rhBMP-2 and one rhBMP-7, and two cohort studies of rhBMP-2 all reported similar risks of reoperation following lumbar spinal fusion with BMP compared with control over 3 to 27 months of follow-up (1.0%–10% vs. 2.0%–10%, respectively). In contrast, one case-control database study of nearly 5000 patients found lower risks of repeat fusion after BMP use compared with those who did not receive BMP. One nonrandomized comparative study also reported a similar risk of reoperation following rhBMP-2 use in the cervical spine compared with controls at 30 months (0% versus 3.0%). The strength of evidence is *high* regarding these between-group comparisons for the off-label use of rhBMP.

All data are summarized in Table 63.

Detailed results

rhBMP-2 on-label use: lumbar spine

- Integrated analysis:

In the integrated analysis of the Burkus 2002 RCT[6] as well as some data from one case series[24] and unpublished trials, Burkus et al. (2003) reported statistically lower risks of reoperation in those who received rhBMP-2 (2.9% (8/277)) compared with ICBG (8.0% (32/402)) patients following anterior lumbar fusion[23] (*P* = .004). No details were reported.

rhBMP-2 off-label use: lumbar spine

- RCTs (2 RCTs):
  A similarly low percentage of rhBMP-2 groups underwent reoperation compared with those in the ICBG groups as reported by three RCTs with a mean follow-up ranging from 17 to 24 months (4.2% versus 4.3%, respectively)[7, 26, 46].

  Details of the timing and causes of supplemental fixation are as follows:

  rhBMP-2:
  - Time not reported: decompression for leg pain (n = 1)[7]
  - Details not reported (n = 12)[46]

  ICBG:
  - Time not reported: pain pump insertion (n = 1)[26]
  - Details not reported (n = 11)[46]

- Cohort studies (2 studies)
  Pooled reoperation risks were similar in the rhBMP-2 and autograft groups as reported by two cohort studies (3.6% (4/112) versus 2% (1/57), respectively) at three to a mean of 27 months follow-up[14, 27]:

  Details of the timing and causes of revision are as follows:

  rhBMP-2:
  - Timing not reported, for the following reasons:
    - Nerve root deficit (n = 1)[27]
    - Malpositioned screws with radiculitis (n = 1)[14]
    - Ectopic bone formation with radiculitis (n = 1)[14]

  Autograft:
  - Timing not reported, for the following reasons:
    - Nerve root deficit (n = 1)[27]

rhBMP-7 off-label use: lumbar spine

- RCTs (1 RCT)
  Johnsson et al. (2002) reported identical risks of reoperation (10%) in both treatment groups by 12 months follow-up[20]. Two patients (one in each group) received decompression; no other details were reported.

BMP (unspecified) use in the lumbar spine

- Cohort studies (1 study)
  In a case-control study of the MarketScan base of BMP patients (and matched controls) with at least 12 months follow-up, Cahill et al. (2011) reported lower risks of repeat fusion at up to three years in the BMP group compared with the no BMP group (6.8% (161/2372) versus 9.2% (218/3272), respectively), with an adjusted hazard ratio of 0.74 (95% CI, 0.58, 0.93) (P = .01)[138]. Similarly, at one and two years, cumulative risks of repeat fusion were 2.3% and 5.2% in the BMP group compared with 3.4% and 6.6% in the control group, respectively, with an adjusted odds ratio of 0.66 (95% CI, 0.47, 0.94) (P = .03). Causes of revision were not reported.

rhBMP-2 off-label use: cervical spine

- Cohort studies (1 study)
  Buttermann et al. (2008) reported similar risks of reoperation in both treatment groups (0% for rhBMP-2 versus 3% (1/36) for ICBG)[11]. One control patient underwent reduction and fixation of a fracture.

Table 63. Reoperation: additional procedure at the index level besides a revision, hardware removal, or supplemental fixation

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP on-label use: lumbar spine					
Cohort studies					
rhBMP-2	1[23] (integrated analysis)	24 mos.	277	2.9% (8)	n/a
Control			402	8.0% (32)	n/a
BMP off-label use: lumbar spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	3[7, 26, 46]	17-24 mos.	309	4.2% (13)	0-5%
Control			281	4.3% (12)	0-4.9%
rhBMP-7	1[20]	12 mos.	10	10% (1)	n/a
Control			10	10% (1)	n/a
Cohort studies					
rhBMP-2	2[14, 27]	3-27 mos.	112	3.6% (4)	2-6%
Control			57	2% (1)	0-4%
BMP (unspecified)	1[138]	36 mos.	2372	6.8% (161)	n/a
Control			2372	9.2% (218)	n/a
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
Cohort studies					
rhBMP-2	1[11]	30 mos.	30	0% (0)	n/a
Control			36	3% (1)	n/a

n/a: not applicable; NR: not reported

* Mean calculated as: number of patients with adverse event ÷ total number of patients

Fusion at a different spinal level:

Summary:

On-label

There were no on-label studies reporting on this outcome.

Off-label

Risks of fusion at a different spinal level were similar between patients treated with rhBMP-2 and control in the lumbar spine as reported by two RCTs at 24 months follow-up (5.0% vs. 4.0%, respectively). In the cervical spine, risks were similar following rhBMP-2 (3.8%–5.6%) compared with controls (0%) at 24 to 30 months as reported by one small RCT and two small cohort studies. The strength of evidence is *moderate* for these estimates.

All data are summarized in Table 64.

Table 64. Fusion at a different spinal level

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range of means (%)
BMP off-label use: lumbar spine					
RCTs					
rhBMP-2	2[9, 26]	24 mos.	84	5% (4)	2-9%
Control			85	4% (3)	0-9%
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range of means (%)**
RCTs					
rhBMP-2	1[5]	24 mos.	18	6% (1)	n/a
Control			15	0% (0)	n/a
Cohort studies					
rhBMP-2	2[11, 34]	24-36 mos.	52	4% (2)	3-5%
Control			60	0% (0)	0%

n/a: not applicable; NR: not reported

* Mean calculated as: number of patients with adverse event ÷ total number of patients

Second surgeries (details not reported):

Summary:

On-label

There were no on-label studies reporting on this outcome. The strength of evidence for the risk of a second sugery is *insufficient* with respect to on-label use of rhBMP compared with control.

Off-label

Risks of any other unspecified types of second surgeries were similar between rhBMP-2 and control groups as reported by two cohort studies, one large database study investigating BMP use in the lumbar spine (10.8% vs. 10.5%, respectively) and one in the cervical spine (15.0% vs. 21.0%, respectively). The strength of evidence is *insufficient* regarding these estimates.

All data are summarized in Table 65.

Table 65. Second surgeries (details not reported)

	Studies	Length follow-up	Patients (n)	Mean (%)*	Range (%)
BMP off-label use: lumbar spine					
Cohort studies					
BMP	1[162]	48 mos.	1703	10.8% (183)	n/a
Control			15,119	10.5% (1588)	n/a
BMP off-label use: cervical spine					
	Studies	**Length follow-up**	**Patients (n)**	**Mean (%)***	**Range (%)**
Cohort studies					
rhBMP-2	1[35]	24 mos.	48	15% (7)	n/a
Control			156	21% (32)	n/a

n/a: not applicable; NR: not reported

* Mean calculated as: number of patients with adverse event ÷ total number of patients

4.3.9. Graft site morbidity

Summary: Following ICBG harvesting, hip pain VAS (0-10) ranged from 5.7-8.0 in the perioperative period as reported by four studies[6-9] and from 0.2-2.8 at 12-24 months (last follow-up) as reported by six studies[6-11]. The percentage of patients experiencing pain (definitions varied by study, see table for details) at last follow-up (6-36 months) ranged from 10-66% as reported by nine studies[2, 9, 10, 12-20]. Additional complications included injury to lateral femoral cutaneous nerve, ASIS fractures, superficial infection, deep infection requiring surgery, and hematoma. All data are summarized in Table 66.

Table 66. Graft site morbidity

Author	Study type	ICBG patients (N)	Hip pain	Unhappy with graft site appearance (%)	Other complications
Burkus 2002[6]	RCT	136	VAS (0-10)* Discharge: 6.4 3 mos.: 1.8 12 mos.: 2.1 24 mos.: 0.9	"very unhappy with appearance of graft site": Discharge: 9.7% 3 mos.: 2.2% 24 mos.: 2.6%	Any adverse event: 5.9% (8/136), including: • Injury to lateral femoral cutaneous nerve: 2.2% • Avulsion fractures of ASIS: 1.5% (2/136) • Infection (superficial): 0.7% • Hematoma: 0.7% (1/136)
FDA SSED InFUSE[22]	FDA report	139	NR	NR	Any adverse event: 5.8% (8 events in 8 patients) (details NR)
Boden 2002[7]	RCT	5	VAS (0-10)* Discharge: 8.0 Mean 17 mos.: 2.6	NR	NR
Burkus 2005[12]	RCT	52	% of patients with pain: 24 mos.: 47%	NR	NR
Dawson 2009[25]	RCT	21	NR	NR	Graft site infection: 5% (details NR)
Dimar 2009[8]	RCT	224	VAS (0-10)* Discharge: 5.7 24 mos.: 2.6 % of patients with pain (any): 24 mos.: 60% (108 reporting)	NR	Any adverse event: 7.6% (17/224)
Haid 2004[9]	RCT	33	VAS (0-10)* Discharge: 5.8 24 mos.: 2.8 % of patients with pain: 24 mos.: 60%	"appearance of graft site bothered them some": 24 mos.: 13%	Any adverse event: 6% (2/33), including: • Pain (n = 1) • Hematoma: (n = 1)
Mummaneni 2004[13]	Retro. cohort	33	% of patients with pain: 6 mos: 58% (mean pain score 5 (VAS 0-10) for these patients)	NR	NR
Howard 2011[44]‡	Cross-	53	VAS (0-10)	NR	NR

	sectional		Mean score for all f/u (0-84 mos.): 3.09 (range, 1.89-4.00)‡		
Rihn (2009)[14]	Retro. cohort	33	% of patients with persistent pain: mean 36 mos.: 30%	NR	• Donor site infection (required reoperation): 3%†
Taghavi (2010)[15]	Retro. cohort	20	% of patients with persistent pain: mean 28 mos.: 20%	NR	NR
Vaccaro 2004/2005/2008	RCT	12	% of patients with pain (any): 6 wks.: 58% 3 mos.: 73% 12 mos.: 60% 24 mos.: 66%	NR	NR
Vaccaro, Lawrence (2008)/Hwang (2010)	RCT	87	% of patients with pain (mild/moderate): 36+ mos.: 35%	NR	NR
Johnsson 2002	RCT	10	% of patients with pain: 12 mos.: 10%	NR	NR
Delawi 2010	RCT	16	VAS (0-10) 6 wks.: 3.0 3 mos.: 1.7 6 mos.: 3.8 12 mos.: 2.7 % of patients with pain: 12 mos.: 64% ("mild")	NR	"Complications directly related to ICBG harvesting": 0%
Baskin 2003	RCT	15	Data NR At 24 mos. f/u "some ICBG patients continued to experience residual pain"	Data NR At 24 mos. f/u "some ICBG patients… rate[d] appearance of site as only fair"	NR
Buttermann (2008)	Prosp. cohort	36	VAS (0-10) 12 mos.: 0.2		• Infection (required irrigation & debridement): 3% • ASIS fracture (with elective open reduction and

					internal fixation): 3%
Crawford (2009)	Retro. cohort	36	NR	NR	• Deep infection of site (required operation): 3%†

ASIS: anterior superior iliac spine; n/a: not applicable; NR: not reported
*results converted to a scale ranging from 0-10, with 10 indicating greater pain.
† Adverse events included in reoperation for infection complications table as similar information was provided for the investigational groups.
‡ Howard (2011) harvested the ICBG through a midline lumbar incision (no scar over graft site). Of the 53 patients who underwent ICBG graft harvesting, only a subset of patients (9-60%) reported graft pain scores at any given follow-up. We thus reported the mean and range of scores reported for all follow-ups from < 12 months to 84 months.

4.4. Key Question 4: Differential efficacy or safety in subpopulations

What is the evidence that on- or off-label use of rhBMP-2 or rhBMP-7 for spinal fusion compared with spinal fusion using ICBG or alternative bone graft substitutes has differential efficacy or safety issues in sub-populations? Including consideration of:

- Gender
- Age
- Baseline functional or pain status
- Comorbidities (including but not limited to tobacco use, alcohol use, psychological or psychological)
- Other patient characteristics or evidence-based patient selection criteria
- Provider type, setting or other provider characteristics
- Payor/ beneficiary type: including worker's compensation, Medicaid, state employees

Summary: We found no strong evidence of the differential effectiveness of spinal fusion using rhBMP-2 or rhBMP-7 versus spinal fusion using ICBG or alternative bone graft substitutes in any subpopulation. Although eight studies examined outcomes in various subpopulations, none of these studies pre-specified the subgroup analyses, none of the studies performed a test of interaction as the method of subgroup analysis, and some of the studies were inadequately powered to detect differences in treatment effect. In general, fusion without rhBMP tended to have lower complication risks and fusion with rhBMP tended to have better radiographic outcomes across most subpopulations examined, although in many cases the differences were small.

Subpopulations

We identified eight cohort studies[15, 28, 32, 60, 67, 160-162] (LoE II: 1 study; LoE III: 7 studies) comparing the differential effectiveness of spinal fusion using rhBMP-2 or rhBMP-7 versus spinal fusion using ICBG or alternative bone graft substitutes in subpopulations characterized by: age, sex, smoking status, number of levels treated, complexity of fusion, surgical approach (anterior or posterior), and previous surgeries. None of the studies performed a test of interaction as the method of subgroup analysis and all subgroup analyses were specified post-hoc. We report the results from these eight studies (summary results in Table 67 and detailed data abstraction tables in Appendix F).

Age

There is no evidence of differential effect of rhBMP-2 use and age with respect to overall surgical and perioperative complications. One retrospective cohort (database) study[160] compared surgical and perioperative complications following spinal fusion with and without rhBMP-2 in adult and pediatric scoliosis patients. In general, both adult and pediatric patients *not* receiving rhBMP-2 experienced slightly lower overall complication risks than patients receiving rhBMP-2 (9.3% versus 13.8% in the adult population and 7.0% versus 8.3% in the pediatric population, respectively).

Epidural hematoma/seroma occurred rarely in each population: 0.1% *with* rhBMP-2 versus 0.3% *without* rhBMP-2 in the adult patients, and 0.2% *with* rhBMP-2 versus 0.1% *without* rhBMP-2 in the pediatric patients. There was no evidence of differential effect with respect to superficial infection. Though the risk of deep infection was slightly higher in adults receiving fusion *without* versus *with* rhBMP-2 (2.0% versus 1.8%) and slightly higher in pediatric patients receiving fusion *with* versus *without* rhBMP-2 (1.6% versus 1.3%), the differences are quite small.

Sex

There was no evidence of differential effect of rhBMP-2 use and sex with respect to radiographic outcomes in one retrospective cohort study[28]. In general, male patients had better outcomes *with* rhBMP-2 and females had slightly better outcomes with ICBG (no rhBMP-2).

Smoking status

Two retrospective cohort studies[28, 67] compared radiographic, pain, and function outcomes following spinal fusion with and without rhBMP-2 in smokers and non-smokers. For most outcome measures, patients receiving rhBMP-2 tended to have better outcomes than those not receiving rhBMP-2 regardless of smoking status.

In one study both smokers and non-smokers receiving fusion *with* rhBMP-2 had lower non-union risks (0% versus 7.8%) than patients receiving fusion *without* rhBMP-2 (40% versus 10%)[28].

Another study examined fusion risks using two different fusion criteria, IDE (investigational device exemption) and CT bridging bone[67]. At both follow-up periods (12 and 24 months), both smokers and non-smokers receiving fusion *with* rhBMP-2 had higher fusion risks than patients receiving fusion *without* rhBMP-2. For example, the fusion risk as measured by CT bridging bone at 24 months for patients receiving fusion *with* rhBMP-2 compared with fusion *without* rhBMP-2 was 95.0% versus 75.0% in smokers and 98.1% versus 90.2% in non-smokers. There was no evidence of differential effect regarding CT grade or improvement in pain or function scores.

Number of levels treated

There is no evidence of differential effect of rhBMP-2 use and number of levels treated with respect to time to solid fusion. One retrospective study compared fusion outcomes following spinal fusion with rhBMP-2, BMAA (Bone Marrow Aspirate with Allograft), or autograft[15]. In general, patients receiving fusion at one- or multi-levels *with* rhBMP-2 had lower time to solid fusion (199.8 versus 240.4 days for one-level and multi-level fusion, respectively) than patients receiving BMAA (313.3 versus 282.0 days) or autograft (276.7 versus 263.3 days). There was no evidence of differential effect of rhBMP-2 use and number of levels treated with respect to non-union or fusion risk in this study or one other study[32]. There was also no evidence of differential effect with respect to the retrograde ejaculation risk in one- and two-level fusions in another study[60].

Complexity of fusion

There was no evidence of differential effect of rhBMP-2 use and complexity of surgery with respect to repeat surgeries within one to four years of the index surgery in one large cohort study[162].

Surgical approach (anterior or posterior)

One large retrospective cohort (database) study[161] compared surgical and perioperative complications following spinal fusion with and without rhBMP-2 in patients receiving anterior or posterior cervical surgery. Both anterior and posterior cervical patients *not* receiving rhBMP-2 experienced a lower overall complication risk than patients receiving rhBMP-2 (4.68% *without* rhBMP-2 versus 7.09% *with* rhBMP-2 in the anterior cervical population and 9.95% versus 10.04% in the posterior cervical population). In addition, anterior and posterior cervical patients *not* receiving rhBMP-2 experienced a slightly lower dysphagia/hoarseness risk than patients receiving rhBMP-2 (2.45% *without* rhBMP-2 versus 4.35% *with* rhBMP-2 in the anterior cervical population and 1.63% versus 2.09% in the posterior cervical population). There was no evidence of differential effect of rhBMP-2 use and surgical approach with respect to wound complications in this study.

Previous surgeries

There is no evidence of differential effect of rhBMP-2 use and previous surgery with respect to the risk of repeat operations. One large retrospective cohort (database) study[162] compared the risk of repeat surgeries within one to four years of the index surgery following spinal fusion with and without rhBMP-2 in patients with and without a history of previous surgery. In general, though patients without a history of previous surgery experienced a lower risk of repeat surgeries compared to patients with a history of previous surgery irrespective of whether they received fusion with or without rhBMP-2, the differences are quite small.

Table 67. Differential efficacy or safety in various subpopulations.

Study	Outcome	Subpopulation	rhBMP2 (unless otherwise indicated)	Comparator 1	Comparator 2	Test of interaction
			(mean ± SD (range) unless otherwise indicated)			
Radiographic Outcomes						
Glassman, Carreon (2007)[28] * comparator 1 = ICBG	Non-union (% patients)	Male Female	11.1% (4/36) 3.6% (2/55)	26% (NR) 0% (NR)	n/a n/a	no
		Smokers Non-smokers	0% (0/14) 7.8% (6/77)	40% (2/5) 10% (3/30)	n/a n/a	no
Slosar (2007)[32] comparator 1 = autograft	Non-union (% patients)	1 level 2 level 3 level	0% (0/10) 0% (0/26) 0% (0/9)	11% (1/9) 13% (2/15) 33% (2/6)	n/a n/a n/a	no no
Taghavi (2010)[15] comparator 1=BMAA, comparator 2= autograft	Fusion rate (% patients)	1-level Multi-level	100% (13/13) 100% (11/11)	100% (7/7) 63.6% (7/11)	100% (10/10) 100% (10/10)	no
Glassman, Dimar (2007)[67] † comparator 1= ICGB	Fusion rate at 12 months (% patients) (IDE criteria)	Smokers Non-smokers	94.7% (NR) 96.3% (NR)	75.0% (NR) 89.6% (NR)	n/a n/a	no
	Fusion rate at 24 months (% patients) (IDE criteria)	Smokers Non-smokers	95.2% (20/21) 100.0% (55/55)	76.2% (16/21) 94.1% (48/51)	n/a n/a	no
	Fusion rate at 12 months (% patients) (CT bridging bone)	Smokers Non-smokers	94.4% (NR) 94.4% (NR)	73.7% (NR) 83.3% (NR)	n/a n/a	no
	Fusion rate at 24 months (% patients) (CT bridging bone)	Smokers Non-smokers	95.0% (19/20) 98.1% (52/53)	75.0% (15/20) 90.2% (46/51)	n/a n/a	no
Taghavi (2010)[15] comparator 1=BMAA, comparator 2= autograft	Time to solid fusion (days)	1-level Multi-level	199.8 ± 49.8 240.4 ± 71.3	313.3 ± 34.3 282.0 ± 87.5	276.7 ± 29.8 263.3 ± 79.4	no
Glassman, Carreon (2007)[28] * comparator 1 =	CT grade	Male Female	4.04 ± NR 4.61 ± NR	3.75 ± NR 4.69 ± NR	n/a n/a	no
		Smokers	4.32 ± NR	3.20 ± NR	n/a	no

Study	Outcome	Subpopulation	rhBMP2 (unless otherwise indicated)	Comparator 1	Comparator 2	Test of interaction
			(mean ± SD (range) unless otherwise indicated)			
ICBG		Non-smokers	4.40 ± NR	4.33 ± NR	n/a	
Pain						
Glassman, Dimar (2007)[67] † comparator 1= ICGB	Improvement in ODI score from pre-operative score at 24 months	Smokers Non-smokers	22.1 26.4	21.0 24.6	n/a n/a	no
Function						
Glassman, Dimar (2007)[67] † comparator 1= ICGB	Improvement in SF-36 PCS score from pre-operative score at 24 months	Smokers Non-smokers	7.1 10.2	11.6 11.2	n/a n/a	no
Surgical and perioperative complications						
Cahill (2009)[161] rhBMP (any) comparator 1= fusion without rhBMP2	Overall complication risk (% patients)	Surgical approach: anterior cervical Surgical approach: posterior cervical	7.09% (163/2299) 10.04% (48/478)	4.68% (1158/24768) 9.95% (238/2391)	n/a n/a	no
	Dysphagia/ hoarseness risk (% patients)	Surgical approach: anterior cervical Surgical approach: posterior cervical	4.35% (100/2299) 2.09% (10/478)	2.45% (608/24768) 1.63% (39/2391)	n/a n/a	no
	Wound complication risk (% patients)	Surgical approach: anterior cervical Surgical approach: posterior cervical	1.22% (28/2299) 2.93% (14/478)	0.65% (160/24768) 2.51% (60/2391)	n/a n/a	no
Williams (2011)[160] ‡ rhBMP (any) comparator 1= fusion without rhBMP2	Overall complication risk (% patients)	Adult scoliosis Pediatric scoliosis	13.8% (124/899) 8.8% (139/1576)	9.3% (425/4586) 7.0% (1310/15937)	n/a n/a	no
	Epidural hematoma/seroma risk (% patients)	Adult scoliosis Pediatric scoliosis	0.1% (1/899) 0.2% (3/1576)	0.3% (13/4586) 0.1% (20/15937)	n/a n/a	no
	Superficial infection risk (% patients)	Adult scoliosis Pediatric	1.3% (12/899) 1.1% (18/1576)	0.9% (42/4586) 0.7% (138/15937)	n/a n/a	no

Study	Outcome	Subpopulation	rhBMP2 (unless otherwise indicated) (mean ± SD (range) unless otherwise indicated)	Comparator 1	Comparator 2	Test of interaction
		scoliosis				
	Deep infection risk (% patients)	Adult scoliosis	1.8% (16/899)	2.0% (90/4586)	n/a	no
		Pediatric scoliosis	1.6% (26/1576)	1.3% (235/15937)	n/a	
Adverse Events						
Carragee, Mitsunaga (2011)[60] comparator 1= no rhBMP2	RE complication risk (n, % patients, 90% CI)	1-level	6.7% (3/45, 0.55 – 12.79)	0% (0/110, < 2.4)	n/a	no
		2-level	8.3% (2/24, -0.95 – 17.61)	1.6% (1/64, -0.99 – 4.11)		
Second Surgeries						
Deyo (2011)[162] § rhBMP (any) comparator 1= no rhBMP2	Repeat surgery within 1 year of index surgery (n, % patients)	Previous surgery	3.8% (14/366)	4.6% (100/2181)	n/a	no
		No previous surgery	2.4% (32/1337)	2.7% (343/12938)	n/a	
		Simple fusion	2.6% (26/1014)	2.8% (307/10792)	n/a	no
		Complex fusion	2.9% (20/689)	3.1% (136/4327)	n/a	
	Repeat surgery within 2 years of index surgery (n, % patients)	Previous surgery	8.2% (30/366)	8.5% (186/2181)	n/a	no
		No previous surgery	5.8% (77/1337)	5.6% (726/12938)	n/a	
		Simple fusion	6.1% (62/1014)	5.8% (630/10792)	n/a	no
		Complex fusion	6.5% (45/689)	6.5% (282/4327)	n/a	
	Repeat surgery within 3 years of index surgery (n, % patients)	Previous surgery	12.3% (45/366)	12.1% (264/2181)	n/a	no
		No previous surgery	8.4% (112/1337)	7.9% (1023/12938)	n/a	
		Simple fusion	8.9% (90/1014)	8.2% (881/10792)	n/a	no
		Complex fusion	9.7% (67/689)	9.4% (406/4327)	n/a	
	Repeat surgery within 4 years of index surgery (n, % patients)	Previous surgery	14.5% (53/366)	14.9% (325/2181)	n/a	no
		No previous surgery	9.7% (130/1337)	9.8% (1263/12938)	n/a	
		Simple fusion	10.0% (101/1014)	10.3% (1092/10792)	n/a	no
		Complex fusion	11.9 (82/689)	11.5% (496/4327)	n/a	

SD: standard deviation; ICBG: Iliac Crest Bone Graft; IDE: Investigational Device Exemption; ODI: Oswestry Disability Index; SF-36: Short-Form 36; PCS: Physical Component Summary; RE: Retrograde Ejaculation; CI: Confidence Interval; BMAA: Bone Marrow Aspirate with Allograft

* rhBMP2 group is a mixture of one-level (n = 61) and two-level (n=30) treatments, ICBG control group is one-level treatment only (n = 35). CT grade based on the following criteria: grade 1 (no fusion) and grade 2 (partial unilateral fusion) defined as non-union; grade 3 (partial bilateral fusion) defined as probably fusion; grades 4 and 5 (solid unilateral or bilateral fusion) defined as definite fusion[28].
†Fusion success is defined by the IDE protocol as bilateral bridging trabecular bone on plain radiographs with less than 3° of translation and less than 5° of angulation on flexion-extension views; defined by CT scan criteria as presence of contiguous bridging bone on fine cut CT scan with coronal and sagittal reconstructions[67].
‡Authors focused on intraoperative and immediate postoperative complications, including death, new neurological deficit, wound infection (superficial or deep), pulmonary embolus, deep venous thrombosis, other pulmonary complications, implant related, peripheral nerve deficit, visual deficit, and epidural hematoma. Epidural hematoma and seroma complications are grouped together as “epidural hematoma/seroma”. Scoliosis patients are separated into adult (≥ 21 years) and pediatric[160].
§Previous surgery is defined as having had lumbar surgery prior to the index operation; repeat surgery is defined as any reoccurrence of lumbar surgery following the index operation, with the nature of surgery and spinal levels unknown. Simple fusion is defined as anterior fusion, transverse process or posterior fusion involving one or two disc levels, or an unreported number of disc levels; complex fusion is defined as 360-degree spine fusion by single incision, any combination of anterior with either transverse process or posterior fusion, or any fusion involving more than two disc levels[162].

4.5. Key Question 5: Cost-effectiveness

What evidence of cost implications and cost-effectiveness of on- or off-label use of use of rhBMP-2 or rhBMP-7 exists? Including consideration of:

- Costs (direct and indirect) and cost effectiveness
- Short term and long term

Summary:

There is conflicting evidence about whether the use of rhBMP-2 for on-label lumbar spinal fusion results in better outcomes and/or lower costs than control or standard treatment.

- One well conducted cost effectiveness analysis performed by the AHRQ suggested that when analyzed as part of the treatment cost, on-label use of rhBMP-2 for lumbar spinal fusion results in lower costs per QALY only when it is assumed to be a part of the Medicare reimbursement and no cost differential is calculated. However the more common payer strategy assumes the cost of rhBMP-2 is added to treatment, in which case the group treated with rhBMP-2 had higher QALYs and higher cost, a common outcome for new technologies.
- One CUA concluded on-label use of rhBMP-2 for lumbar spinal fusion unlikely to be cost-effective due to higher costs and similar outcomes compared with the control group.

One cost-utility study showed that off-label use of rhBMP-2 was more cost-effective than ICBG for posterolateral spinal fusion in patients at least 60 years of age.

- One moderately well conducted cost utility analysis determined that off-label use of rhBMP-2 in posterolateral spinal fusion was associated with similar efficacy and somewhat lower complication risks compared with ICBG, resulting in a decreased overall cost of $2319.

No studies were identified that evaluated the cost-effectiveness of rhBMP-2 for use in the cervical spine.

No studies were identified that evaluated the cost-effectiveness of rhBMP-7 use in the spine.

Background and context

We conducted a search for cost effectiveness and cost utility studies that evaluated on- or off-label uses of rhBMP-2 or rhBMP-7 for spinal fusion. Three[39, 42, 182] studies of the nine evaluated at full-text review met our inclusion criteria and are described below. All three studies evaluated rhBMP-2; no studies were identified that assessed the cost effectiveness of rhBMP-7. Details of the included studies can be found in Appendix F.

Results

AHRQ BMP HTA cost utility analysis (CEA):
The AHRQ assessed the cost effectiveness of rhBMP-2 as part of their HTA evaluating BMP use in spinal fusion. Outcomes data were taken from the randomized trial by Burkus et al., (2002)[6], which served as the pivotal trial for FDA approval of InFUSE (Medtronic). In this trial, patients with single-level DDD and disabling symptoms of at least six months duration underwent single-level open ALIF with an LT-Cage device filled with either InFUSE (an absorbable collagen sponge infused with rhBMP-2) (n = 143) or ICBG (n = 136) and were followed for two years.

AHRQ's cost utility analysis utilized stationary Markov models, which assess the impact of changes in health states over time. A cycle length of one week was used. The analysis was based on three health states for the rhBMP-2 group (pre-fusion, secondary intervention, and fusion) and six for the ICBG group (same three health states as the BMP group plus the presence or absence of donor site pain). The minimum time to fusion and union was assumed to be six weeks. Utility estimates for pre-fusion without donor site pain (S1) and fusion without donor site pain (S6) were based on preoperative and 6-month unpublished SF-36 data (collected by Burkus (AHRQ refs 72, 182) and described in the NHS cost effectiveness analysis[42] (see below). Utility and relevant outcomes data are summarized as follows:

- S1: Pre-fusion without donor site pain (utility = 0.54) (source: unpublished data).
- S2: Pre-fusion with donor site pain (utility = 0.52) (source: S1 utility reduced by 0.02)
- S3: Secondary intervention without donor site pain (utility = 0.49) (source: S1 utility reduced by 0.05)
- S4: Secondary intervention with donor site pain (utility = 0.47) (source: S3 utility reduced by 0.02)
- S5: Fusion with donor site pain (utility = 0.60) (source: S6 utility reduced by 0.02)
- S6: Fusion without donor site pain (utility = 0.62) (source: unpublished data)

- Fusion rates were similar between groups and decreased slightly over the course of the study: in the rhBMP-2 group, 97.0%, 96.9%, and 94.5% of patients had radiographic evidence of fusion at 6, 12, and 24 months, respectively, compared with 95.8%, 92.6%, and 88.7% of patients in the ICBG at the same follow-ups.

- Pre-fusion probabilities were derived from radiographic fusion success probabilities; no data were reported.

- The percentage of ICBG patients experiencing donor site pain decreased from 100% at discharge to 83% at 6 weeks, 56% at 3 months, 43% at 6 months, 35% at 12

months, and 32% at 24 months. Probabilities of donor site pain for each week were derived from these data.

- 12.6% of the rhBMP-2 group underwent secondary interventions compared with 13.2% of the control group; the risk ratio was 0.95 (95% CI, 0.51, 1.75). Undergoing a secondary intervention was considered to be a temporary state by the AHRQ that lasted one week. The area spend in the secondary intervention state was calculated by dividing the percentage of patients undergoing secondary interventions by the number of weeks past the minimum time to fusion.

The analysis was performed from the (CMS) payer perspective and used direct costs as reported by Medicare during the year 2007. Cost categories included initial hospitalization and secondary interventions (including removals, supplemental fixations, and reoperations).

The quality adjusted life year (QALY) serves as the effectiveness measure and was calculated from the Burkus data, based on fusion, secondary intervention to obtain spinal fusion or to address complications, and donor-site pain.

Incremental cost-effectiveness ratio (ICER) was calculated as the difference in total costs between treatment and control divided by the difference in QALYs.

Results:

Base case: assumes that initial hospital costs are identical in both groups (i.e., that rhBMP-2 cost was treated as a bundled part of the diagnosis-related groups (DRG) payment system and thus incurred no additional cost). RhBMP-2 was found to be the dominant treatment strategy and was associated with cost savings of $94 and an increase of 0.024 QALYs (over 24 months) compared with ICBG.

- Total cost (104 weeks): rhBMP-2: $31,159; ICBG: $31,253.
- Lower cost for the rhBMP-2 group was attributed to the lower probability of secondary intervention (removals, supplementary fixations, and reoperations).
- Total QALYs: rhBMP-2: 1.218; ICBG: 1.194.
- Higher QALY for the rhBMP-2 group was attributed to donor site pain in the control group.

One-way sensitivity analyses: A number of one-way sensitivity analyses were performed on the base case (i.e., BMP incurring no additional cost). The estimate was sensitive to risk ratio and cost of BMP. BMP remained the dominant intervention under variation of hazard ratios of fusion and risk ratio of secondary interventions.

- Upper value of the risk ratio for secondary interventions used, rhBMP-2 was not the dominant treatment strategy and was associated with an increased cost of $2,153 and an ICER of $89,765 per QALY gained.

- ***BMP as added cost ($3000):*** When rhBMP-2 was assumed to incur an additional cost of $3000 (a mean estimate) it was no longer the dominant treatment strategy. In this scenario, rhBMP-2 was associated with a cost increase of $2,906 over ICBG and an ICER of $121,160 per QALY gained.

Two-way sensitivity analyses were also performed by varying utility values as done in one-way sensitivity analysis (see above) and treating rhBMP-2 as an added cost of $3000.

- In no scenario was rhBMP-2 the dominant treatment. The lowest ICER occurred when the disutility of donor-site pain was assumed to be larger (decrease of utility value by 0.05 versus 0.02 as done in the base case), and was associated with a cost of $56,959 per QALY gained, with an assumed cost of BMP of $3000.
- Decreasing the assumed cost of BMP to $1000 lowered the ICER to $37,785 per QALY gained.
- Increasing the assumed cost of BMP to $5000 and $8000 increases the ICERs to $204,536 and $329,599 per QALY gained, respectively.

Based on the results of the cost effectiveness analysis, the authors recommend rhBMP-2 as a treatment strategy only when rhBMP-2 is not an added cost. The majority of one-way sensitivity analyses of the base case similarly found that rhBMP-2 was associated with lower costs and increase QALYs. However, when treated as an additional cost, rhBMP-2 was no longer recommended due to the fact that the cost-effectiveness ratio is very sensitive to cost. The relatively low increase in QALYs (0.018–0.051) coupled with the increased costs meant that the ICERs for rhBMP-2 use were high (range: $37,785 – $329,599 per QALY gained, depending on the cost of BMP and the analysis done).

This is a well-conducted study (QHES score: 95/100). Its strengths are in its use of clinical trial data, inclusion of states in which patients experienced donor site pain, as well as the fact that because it is a Markov model, it takes into account changes in health states over time. One weakness is the use of free publicly available CMS cost estimates; limited access cost sources may provide more accurate cost estimates.

NHS HTA (UK) cost utility analysis (2007)[42]:
As part of a systematic review, National Health Services (NHS) (UK) assessed the cost-effectiveness of rhBMP-2 by modifying an economic model provided by ABACUS International. ABACUS is a European consulting firm that specializes in health economics. Medtronic funded development of the economic model.

The original economic model that ABACUS provided to the NHS was generated in 2005 and modified in 2006. Data from this model were based primarily on the Burkus 2003 integrated analysis[23] of 679 patients (including data from one published RCT (Burkus 2002[6]; N = 279, the same data utilized by the AHRQ analyses) and two unpublished nonrandomized studies (N = 400)) who received InFUSE or ICBG in anterior open or laparoscopic spinal fusions. The ABACUS model considered the following: the number of anterior level fusions per year in England (based on a published source), the estimated costs at the time for a standard anterior open or laparoscopic spinal fusion (based on the National Tariff HRG Code for

spinal fusion surgery) with or without rhBMP-2 (additional cost of £1790 per case provided by Wyeth Pharmaceuticals), surgical parameters (operating room time, hospital stay), risks and costs of repeat operation (revisions, removals, supplemental fixations, and reoperations; based on data from Burkus[23]), preoperative work status and time to return to work[23], fusion rates[23], and unpublished utility values based on reported SF-36 scores[23]. This original model estimated that for the approximately 1000 anterior open or laparoscopic spinal fusion procedures conducted each year in England, use of BMP-2 instead of ICBG would decrease operating room time by 900 hours, length of stay by 1143 days, and the number of reoperations by 78. By 24 months, BMP-2 use was associated with 56 more QALYs (the cost per QALY was not reported). The method by which the QALY was calculated was not specified. Furthermore, because patients in the integrated analysis who received rhBMP-2 returned to work significantly earlier than those who were treated with ICBG (116 versus 154 days; $P = .0156$), the ABACUS model concluded that rhBMP-2 use saved approximately £4.5 million in paid sick leave. Overall, this model suggested that use of rhBMP-2 would result in a final annul savings to the UK of over £4 million (£4393 per patient).

Using the ABACUS model, the NHS conducted a modified cost-utility analysis of rhBMP-2. In the updated model, data (surgery parameters, reoperation risks, return to work, and fusion rates) were based on the Burkus 2002 RCT[6], which evaluated use of InFUSE compared with ICBG in open (but not laparoscopic) spinal fusions (N = 279). Costs were updated. Return to work rates were adjusted based on preoperative work status.

The modified ABACUS model estimated that for an estimated 1024 spinal fusion procedures performed annually in England, use of rhBMP-2 compared with ICBG would decrease operating room time by 410 hours, length of stay by 205 days, and the number of reoperations by 7. Based on data from the Burkus 2002 RCT[6], patients treated with rhBMP-2 tended to return to work later than patients who received ICBG (59% versus 64% of patients after adjusting for preoperative work status; how these estimates were derived was not reported). Over a 24 month period, use of rhBMP-2 was associated with an additional 11 QALYs compared with ICBG. The results suggested that the use of rhBMP would be associated with a final (adjusted) cost that was £1.3 million per year higher than that associated with ICBG. Thus, the cost per QALY gained with rhBMP-2 was estimated at £120,390. The probability that rhBMP is cost-effective at a willingness to pay threshold of £30,000 per QALY was 6.4%. This is a very low probability given the cost.
The authors determined that given the economic evaluation, the use of rhBMP for spinal fusion is associated with higher costs compared with ICBG. The AHRQ report noted that the "way in which the ABACUS models calculated QALYs is opaque and would be difficult to reproduce. A request to examine the MS Excel files used by ABACUS before completion of this [AHRQ] analysis was declined". The strengths of this study were its use of several data sources; weaknesses included lack of further information about models used and way calculations were carried out. This was a moderately conducted study (QHES score: 72/100).

Carreon (2009) cost utility analysis (CUA)[182]:
The authors performed a cost-utility analysis using actual cost data from their randomized clinical trial of patients over age 60 who were treated with either InFUSE (rhBMP-2/ACS) (n

= 50) or ICBG (n = 52) for single or multilevel posterolateral lumbar fusions. The mean number of fused levels per patient was 1.96 versus 1.98, respectively. Improvement in scores for the ODI, SF-36, SF-6D, and back and leg pain scores over two years were similar between groups (no significant differences). Rates of fusion and revision similar were not significantly different between groups. The percentage of patients with complications requiring medical or surgical treatment was similar in both groups (12% (BMP) vs. 15% (ICBG); $P = .775$), however more patients in the ICBG group had persistent symptoms (i.e., low back or leg pain) that required additional treatment (20% vs. 38%; $P = .051$).

The cost utility analysis was based on actual costs (both inpatient and outpatient) and included actual reimbursements. The mean total two-year cost for care per patient (without the costs for complication and additional treatment) was $2295 more for rhBMP-2 versus ICBG, with a total per patient cost of $36,530 for the index procedure in the rhBMP-2 group and that of $34,235 in the ICBG group. The mean cost to treat a major complication was $10,888; the mean cost for revision surgery for non-fusion was $46,852, and the mean cost for additional treatment for spine-related events was $5,892.

Preoperative and 2-year postoperative utility scores were determined using the SF-6D, derived from the SF-36 and valued by a representative sample of the UK general population. The SF-6D scores were similar between groups at all follow-ups. A decision tree was created and based on the probability of complications, need for additional treatments and revision surgery, costs associated with initial treatment and treatment for complications, and utility scores.

Results of the cost utility decision tree analysis showed that in RCT of patients who were at least 60 years of age and underwent single- or multilevel posterolateral spinal fusion:

- The cost of using rhBMP-2 was $39,967 with 0.11 mean improvement in the SF-6D. No data were provided on the clinical significance of this change.
- The cost of using ICBG was $42,286 with a mean improvement of 0.10 in SF-6D

Thus, results from the cost utility analysis suggest that rhBMP-2 results in lower costs than ICBG for posterolateral fusion when used in patients 60 years of age or older. The authors noted that the costs associated with treating complications and managing persistent back or leg pain symptoms were considerably higher in the ICBG group than in the rhBMP-2 group and offset the initial greater cost of using rhBMP-2. This is a relatively well-conducted study (QHES score = 86/100). Its strength is its use of randomized clinical trial data. The weaknesses include simplifying the economic analysis by assuming a single payer model and exclusion of indirect costs. Furthermore, rhBMP-2 use was off-label, as InFUSE has not received FDA approval for posterolateral fusion.

5. Summary by Key Question

Information on determination of overall strength of evidence (SoE) is found in Appendix D. Summaries for individual key questions are found in the Executive Summary and in the corresponding sections of the report. The following tables summarize the overall strength of evidence for each key question.

Table 68. Summary of evidence for Key Question 1

Key Question 1: Validated instruments for measuring treatment outcomes					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
Measures		The most commonly used outcome measures in comparative studies evaluating BMP use in lumbar or cervical spinal fusion were identified. The following outcome measures have undergone psychometric analysis in spine patients: Measures: • ODI (18 studies) • SF-36 (17 studies) • Pain assessed by VAS (14 studies) • NDI (1 cervical study)			
Validity, reliability, and responsiveness		One outcome measure (SF-36) has been shown to have criterion validity and reliability in patients undergoing spinal fusion by one study. All four outcome measures have been shown to have a degree of validity, reliability, and responsiveness in various spine populations, some of which might be eligible for fusion.			
MCID		For the ODI (scale 0-100), the MCID has been variously defined in fusion patients as 10-22.9 depending on the study population and calculation method. However, there is some cause for concern regarding the definition. For VAS pain (scale 0-10), the MCID has been defined by one study as 1.8-1.9. However, there is some cause for concern regarding the definition. No studies were found that examined the MCID of the SF-36 or NDI in any spine population.			

MCID: minimal clinically important difference; NDI: Neck Disability Index; ODI: Oswestry Disability Index; SF-36: Short-Form 36; SoE: Strength of Evidence; VAS: Visual Analogue Scale

Table 69. Summary of evidence for Key Question 2

Key Question 2: Efficacy and effectiveness					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
On-label use of rhBMP-2 in the lumbar spine					
Efficacy		***Study characteristics*** • Evidence base: 2 RCTs[6, 21] (LoE IIb). Study size ranged from 14 to 279 patients. • Interventions: Primary single-level open anterior lumbar fusion with either rhBMP-2/ACS (InFUSE) (n = 154) or iliac crest bone autograft (ICBG) (n = 139). RhBMP-2 was used at a dose ranging from 4.2-8.4 mg/patient. • Population: Patients with DDD, radiculitis, and/or up to 25% spondylolisthesis who were refractory to conservative care. • Length of follow-up: 24 months. • Sponsorship: Both studies were sponsored by Medtronic. • These studies served as the pilot and pivotal trials in the 2002 FDA Summary of Strength and Effectiveness Data (SSED) for InFUSE (P000058) [22]. • The studies were similar in design, thus we were able to pool outcomes data. • Additional details: Table 8 and surrounding text.			
	Low	***Conclusions*** The following outcomes were similar in both treatment groups: mean operative time (2 RCTs), length of hospital stay (2 RCTs), fusion (2 RCTs), ODI outcomes (2 RCTs), back and leg pain outcomes (1 RCT), SF-36 physical function scores (1 RCT), patient satisfaction (2 RCTs), return to work (2 RCTs), and neurological success (1 RCT). The following outcomes were improved in patients treated with rhBMP-2 compared with ICBG: perioperative blood loss (2 RCTs).	+	-	-
Effectiveness		***Study characteristics*** • Evidence base: 1 integrated analysis[23] (LoE II) based on the following studies: one RCT[6] (reported in efficacy) (n = 279), one case series[24] (n = 22), and one unpublished study (n = 378). • Interventions: Primary single-level open (41%) or laproscopic (59%) anterior lumbar fusion with either rhBMP-2/ACS (InFUSE) (n = 277) or iliac crest bone autograft (ICBG) (n = 402).			

Key Question 2: Efficacy and effectiveness					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
		• Population: Patients with DDD and radiculitis who were refractory to conservative care. • Length of follow-up: 24 months. • Sponsorship: Both studies have been reported to be sponsored by Medtronic. • Additional details: Table 13 and surrounding text.			
	Low	***Conclusions*** The following outcomes were improved in patients treated with rhBMP-2 compared with ICBG: perioperative outcomes (operating time, blood loss, and length of hospital stay). The following outcomes were similar in both treatment groups: fusion, ODI outcomes, SF-36 pain index and physical component subscale scores, and return to work.	+	-	-
On-label use of rhBMP-7 in the lumbar spine					
Efficacy		No studies were identified that evaluated the efficacy of on-label use of rhBMP-7 in the lumbar spine.			
Effectiveness		No studies were identified that evaluated the effectiveness of on-label use of rhBMP-7 in the lumbar spine.			
Off-label use of rhBMP-2 in the lumbar spine					
Efficacy		***Study characteristics*** • Evidence base: 6 RCTs[1, 7-9, 12, 25, 26]: LoE IIa (1 study)[1, 12], LoE IIb (5 studies)[7-9, 25, 26]. Study size ranged from 27 to 463 patients. • Interventions: Various. Patients underwent primary single- (or in one study, multi-) level posterior (four studies), anterior (one study), or posterolateral (one study) lumbar fusion with either rhBMP-2/ACS (InFUSE) or iliac crest bone autograft (ICBG). Due to heterogeneity in surgical procedures (i.e., approach, use of ceramic granules, use of cage versus allograft dowel versus no device, single- versus multilevel fusion), we did not pool outcomes data from the six studies. Patients received BMP in a variety of forms: rhBMP-2/CRM; InFUSE; and AMPLIFY. Doses of rhBMP-2 varied and ranged from 4.2-40 mg per patient (when reported). • Population: Patients with DDD, radiculitis, and/or up to 25% spondylolisthesis who were refractory to conservative care. • Length of follow-up: 17 (mean) - 24 months.			

Key Question 2: Efficacy and effectiveness					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
		• Sponsorship: Medtronic (5 RCTs)[1, 7-9, 12, 25]; Norton Healthcare grant (1 RCT)[26]. • Additional details: Table 15 and surrounding text.			
	High	***Conclusions*** The following outcomes were similar in both treatment groups: length of hospital stay (5 RCTs), ODI outcomes (3 – 6 RCTs), leg pain (6 RCTs), SF-36 scores (6 RCTs), and work status (4 RCTs).	+	+	+
	Low	The following outcomes were similar in both treatment groups: patient satisfaction (2 RCTs), neurological status (1 RCT), and overall success (1 RCT).	+	-	-
	Moderate	The following outcomes were reported as either similar or improved in patients treated with rhBMP-2 compared with ICBG: • mean operative time (similar in 3 RCTs; statistically improved in 3 RCTs), • perioperative blood loss (similar in 4 RCTs; statistically improved in 2 RCTs), • fusion (similar in 3 RCTs; statistically improved in 3 RCTs), • back pain (similar in 5 RCTs; clinically improved in 1 RCT)	+	+	-
Effectiveness		***Study characteristics*** • Evidence base: 8 cohort studies[13, 15, 27-32], (including 2 prospective cohort studies[30, 32], 1 prospective case control study[31], 3 retrospective cohort studies[13, 15, 29], and 2 retrospective cohort studies with historical controls[27, 28]): LoE II (1 study)[32]; LoE III (7 studies)[13, 15, 27-31]. Study size ranged from 36-126 patients. • Interventions: Primary or revision single- or multi-level anterior (two studies), posterior (two studies), transforaminal (one study), or posterolateral (three studies) lumbar fusion with rhBMP-2 or iliac crest bone autograft (ICBG), allograft chips, or local or rib autograft. Due to heterogeneity in control treatment, patient diagnosis, and surgical procedures (i.e., approach, use of local autograft or ICBG or bone graft extenders, use of cage versus allograft dowel versus no device, single- versus multilevel design, primary versus revision surgery), we did not pool outcomes data. Doses of rhBMP-2 varied and ranged from 3-36 mg per patient (when reported).			

Key Question 2: Efficacy and effectiveness					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
		• Population: Six of the cohort studies included patients with DDD, radiculitis; some of these studies also include those with up to grade 1 or 2 spondylolisthesis, scoliosis, instability, nonunion, or adjacent segment degeneration. One study treated patients with symptomatic pseudarthrosis following previous PLIF for DDD; one study evaluated patients with scoliosis with degeneration distal to a prior long idiopathic scoliosis fusion site. • Length of follow-up: Mean of 9-39 months. • Sponsorship: Medtronic (1 study)[32], Medtronic and Norton Healthcare grants (1 study)[28], no funding (2 studies)[15, 30], no direct funding but benefits may have been received (1 study)[27], or funding not reported (3 studies)[13, 29, 31]. • Additional details: Table 19 and surrounding text. ***Conclusions***			
	Low	The following outcomes were similar in both treatment groups: fusion (similar in 7 studies, improved in 1 study) and pain (5 studies)	-	+	+
	Insufficient	The following outcomes were similar in both treatment groups: operative time (1 study), ODI scores (2 studies), function (2 studies), patient satisfaction (2 studies), overall patient-reported clinical outcome (1 study), medication use (1 study), and mental health/self image (1 study).	-	-	-
	Insufficient	The following outcomes were statistically improved in patients treated with rhBMP-2 compared with control: perioperative blood loss (1 study)	-	-	-
Off-label use of rhBMP-7 in the lumbar spine					
Efficacy		***Study characteristics*** • Evidence base: 1 RCT[5]: LoE IIb. There were 33 patients enrolled in the study. • Interventions: Primary one- or two- level ACDF with InFUSE (n = 18) or ICBG (n = 15). RhBMP-2 was used at a dose of 0.6-1.2 mg per patient. • Population: Patients with degenerative cervical disease with radiculopathy and/or myelopathy. • Length of follow-up: 24 months. • Sponsorship: No direct funding but benefits may have been received.			

Key Question 2: Efficacy and effectiveness					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
		• Additional details: Table 30 and surrounding text.			
	High	The following outcomes were similar in both treatment groups: length of hospital stay (3 RCTs), fusion (5 RCTs), and ODI outcomes (3 RCTs)	+	+	+
	Low	The following outcomes were similar in both treatment groups: back pain (1 RCT), leg pain (1 RCT), SF-36 physical component subscale scores (1 RCT), neurologic success (1 RCT), or overall success (1 RCT).	+	-	-
	Low	The following outcome was reported as either similar or improved in patients treated with rhBMP-7 compared with ICBG or local autograft: • operative time (similar in 2 RCTs, statistically improved in 1 RCT) • perioperative blood loss (similar in 1 RCT, statistically improved in 1 RCT)	+	-	-
Effectiveness		No studies were identified that evaluated the effectiveness of off-label use of rhBMP-7 in the lumbar spine.			
Off-label use of rhBMP-2 in the cervical spine					
Efficacy		***Study characteristics*** • Evidence base: 1 RCT[5]: LoE IIb. There were 33 patients enrolled in the study. • Interventions: Primary one- or two- level ACDF with InFUSE (n = 18) or ICBG (n = 15). RhBMP-2 was used at a dose of 0.6-1.2 mg per patient. • Population: Patients with degenerative cervical disease with radiculopathy and/or myelopathy. • Length of follow-up: 24 months. • Sponsorship: No direct funding but benefits may have been received. • Additional details: Table 30 and surrounding text.			
	Low	***Conclusions*** The following outcomes were similar in both treatment groups: operative time, perioperative blood loss, length of hospital stay, fusion, neck pain, SF-36 scores, patient satisfaction, and neurological success. The following outcomes were statistically improved in patients treated with rhBMP-2	+	-	-

Key Question 2: Efficacy and effectiveness					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
		compared with ICBG: NDI and arm pain scores.			
Effectiveness		***Effectiveness*** • Evidence base: 5 cohort studies[11, 33-36] (including 1 prospective cohort study[11], 3 retrospective cohort studies[33-35], and 1 retrospective case-control database study[36]): all LoE III. Study size ranged from 58-775 patients. • Interventions: Primary or revision single- or multi-level anterior (two studies), posterior (two studies) cervical fusion with rhBMP-2 or iliac crest bone autograft (ICBG) (two studies), allograft and demineralized bone matrix (one study), a combination of autograft and/or allograft materials (one study). One study did not report surgical approach or the details of the control treatment (referred to as "non-BMP"). BMP was used a dose that ranged from 0.9 to 12 mg per patient (when reported). Due to heterogeneity in control treatments and surgical procedures (i.e., approach, use of local autograft or ICBG or allograft, single- versus multilevel design, primary versus revision surgery), we were not able to pool outcomes data. • Population: Two of the cohort studies included patients with DDD; another included patients with DDD, herniated nucleus pulposus, or stenosis. A fourth study treated patients for stenosis, spondylosis, or nonunion from a previous fusion. The fifth study did not report patient diagnoses. • Length of follow-up: 1 – 36 months. • Sponsorship: : No funding received (1 study)[11]; funding received but source not stated (1 study)[36]; no direct funding but benefits may have been received (2 studies)[33, 35], and funding not reported (1 study)[34]. • Additional details: Table 35 and surrounding text.			
	Low	***Conclusions*** The following outcome was similar in both treatment groups: perioperative blood loss (3 studies).	-	+	+

Key Question 2: Efficacy and effectiveness					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
	Insufficient	The following outcomes were similar in both treatment groups: operative time (2 studies), ODI outcomes (2 studies), and arm pain (2 studies).	-	-	-
	Insufficient	The following outcome was reported as <u>either similar or improved</u> in patients treated with rhBMP-2 compared with control: • fusion (similar in 1 cohort study, statistically improved in 1 cohort study)	-	-	-
	Insufficient	The following outcomes were reported as <u>either similar or worse</u> in patients treated with rhBMP-2 compared with control: • length of hospital stay (4 studies reported similar outcomes while 1 large study reported longer hospital stays in the rhBMP-2 group compared with the control group) • neck pain (2 studies reported similar outcomes while 1 study reported more rhBMP-2 patients with persistent neck pain at final follow-up). One study was funded but the source was not stated; authors from two studies may have received financial or other benefits related to the study; one study was not funded; funding for the remaining study was not reported.	-	+	-
Off-label use of rhBMP-7 in the cervical spine					
<u>Efficacy</u>		No studies were identified that evaluated the efficacy of off-label use of rhBMP-7 in the cervical spine.			
<u>Effectiveness</u>		No studies were identified that evaluated the efficacy of off-label use of rhBMP-7 in the cervical spine.			

DBM: demineralized bone matrix; SoE: Strength of Evidence

Table 70. Summary of evidence for Key Question 3

Key Question 3: Safety					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
Bone overgrowth		**On-label:** no on-label comparative studies reporting on this outcome.			
	Low (downgraded from moderate due to variability in estimates)	**Off-label**: inconsistent results reported • no cases of bone overgrowth in either group (2 RCTS and 2 cohort studies) • incidence of 75% in the rhBMP group vs. 13% in the control (1 RCT), and 21% vs. 8% of spinal levels (1 cohort study)	+	+	-
Osteoclast activity (resorption, osteolysis, graft migration/ loosening/ subsidence)	**Low**	**On-label:** similar risks in both groups, 1.3% vs. 0.0% (FDA pilot and pivotal RCTs for InFUSE)	+	-	-
	Moderate (downgraded from high due to effect size in cohort study)	**Off-label:** similar or possible higher risk for rhBMP • similar with risks ≤ 6% in each group (3 RCTs) • higher risk in the rhBMP-2 group, 62% of spinal levels vs. 10% in the controls (1 cohort study)	+	+	+
Local wound complications, superficial	**Insufficient**	**On-label:** insufficient risk estimates (1 very small pilot study, N = 14)	-	-	-
	Moderate	**Off-label:** similar risks (<10%) in both groups (2 RCTs, 5 cohorts)	+	-	+
Local wound complications, superficial or deep (unspecified)	**Low**	**On-label:** similar risks in both groups, 12.2% vs. 11.5% (FDA pilot and pivotal RCTs for InFUSE)	+	-	-
	Moderate (downgraded from high due to variability in estimates)	**Off-label**: similar but variable risks in both groups, 0–20% (4 RCTs, 2 cohort studies)	+	+	+
Local wound complications, deep; surgery for deep wound complications		**On-label:** no on-label comparative studies reporting on this outcome.			
	Low	**Off-label:** similar risks in both groups, ≤10% in each group (1 RCT, 4 cohort studies)	-	+	+
Dysphagia/ neck swelling	**Insufficient**	**On-label:** higher risk of "respiratory" complications in rhBMP-2 patients, 8.6% vs. 1.7% (FDA summary on InFUSE)	-	-	-
	Moderate (downgraded from high due to the small sample size of 2 RCTs)	**Off-label (lumbar):** similar risks of "respiratory" complications in both groups, <7% (3 RCTs, 2 cohorts)	+	+	+
Dysphagia/ neck swelling (cont.)	**Moderate** (upgraded from low due to effect size)	**Off-label (cervical)**: higher risks in the rhBMP groups • 35% vs. 9% (pooled) (4 cohort studies) • ~2 fold increase (2 large database studies)	-	+	+
Retrograde ejaculation	**Low** (upgraded from insufficient due to effect size)	**On-label:** higher risk in rhBMP-2 groups, 7.9% vs. 1.4% (FDA summary on InFUSE)	+	-	-

Key Question 3: Safety					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
	Low (upgraded from insufficient due to high risk differences)	**Off-label**: higher risk in rhBMP-2 groups, 7.2% vs. 0.6% (1 cohort study)	-	-	-
Ileus/bowel obstruction	**Insufficient**	**On-label:** insufficient risk estimates (1 very small pilot study, N = 14)	-	-	-
	Insufficient	**Off-label:** insufficient risk estimates (1 retrospective cohort study)	-	-	-
Urinary retention	**Insufficient**	**On-label:** insufficient risk estimates (1 very small pilot study, N = 14) **Off-label:** No comparative studies reported on this outcome.	-	-	-
Radiculitis (adverse event)	**Insufficient**	**On-label**: similar risks in both groups, 23% vs. 22% (FDA SSED for InFUSE)	-	-	-
	Low	**Off-label**: similar or lower risks in the rhBMP groups • Risks similar for rhBMP-2 compared with controls, 0-2% (1 RCT, 1 cohort study) • Risk lower for rhBMP-7 compared with controls, 6% vs. 13% (1 RCT)	+	-	-
Dural injury or CSF leak	**Insufficient**	**On-label**: similar low risks in both groups, 0% vs. 0.7% (FDA SSED for InFUSE)	-	-	-
	High	**Off-label**: similar but variable risks in both groups, 2.4–11% (3 RCTs, 7 cohort studies)	+	+	+
Neurological, unspecified/other (adverse event)	**Insufficient**	**On-label:** similar risks in both groups, 12.5% vs. 15.1% (FDA SSED for InFUSE)	-	-	-
	High	**Off-label:** similar but variable risks in both groups, 4.0–26.0% (4 RCTs, 3 cohort studies, FDA SSPB*)	+	+	+
Antibody responses to BMP	**Low**	**On-label:** similar low risks in both groups, 0.7% vs. 0.8% (1 RCT)	+	-	-
	High	**Off-label (rhBMP-2):** similar low risks in both groups, 0–0.7% (4 RCTs)	+	+	+
	Low	**Off-label (rhBMP-7):** higher risk in rhBMP group, 93.7% vs. 21% (no clinical sequelae) (1 RCT)	+	-	-
Antibody responses to collagen	**Low**	**On-label:** similar low risks in both groups, 0.7% vs. 0.8% (1 RCT)	+	-	-
	Low	**Off-label:** similar risks in both groups, 9% vs. 11% (2 RCTs)	+	-	-
Cancer	**Low**	**On-label:** similar low risks in both groups, 0.7% vs. 0.7% (1 pivotal RCT of the FDA SSED for	+	-	-

Key Question 3: Safety					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
		InFUSE)			
	Moderate (downgraded due to poor methodology for assessing cancer risks)	**Off-label:** higher cancer risks in the rhBMP-2 and rhBMP-7 groups at 1, 2, 4, and 5 years; 3.8–16.7% vs. 0.9–7.6% (3 RCTs including the pivotal RCT of the FDA SSED for InFUSE, 1 cohort study)	+	+	+
Cardio/vascular	**Low**	**On-label:** similar but variable risks in both groups, 4.2–10.1% vs. 2.2–12.2%, (1 RCT + largely overlapping FDA SSED)	+	-	-
	High	**Off-label:** similar but variable risks in both groups, 3.9–18.3% vs. 2.0–22.1%, (4 RCTs, 3 cohort studies to include the FDA SSPB* which may partially overlap)	+	+	+
Deep vein thrombosis	**Low**	**On-label:** similar low risks in both groups, 0% vs. 1.5% (1 RCT)	+	-	-
	Low	**Off-label:** similar but variable risks in both groups, 0–9% versus 1.9–12% (1 RCT, 2 cohort studies)	+	-	-
Death	**Low**	**On-label:** similar low risks in both groups at 24 months, 0% vs. 0.7% (1 RCT + largely overlapping FDA SSED)	+	-	-
	High	**Off-label, lumbar:** similar but variable risks in both groups at 24 to 36 months, 1.6–5.3% vs. 1.7–6.0% (4 RCTs, 2 cohort studies)	+	+	+
	Insufficient	**Off-label, cervical:** higher risk in rhBMP group up to 90 days post-operative, 4.2% vs. 1.7%, *P* = .047 (1 RCT); • causes of death were not reported, no demographic or surgical details provided – thus, significance of this result should be interpreted with caution given an absence of controlling for possible confounding between treatment groups	+	-	-
Revision	**Insufficient**	**On-label:** similar low risks in both groups at 24 months, 0.4% vs. 2.0% (1 integrated analysis)	-	-	-
	High	**Off-label:** similar or lower risks in the rhBMP groups over 17 to 48 months follow-up • Pooled risks similar in both groups, 6.0% vs. 6.2% (7 RCTs) • Lower risks in the rhBMP groups (0–3%) vs. controls (4–10%) (3 cohort studies) • Overall risks were slightly higher with rhBMP-7 use (9.5% vs. 11%) compared with rhBMP2 use (3.8% vs. 4.8%)	+	+	+
Hardware removal	**Low**	**On-label:** similar low risks in both groups at 24 months, 0–1.7% (1 RCT + 1 partially overlapping integrated analysis)	+	-	-

Key Question 3: Safety					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
	Moderate	**Off-label:** lower or similar risk in rhBMP groups over 3 to 28 months follow-up • Pooled risks lower in rhBMP-2 group (2.8%) vs. controls (7.2%) at 24 months (4 RCTs) • Risks identical between groups (8.0%) (2 cohort studies)	+	+	-
Supplemental fixation	**Low**	**On-label:** lower risks in rhBMP groups at 24 months, 5.2% vs. 10.8% (2 RCTs: 1 small pilot RCT and one pivotal RCT)	+	-	-
	Moderate	**Off-label (rhBMP-2, lumbar):** lower risk in rhBMP groups, 2.5–6.7% vs. 6.2–9.5% over 2 to 36 months follow-up (2 RCTs, 4 cohorts)	+	-	+
	Low (downgraded from moderate due to small sample sizes of the RCTs)	**Off-label (rhBMP-7, lumbar):** higher risk in rhBMP groups at 24 months, 10% vs. 0% (2 small RCTs)	+	-	+
	Insufficient	**Off-label (rhBMP-2, cervical):** lower risks in rhBMP groups at 30 months, 0% vs. 3.0% (1 large database study)	-	-	-
Reoperation	**Insufficient**	**On-label:** lower risks in the rhBMP-2 group at 24 months, 2.9% vs. 8.0% (1 integrated analysis)	-	-	-
	High	**Off-label:** similar but variable risks in both groups, 0–10% vs. 2.0–10% over 3 to 30 months follow-up (3 RCTs, 3 cohort studies)	+	+	+
Fusion at a different spinal level		**On-label:** no on-label studies reporting on this outcome.			
	Moderate (downgraded from high due to small sample sizes)	**Off-label:** similar but variable risks in both groups, 3.8–5.6% vs. 0–4.0% at 24 to 30 months follow-up (3 RCTs, 2 cohort studies)	+	+	+
Second surgeries (details not reported)		**On-label:** no on-label studies reporting on this outcome.			
	Insufficient	**Off-label:** similar but variable risks in both groups, 10.8–15.0% vs. 10.5–21.0% at 24 to 48 months follow-up (2 cohort studies)	-	-	+

*The FDA SSPB for OP-1 was to be considered of low quality because no information was presented for 278/326 patients included in the safety data tables (i.e., study data only presented for the pilot trial, which included 48 patients and presumably is part of the 326 patients evaluated for safety).

.

Table 71. Summary of evidence for Key Question 4

Key Question 4: Differential efficacy, effectiveness, and safety					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
• **Age** • **Sex** • **Smoking status** • **Number of levels treated** • **Complexity of fusions** • **Surgical approach** • **Previous surgeries**	**Insufficient**	We found no strong evidence of the differential effectiveness of spinal fusion using rhBMP-2 or rhBMP-7 versus spinal fusion using ICBG or alternative bone graft substitutes in any subpopulation. Although these eight studies examined outcomes in various subpopulations, none of these studies pre-specified the subgroup analyses, none of the studies performed a test of interaction as the method of subgroup analysis, and some of the studies were inadequately powered to detect differences in treatment effect. In general, fusion without rhBMP tended to have lower complication risks, while fusion with rhBMP tended to have better radiographic outcomes across most subpopulations examined, although in many cases the differences were small.	-	-	-
• **Baseline functional or pain status** • **Provider type** • **Payor/ beneficiary type**		No studies were identified that evaluated the differential effectiveness of spinal fusion with rhBMP-2 or rhBMP-7 based on baseline function or pain, provider type, or payor/ beneficiary type.			

Table 72. Summary of evidence for Key Question 5

Key Question 5: Cost effectiveness					
	SoE	**Conclusions/Comments**	**Quality**	**Quantity**	**Consistency**
Cost-effectivenesss	**Low**	<u>RhBMP-2 use in lumbar spine (on-label):</u> Conflicting evidence (2 studies): • One study concluded that when analyzed as part of the treatment cost, on-label use of rhBMP-2 results in lower costs per QALY only when it is assumed to be a part of the Medicare reimbursement and no cost differential is calculated. When the cost of BMP is added to the treatment (more common policy), BMP use associated with higher QALYs and higher cost, a common outcome for new technologies. • Another study found that rhBMP-2 use was unlikely to be cost-effective due to higher costs and similar outcomes compared with the control group.	+	-	-
	Low	<u>RhBMP-2 use in lumbar spine (off-label):</u> Off-label use of rhBMP-2 was more cost-effective than ICBG for posterolateral spine fusion in patients ≥ 60 years of age (1 study). No studies were identified that evaluated the cost-effectiveness of spinal fusion with rhBMP-7 in the lumbar or cervical spine or with rhBMP-2 in the cervical spine.	+	-	-

QALY: quality-adjusted life year

References

1. Burkus JK, Sandhu HS, Gornet MF. Influence of rhBMP-2 on the healing patterns associated with allograft interbody constructs in comparison with autograft. Spine 2006;31:775-81.
2. Hwang CJ, Vaccaro AR, Hong J, et al. Immunogenicity of osteogenic protein 1: results from a prospective, randomized, controlled, multicenter pivotal study of uninstrumented lumbar posterolateral fusion. J Neurosurg Spine 2010;13:484-93.
3. FDA (US. Food and Drug Administration). Summary of safety and probable benefit (SSPB): H020008: OP-1 Putty (Stryker Biotech; Hopkinton, MA); http://www.accessdata.fda.gov/cdrh_docs/pdf2/H020008b.pdf (accessed August, 2011). 2004.
4. Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best Pract Res Clin Rheumatol 2005;19:593-607.
5. Baskin DS, Ryan P, Sonntag V, Westmark R, Widmayer MA. A prospective, randomized, controlled cervical fusion study using recombinant human bone morphogenetic protein-2 with the CORNERSTONE-SR allograft ring and the ATLANTIS anterior cervical plate. Spine 2003;28:1219-24; discussion 25.
6. Burkus JK, Gornet MF, Dickman CA, Zdeblick TA. Anterior lumbar interbody fusion using rhBMP-2 with tapered interbody cages. J Spinal Disord Tech 2002;15:337-49.
7. Boden SD, Kang J, Sandhu H, Heller JG. Use of recombinant human bone morphogenetic protein-2 to achieve posterolateral lumbar spine fusion in humans: a prospective, randomized clinical pilot trial: 2002 Volvo Award in clinical studies. Spine 2002;27:2662-73.
8. Dimar JR, 2nd, Glassman SD, Burkus JK, Pryor PW, Hardacker JW, Carreon LY. Clinical and radiographic analysis of an optimized rhBMP-2 formulation as an autograft replacement in posterolateral lumbar spine arthrodesis. J Bone Joint Surg Am 2009;91:1377-86.
9. Haid RW, Jr., Branch CL, Jr., Alexander JT, Burkus JK. Posterior lumbar interbody fusion using recombinant human bone morphogenetic protein type 2 with cylindrical interbody cages. Spine J 2004;4:527-38; discussion 38-9.
10. Delawi D, Dhert WJ, Rillardon L, et al. A prospective, randomized, controlled, multicenter study of osteogenic protein-1 in instrumented posterolateral fusions: report on safety and feasibility. Spine (Phila Pa 1976) 2010;35:1185-91.
11. Buttermann GR. Prospective nonrandomized comparison of an allograft with bone morphogenic protein versus an iliac-crest autograft in anterior cervical discectomy and fusion. Spine J 2008;8:426-35.
12. Burkus JK, Sandhu HS, Gornet MF, Longley MC. Use of rhBMP-2 in combination with structural cortical allografts: clinical and radiographic outcomes in anterior lumbar spinal surgery. J Bone Joint Surg Am 2005;87:1205-12.
13. Mummaneni PV, Pan J, Haid RW, Rodts GE. Contribution of recombinant human bone morphogenetic protein-2 to the rapid creation of interbody fusion when used in transforaminal lumbar interbody fusion: a preliminary report. Invited submission from

the Joint Section Meeting on Disorders of the Spine and Peripheral Nerves, March 2004. J Neurosurg Spine 2004;1:19-23.
14. Rihn JA, Patel R, Makda J, et al. Complications associated with single-level transforaminal lumbar interbody fusion. Spine J 2009;9:623-9.
15. Taghavi CE, Lee KB, Keorochana G, Tzeng ST, Yoo JH, Wang JC. Bone morphogenetic protein-2 and bone marrow aspirate with allograft as alternatives to autograft in instrumented revision posterolateral lumbar spinal fusion: a minimum two-year follow-up study. Spine (Phila Pa 1976) 2010;35:1144-50.
16. Vaccaro AR, Anderson DG, Patel T, et al. Comparison of OP-1 Putty (rhBMP-7) to iliac crest autograft for posterolateral lumbar arthrodesis: a minimum 2-year follow-up pilot study. Spine (Phila Pa 1976) 2005;30:2709-16.
17. Vaccaro AR, Patel T, Fischgrund J, et al. A pilot study evaluating the safety and efficacy of OP-1 Putty (rhBMP-7) as a replacement for iliac crest autograft in posterolateral lumbar arthrodesis for degenerative spondylolisthesis. Spine (Phila Pa 1976) 2004;29:1885-92.
18. Vaccaro AR, Whang PG, Patel T, et al. The safety and efficacy of OP-1 (rhBMP-7) as a replacement for iliac crest autograft for posterolateral lumbar arthrodesis: minimum 4-year follow-up of a pilot study. Spine J 2008;8:457-65.
19. Vaccaro AR, Lawrence JP, Patel T, et al. The safety and efficacy of OP-1 (rhBMP-7) as a replacement for iliac crest autograft in posterolateral lumbar arthrodesis: a long-term (>4 years) pivotal study. Spine 2008;33:2850-62.
20. Johnsson R, Stromqvist B, Aspenberg P. Randomized radiostereometric study comparing osteogenic protein-1 (BMP-7) and autograft bone in human noninstrumented posterolateral lumbar fusion: 2002 Volvo Award in clinical studies. Spine 2002;27:2654-61.
21. Boden SD, Zdeblick TA, Sandhu HS, Heim SE. The use of rhBMP-2 in interbody fusion cages. Definitive evidence of osteoinduction in humans: a preliminary report. Spine 2000;25:376-81.
22. FDA (US. Food and Drug Administration). Summary of safety and effectiveness (SSED): P000058: InFUSE Bone Graft/LT-CAGE Lumbar Tapered Fusion Device (Medtronic Sofamor Danek; Memphis TN); www.accessdata.fda.gov/cdrh_docs/pdf/P000058b.pdf (accessed August, 2011). 2002.
23. Burkus JK, Heim SE, Gornet MF, Zdeblick TA. Is INFUSE bone graft superior to autograft bone? An integrated analysis of clinical trials using the LT-CAGE lumbar tapered fusion device. J Spinal Disord Tech 2003;16:113-22.
24. Kleeman TJ, Ahn UM, Talbot-Kleeman A. Laparoscopic anterior lumbar interbody fusion with rhBMP-2: a prospective study of clinical and radiographic outcomes. Spine (Phila Pa 1976) 2001;26:2751-6.
25. Dawson E, Bae HW, Burkus JK, Stambough JL, Glassman SD. Recombinant human bone morphogenetic protein-2 on an absorbable collagen sponge with an osteoconductive bulking agent in posterolateral arthrodesis with instrumentation. A prospective randomized trial. J Bone Joint Surg Am 2009;91:1604-13.
26. Glassman SD, Carreon LY, Djurasovic M, et al. RhBMP-2 versus iliac crest bone graft for lumbar spine fusion: a randomized, controlled trial in patients over sixty years of age. Spine 2008;33:2843-9.

27. Crawford CH, 3rd, Bridwell KH, Cho W, et al. Extension of prior idiopathic scoliosis fusions to the sacrum: a matched cohort analysis of sixty patients with minimum two-year follow-up. Spine (Phila Pa 1976) 2010;35:1843-8.
28. Glassman SD, Carreon L, Djurasovic M, et al. Posterolateral lumbar spine fusion with INFUSE bone graft. Spine J 2007;7:44-9.
29. Lee KB, Taghavi CE, Hsu MS, et al. The efficacy of rhBMP-2 versus autograft for posterolateral lumbar spine fusion in elderly patients. Eur Spine J 2010;19:924-30.
30. Pradhan BB, Bae HW, Dawson EG, Patel VV, Delamarter RB. Graft resorption with the use of bone morphogenetic protein: lessons from anterior lumbar interbody fusion using femoral ring allografts and recombinant human bone morphogenetic protein-2. Spine 2006;31:E277-84.
31. Singh K, Smucker JD, Gill S, Boden SD. Use of recombinant human bone morphogenetic protein-2 as an adjunct in posterolateral lumbar spine fusion: a prospective CT-scan analysis at one and two years. J Spinal Disord Tech 2006;19:416-23.
32. Slosar PJ, Josey R, Reynolds J. Accelerating lumbar fusions by combining rhBMP-2 with allograft bone: a prospective analysis of interbody fusion rates and clinical outcomes. Spine J 2007;7:301-7.
33. Crawford CH, 3rd, Carreon LY, McGinnis MD, Campbell MJ, Glassman SD. Perioperative complications of recombinant human bone morphogenetic protein-2 on an absorbable collagen sponge versus iliac crest bone graft for posterior cervical arthrodesis. Spine 2009;34:1390-4.
34. Vaidya R, Carp J, Sethi A, Bartol S, Craig J, Les CM. Complications of anterior cervical discectomy and fusion using recombinant human bone morphogenetic protein-2. Eur Spine J 2007;16:1257-65.
35. Xu R, Bydon M, Sciubba DM, et al. Safety and efficacy of rhBMP2 in posterior cervical spinal fusion for subaxial degenerative spine disease: Analysis of outcomes in 204 patients. Surg Neurol Int 2011;2:109.
36. Yaremchuk KL, Toma MS, Somers ML, Peterson E. Acute airway obstruction in cervical spinal procedures with bone morphogenetic proteins. Laryngoscope 2010;120:1954-7.
37. Hadjidakis DJ, Androulakis, II. Bone remodeling. Annals of the New York Academy of Sciences 2006;1092:385-96.
38. Marsh DR, Li G. The biology of fracture healing: optimising outcome. British medical bulletin 1999;55:856-69.
39. AHRQ (Agency for Healthcare Research and Quality). Technology Assessment Report: Bone morphogenetic protein: the state of the evidence of on-label and off-label use; 2010.
40. Tosounidis T, Kontakis G, Nikolaou V, Papathanassopoulos A, Giannoudis PV. Fracture healing and bone repair: An update. Trauma (GBR) Trauma 2009;11:145-56.
41. Tehranzadeh J, Ton JD, Rosen CD. Advances in spinal fusion. Seminars in ultrasound, CT, and MR 2005;26:103-13.
42. Garrison KR, Donell S, Ryder J, et al. Clinical effectiveness and cost-effectiveness of bone morphogenetic proteins in the non-healing of fractures and spinal fusion: a systematic review. Health Technol Assess 2007;11:1-150, iii-iv.
43. Hu SS, Tribus CB, Diab M, Ghanayem AJ. Spondylolisthesis and spondylolysis. J Bone Joint Surg Am 2008;90:656-71.
44. Howard JM, Glassman SD, Carreon LY. Posterior iliac crest pain after posterolateral fusion with or without iliac crest graft harvest. Spine J 2011;11:534-7.

45. Bauer TW. An overview of the histology of skeletal substitute materials. Archives of pathology & laboratory medicine 2007;131:217-24.
46. FDA (US. Food and Drug Administration). Executive Summary for P050036: Medtronic's AMPLIFY rhBMP-2 Matrix. Orthopaedic and Rehabilitation Devices Advisory Pannel; http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/medicaldevices/medicaldevicesadvisorycommittee/orthopaedicandrehabilitationdevicespanel/ucm220079.pdf (accessed February, 2012).
. 2010.
47. Carragee EJ, Hurwitz EL, Weiner BK. A critical review of recombinant human bone morphogenetic protein-2 trials in spinal surgery: emerging safety concerns and lessons learned. Spine J 2011;11:471-91.
48. Vaidya R. Transforaminal interbody fusion and the "off label" use of recombinant human bone morphogenetic protein-2. Spine J 2009;9:667-9.
49. Resnick DK. Spinal Fusion for Discogenic Back Pain: Patient Selection, Operative Techniques, and Outcomes. Techniques in Neurosurgery Techniques in Neurosurgery 2003;8:176-90.
50. Nandi SK, Roy S, Mukherjee P, Kundu B, De DK, Basu D. Orthopaedic applications of bone graft & graft substitutes: a review. The Indian journal of medical research 2010;132:15-30.
51. Bessa PC, Casal M, Reis RL. Bone morphogenetic proteins in tissue engineering: the road from the laboratory to the clinic, part I (basic concepts). Journal of tissue engineering and regenerative medicine 2008;2:1-13.
52. Wozney JM. Overview of bone morphogenetic proteins. Spine (Phila Pa 1976) 2002;27:S2-8.
53. Rihn JA, Gates C, Glassman SD, Phillips FM, Schwender JD, Albert TJ. The use of bone morphogenetic protein in lumbar spine surgery. J Bone Joint Surg Am 2008;90:2014-25.
54. Mroz TE, Wang JC, Hashimoto R, Norvell DC. Complications related to osteobiologics use in spine surgery: a systematic review. Spine (Phila Pa 1976) 2010;35:S86-104.
55. Joseph V, Rampersaud YR. Heterotopic bone formation with the use of rhBMP2 in posterior minimal access interbody fusion: a CT analysis. Spine (Phila Pa 1976) 2007;32:2885-90.
56. Anderson DG, Sayadipour A, Shelby K, Albert TJ, Vaccaro AR, Weinstein MS. Anterior interbody arthrodesis with percutaneous posterior pedicle fixation for degenerative conditions of the lumbar spine. Eur Spine J 2011;20:1323-30.
57. Sethi A, Craig J, Bartol S, et al. Radiographic and CT evaluation of recombinant human bone morphogenetic protein-2-assisted spinal interbody fusion. AJR Am J Roentgenol 2011;197:W128-33.
58. Vaidya R, Sethi A, Bartol S, Jacobson M, Coe C, Craig JG. Complications in the use of rhBMP-2 in PEEK cages for interbody spinal fusions. J Spinal Disord Tech 2008;21:557-62.
59. McClellan JW, Mulconrey DS, Forbes RJ, Fullmer N. Vertebral bone resorption after transforaminal lumbar interbody fusion with bone morphogenetic protein (rhBMP-2). J Spinal Disord Tech 2006;19:483-6.

60. Carragee EJ, Mitsunaga KA, Hurwitz EL, Scuderi GJ. Retrograde ejaculation after anterior lumbar interbody fusion using rhBMP-2: a cohort controlled study. Spine J 2011;11:511-6.
61. Kendirci M, Hellstrom W. Retrograde ejaculation: Etiology, diagnosis, and management. Current Sexual Health Reports 2006;3:133-8.
62. Wong DA, Kumar A, Jatana S, Ghiselli G, Wong K. Neurologic impairment from ectopic bone in the lumbar canal: a potential complication of off-label PLIF/TLIF use of bone morphogenetic protein-2 (BMP-2). Spine J 2008;8:1011-8.
63. FDA Public Health Notification: Life-threatening complications associated with recombinant human bone morphogenetic protein in cervical spine fusion. 2008. (Accessed 2012, at http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/PublicHealthNotifications/ucm062000.htm.)
64. Smucker JD, Rhee JM, Singh K, Yoon ST, Heller JG. Increased swelling complications associated with off-label usage of rhBMP-2 in the anterior cervical spine. Spine 2006;31:2813-9.
65. Riley LH, 3rd, Vaccaro AR, Dettori JR, Hashimoto R. Postoperative dysphagia in anterior cervical spine surgery. Spine (Phila Pa 1976) 2010;35:S76-85.
66. Robin BN, Chaput CD, Zeitouni S, Rahm MD, Zerris VA, Sampson HW. Cytokine-mediated inflammatory reaction following posterior cervical decompression and fusion associated with recombinant human bone morphogenetic protein-2: a case study. Spine (Phila Pa 1976) 2010;35:E1350-4.
67. Garrett MP, Kakarla UK, Porter RW, Sonntag VK. Formation of painful seroma and edema after the use of recombinant human bone morphogenetic protein-2 in posterolateral lumbar spine fusions. Neurosurgery 2010;66:1044-9; discussion 9.
68. Agarwal R, Williams K, Umscheid CA, Welch WC. Osteoinductive bone graft substitutes for lumbar fusion: a systematic review. J Neurosurg Spine 2009;11:729-40.
69. Australian Safety and Efficacy Register of New Interventional Procedures - Surgical (ASERNIP-S). Horizon Scanning Report OP-1 Putty® for posterolateral lumbar fusion. In: Commonwealth of Australia, ed.; 2006.
70. Rihn JA, Kirkpatrick K, Albert TJ. Graft options in posterolateral and posterior interbody lumbar fusion. Spine (Phila Pa 1976) 2010;35:1629-39.
71. Low back - lumbar & thoracic (acute & chronic): NGC-8517. Work Loss Data Institute, 2011. (Accessed 10/27/2011, at http://www.guideline.gov/content.aspx?id=33184&search=bmp.)
72. Neck and upper back (acute & chronic): NGC-8518. Work Loss Data Institute, 2011. (Accessed 10/27/2011, at http://www.guideline.gov/content.aspx?id=33185&search=bmp.)
73. Resnick DK, Choudhri TF, Dailey AT, et al. Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 16: bone graft extenders and substitutes. J Neurosurg Spine 2005;2:733-6.
74. Recombinant Human Bone Morphogenetic Protein-2 For Spinal Surgery and Treatment of Open Tibial Fractures. 2005. (Accessed 10/28/2011, at http://www.ctaf.org/content/assessment/detail/568.)
75. Washington State Department of Labor & Industries. Bone Morphogenic Protein for the Treatment of Long Bone Fractures and for Use In Spinal Fusion Procedures; 2003.

76. Evidence based review: Bone Morphogenetic Proteins- 2 and 7. 2006. (Accessed 10/31/2011, at http://www.acc.co.nz.)
77. InFUSE Bone Grafts for Spinal Fusion Surgery: Clinical Effectiveness and Indications for Use in Canada. 2009. (Accessed at http://www.cadth.ca/media/pdf/L0107_Morphogenetic_Bone_final.pdf.)
78. Medical Advisory Secretariat. Bone morphogenetic proteins & spinal surgery for degenerative disc disease: an evidence-based analysis. Ontario Health Technology Assessment Series 2004;4.
79. Ryken TC, Heary RF, Matz PG, et al. Techniques for cervical interbody grafting. J Neurosurg Spine 2009;11:203-20.
80. Matz PG, Anderson PA, Kaiser MG, et al. Introduction and methodology: guidelines for the surgical management of cervical degenerative disease. J Neurosurg Spine 2009;11:101-3.
81. Annex B: Key to evidence statements and grades of recommendations 2011. (Accessed 12/13/2011, at http://www.sign.ac.uk/guidelines/fulltext/50/annexb.html.)
82. Clinical policy bulletin: bone and tendon graft substitutes and adjuncts, number 0411. 2011. (Accessed 10/14/2011, at http://www.aetna.com/cpb/medical/data/400_499/0411.html.)
83. Bone graft substitutes for use in bone repair, number 0118. 2011. (Accessed 10/14/2011, at http://www.cigna.com/assets/docs/health-care-professionals/coverage_positions/mm_0118_coveragepositioncriteria_recombinant_human_bone_morphogenetic_protein.pdf.)
84. National medical policy: bone morphogenetic protein, number NMP243. 2011. (Accessed 10/14/2011, at https://www.healthnet.com/static/general/unprotected/pdfs/national/policies/Bone_Morphogenetic_Protein_Feb_11.pdf.)
85. Bone morphogenetic protein, number 7.01.100. 2010. (Accessed 10/14/2011, at https://www.premera.com/stellent/groups/public/documents/medicalpolicy/cmi_022041.htm.)
86. InFUSE bone graft (recombinant human bone morphogenetic protein - 2), MPM 9.4. 2010. (Accessed 10/14/2011, at http://www.phs.org/idc/groups/public/@phs/@php/documents/phscontent/wcmprod1030196.pdf.)
87. Phillips B, Ball C, Sakett D, et al. Levels of evidence and grades of recommendation. 2001. (Accessed December, 2006, at http://www.cebm.net/levels_of_evidence.asp.)
88. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490.
89. West S, King V, Carey TS, et a. Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center, Contract No. 290-97-0011): Agency for Healthcare Research and Quality. Rockville, MD; 2002.
90. Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference. Low back pain: outcome measures. J Rheumatol 2001;28:431-8.
91. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980;66:271-3.

92. Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine (Phila Pa 1976) 2000;25:3115-24.
93. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther 1991;14:409-15.
94. Chapman JR, Hanson BP, Dettori JR, Norvell DC. Spine Outcomes Measures and Instruments. New York: Thieme; 2007.
95. Longo UG, Loppini M, Denaro L, Maffulli N, Denaro V. Rating scales for low back pain. British medical bulletin 2010;94:81-144.
96. Million R, Hall W, Nilsen KH, Baker RD, Jayson MI. Assessment of the progress of the back-pain patient 1981 Volvo Award in Clinical Science. Spine (Phila Pa 1976) 1982;7:204-12.
97. King JT, Jr., Roberts MS. Validity and reliability of the Short Form-36 in cervical spondylotic myelopathy. Journal of neurosurgery 2002;97:180-5.
98. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
99. Christensen FB, Laursen M, Gelineck J, Eiskjaer SP, Thomsen K, Bunger CE. Interobserver and intraobserver agreement of radiograph interpretation with and without pedicle screw implants: the need for a detailed classification system in posterolateral spinal fusion. Spine (Phila Pa 1976) 2001;26:538-43; discussion 43-4.
100. Kaiser MG, Mummaneni PV, Matz PG, et al. Radiographic assessment of cervical subaxial fusion. J Neurosurg Spine 2009;11:221-7.
101. Resnick DK, Choudhri TF, Dailey AT, et al. Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 4: radiographic assessment of fusion. J Neurosurg Spine 2005;2:653-7.
102. Park Y, Ha JW, Lee YT, Sung NY. The effect of a radiographic solid fusion on clinical outcomes after minimally invasive transforaminal lumbar interbody fusion. Spine J 2011;11:205-12.
103. Tsai TH, Huang TY, Lieu AS, et al. Functional outcome analysis: instrumented posterior lumbar interbody fusion for degenerative lumbar scoliosis. Acta neurochirurgica 2011;153:547-55.
104. Kanayama M, Hashimoto T, Shigenobu K, Yamane S, Bauer TW, Togawa D. A prospective randomized study of posterolateral lumbar fusion using osteogenic protein-1 (OP-1) versus local autograft with ceramic bone substitute: emphasis of surgical exploration and histologic assessment. Spine 2006;31:1067-74.
105. International Standards for Neurological Classification of Spinal Cord Injury. American Spinal Injury Association,, 2011. (Accessed 11/01/2011, at http://www.asia-spinalinjury.org/publications/59544_sc_Exam_Sheet_r4.pdf.)
106. Kirkaldy-Willis WH, Paine KW, Cauchoix J, McIvor G. Lumbar spinal stenosis. Clin Orthop Relat Res 1974:30-50.
107. 0 – 10 Numeric Rating Scale. 2003. (Accessed at http://painconsortium.nih.gov/pain_scales/index.html.)
108. Chapman JR, Dettori JR, Norvell DC, eds. Spine Classifications and Severity Measures. New York: Thieme Verlag; 2009.
109. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976) 2000;25:2940-52; discussion 52.

110. Salehi SA, Tawk R, Ganju A, LaMarca F, Liu JC, Ondra SL. Transforaminal lumbar interbody fusion: surgical technique and results in 24 patients. Neurosurgery 2004;54:368-74; discussion 74.
111. SRS Outcomes. 2011. (Accessed 11/1/2011, at http://www.srs.org/professionals/SRS_outcomes/index.htm.)
112. Beurskens AJ, de Vet HC, Koke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. Pain 1996;65:71-6.
113. Fisher K, Johnson M. Validation of the Oswestry low back pain disability questionnaire, its sensitivity as a measure of change following treatment and its relationship with other aspects of the chronic pain experience. Physiother Theory Pract 1997;13:67-80.
114. Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Quebec Back Pain Disability Scale: conceptualization and development. Journal of clinical epidemiology 1996;49:151-61.
115. Grevitt M, Khazim R, Webb J, Mulholland R, Shepperd J. The short form-36 health survey questionnaire in spine surgery. J Bone Joint Surg Br 1997;79:48-52.
116. Gronblad M, Hupli M, Wennerstrand P, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. The Clinical journal of pain 1993;9:189-95.
117. Haas M, Nyiendo J. Diagnostic utility of the McGill Pain Questionnaire and the Oswestry Disability Questionnaire for classification of low back pain syndromes. J Manipulative Physiol Ther 1992;15:90-8.
118. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain 1975;1:277-99.
119. Strong J, Ashton R, Large RG. Function and the patient with chronic low back pain. The Clinical journal of pain 1994;10:191-6.
120. Loisel P, Poitras S, Lemaire J, Durand P, Southiere A, Abenhaim L. Is work status of low back pain patients best described by an automated device or by a questionnaire? Spine (Phila Pa 1976) 1998;23:1588-94; discussion 95.
121. Ackelman BH, Lindgren U. Validity and reliability of a modified version of the neck disability index. Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine 2002;34:284-7.
122. Cleland JA, Fritz JM, Whitman JM, Palmer JA. The reliability and construct validity of the Neck Disability Index and patient specific functional scale in patients with cervical radiculopathy. Spine 2006;31:598-602. .
123. Anagnostis C, Gatchel RJ, Mayer TG. The pain disability questionnaire: a new psychometrically sound measure for chronic musculoskeletal disorders. Spine 2004;29:2290-302; discussion 303.
124. Hagg O, Fritzell P, Oden A, Nordwall A. Simplifying outcome measurement: evaluation of instruments for measuring outcome after fusion surgery for chronic low back pain. Spine 2002;27:1213-22.
125. Guilfoyle MR, Seeley H, Laing RJ. The Short Form 36 health survey in spine disease--validation against condition-specific measures. British journal of neurosurgery 2009;23:401-5.
126. Walsh TL, Hanscom B, Lurie JD, Weinstein JN. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the

Oswestry Disability Index, MODEMS, and the SF-36. Spine (Phila Pa 1976) 2003;28:607-15.
127. Parker SL, Adogwa O, Paul AR, et al. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. J Neurosurg Spine 2011;14:598-604.
128. Mirza SK, Deyo RA. Systematic review of randomized trials comparing lumbar fusion surgery to nonoperative care for treatment of chronic back pain. Spine (Phila Pa 1976) 2007;32:816-23.
129. Taylor SJ, Taylor AE, Foy MA, Fogg AJ. Responsiveness of common outcome measures for patients with low back pain. Spine (Phila Pa 1976) 1999;24:1805-12.
130. Ross R, Mirza AH, Norris HE, Khatri M. Survival and clinical outcome of SB Charite III disc replacement for back pain. J Bone Joint Surg Br 2007;89:785-9.
131. ECRI Institute. Spinal fusion and discography for chronic low back pain and uncomplicated lumbar degenerative disc disease Health Technology Assessment. Olympia, WA; 2007.
132. Hagg O, Fritzell P, Nordwall A, Swedish Lumbar Spine Study G. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J 2003;12:12-20.
133. Young BA, Walker MJ, Strunce JB, Boyles RE, Whitman JM, Childs JD. Responsiveness of the Neck Disability Index in patients with mechanical neck disorders. Spine J 2009;9:802-8.
134. Cleland JA, Childs JD, Whitman JM. Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. Archives of physical medicine and rehabilitation 2008;89:69-74.
135. Beurskens AJ, de Vet HC, Koke AJ, van der Heijden GJ, Knipschild PG. Measuring the functional status of patients with low back pain. Assessment of the quality of four disease-specific questionnaires. Spine (Phila Pa 1976) 1995;20:1017-28.
136. Grotle M, Brox JI, Vollestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. Spine (Phila Pa 1976) 2004;29:E492-501.
137. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008;33:90-4.
138. Cahill KS, Chi JH, Groff MW, McGuire K, Afendulis CC, Claus EB. Outcomes for Single-Level Lumbar Fusion: The Role of Bone Morphogenetic Protein. Spine (Phila Pa 1976) 2011.
139. Geibel PT, Boyd DL, Slabisak V. The use of recombinant human bone morphogenic protein in posterior interbody fusions of the lumbar spine: a clinical series. J Spinal Disord Tech 2009;22:315-20.
140. Mannion RJ, Nowitzke AM, Wood MJ. Promoting fusion in minimally invasive lumbar interbody stabilization with low-dose bone morphogenic protein-2--but what is the cost? Spine J 2011;11:527-33.
141. Villavicencio AT, Burneikiene S, Nelson EL, Bulsara KR, Favors M, Thramann J. Safety of transforaminal lumbar interbody fusion and intervertebral recombinant human bone morphogenetic protein-2. J Neurosurg Spine 2005;3:436-43.

142. Meisel HJ, Schnoring M, Hohaus C, et al. Posterior lumbar interbody fusion using rhBMP-2. Eur Spine J 2008;17:1735-44.
143. Furlan JC, Perrin RG, Govender PV, et al. Use of osteogenic protein-1 in patients at high risk for spinal pseudarthrosis: a prospective cohort study assessing safety, health-related quality of life, and radiographic fusion. Invited submission from the Joint Section on Disorders of the Spine and Peripheral Nerves, March 2007. J Neurosurg Spine 2007;7:486-95.
144. Vaccaro AR, Patel T, Fischgrund J, et al. A pilot safety and efficacy study of OP-1 putty (rhBMP-7) as an adjunct to iliac crest autograft in posterolateral lumbar fusions. Eur Spine J 2003;12:495-500.
145. Brower RS, Vickroy NM. A case of psoas ossification from the use of BMP-2 for posterolateral fusion at L4-L5. Spine (Phila Pa 1976) 2008;33:E653-5.
146. Chen NF, Smith ZA, Stiner E, Armin S, Sheikh H, Khoo LT. Symptomatic ectopic bone formation after off-label use of recombinant human bone morphogenetic protein-2 in transforaminal lumbar interbody fusion. J Neurosurg Spine 2010;12:40-6.
147. Deutsch H. High-dose bone morphogenetic protein-induced ectopic abdomen bone growth. Spine J 2010;10:e1-4.
148. Muchow RD, Hsu WK, Anderson PA. Histopathologic inflammatory response induced by recombinant bone morphogenetic protein-2 causing radiculopathy after transforaminal lumbar interbody fusion. Spine J 2010;10:e1-6.
149. Kim PD, Ludwig S, Poelstra K, Duggan B, Scalea T, Gelb D. Ectopic bone formation in the pelvis after combined anterior and posterior fusion of the spine with osteogenic protein-1 use: a case report. J Spinal Disord Tech 2010;23:215-20.
150. Boakye M, Mummaneni PV, Garrett M, Rodts G, Haid R. Anterior cervical discectomy and fusion involving a polyetheretherketone spacer and bone morphogenetic protein. J Neurosurg Spine 2005;2:521-5.
151. Klimo P, Jr., Peelle MW. Use of polyetheretherketone spacer and recombinant human bone morphogenetic protein-2 in the cervical spine: a radiographic analysis. Spine J 2009;9:959-66.
152. Tumialan LM, Pan J, Rodts GE, Mummaneni PV. The safety and efficacy of anterior cervical discectomy and fusion with polyetheretherketone spacer and recombinant human bone morphogenetic protein-2: a review of 200 patients. J Neurosurg Spine 2008;8:529-35.
153. Burkus JK, Gornet MF, Schuler TC, Kleeman TJ, Zdeblick TA. Six-year outcomes of anterior lumbar interbody arthrodesis with use of interbody fusion cages and recombinant human bone morphogenetic protein-2. J Bone Joint Surg Am 2009;91:1181-9.
154. Vaidya R, Weir R, Sethi A, Meisterling S, Hakeos W, Wybo CD. Interbody fusion with allograft and rhBMP-2 leads to consistent fusion but early subsidence. J Bone Joint Surg Br 2007;89:342-5.
155. Helgeson MD, Lehman RA, Jr., Patzkowski JC, Dmitriev AE, Rosner MK, Mack AW. Adjacent vertebral body osteolysis with bone morphogenetic protein use in transforaminal lumbar interbody fusion. Spine J 2011;11:507-10.
156. Knox JB, Dai JM, 3rd, Orchowski J. Osteolysis in transforaminal lumbar interbody fusion with bone morphogenetic protein-2. Spine (Phila Pa 1976) 2011;36:672-6.

157. Balseiro S, Nottmeier EW. Vertebral osteolysis originating from subchondral cyst end plate defects in transforaminal lumbar interbody fusion using rhBMP-2. Report of two cases. Spine J 2010;10:e6-e10.
158. Lewandrowski KU, Nanson C, Calderon R. Vertebral osteolysis after posterior interbody lumbar fusion with recombinant human bone morphogenetic protein 2: a report of five cases. Spine J 2007;7:609-14.
159. Shields LB, Raque GH, Glassman SD, et al. Adverse effects associated with high-dose recombinant human bone morphogenetic protein-2 use in anterior cervical spine fusion. Spine (Phila Pa 1976) 2006;31:542-7.
160. Williams BJ, Smith JS, Fu KM, et al. Does BMP increase the incidence of perioperative complications in spinal fusion? A comparison of 55,862 cases of spinal fusion with and without BMP. Spine (Phila Pa 1976) 2011.
161. Cahill KS, Chi JH, Day A, Claus EB. Prevalence, complications, and hospital charges associated with use of bone-morphogenetic proteins in spinal fusion procedures. JAMA 2009;302:58-66.
162. Deyo RA, Ching A, Matsen L, et al. Use of Bone Morphogenetic Proteins in Spinal Fusion Surgery for Older Adults with Lumbar Stenosis: Trends, Complications, Repeat Surgery, and Charges. Spine (Phila Pa 1976) 2011.
163. Carreon LY, Glassman SD, Brock DC, Dimar JR, Puno RM, Campbell MJ. Adverse events in patients re-exposed to bone morphogenetic protein for spine surgery. Spine (Phila Pa 1976) 2008;33:391-3.
164. Glassman SD, Gum JL, Crawford CH, 3rd, Shields CB, Carreon LY. Complications with recombinant human bone morphogenetic protein-2 in posterolateral spine fusion associated with a dural tear. Spine J 2011;11:522-6.
165. Glassman SD, Howard J, Dimar J, 2nd, Sweet A, Wilson G, Carreon L. Complications With rhBMP-2 in Posterolateral Spine Fusion: A Consecutive Series of One Thousand Thirty-Seven Cases. Spine (Phila Pa 1976) 2010.
166. Luhmann SJ, Bridwell KH, Cheng I, Imamura T, Lenke LG, Schootman M. Use of bone morphogenetic protein-2 for adult spinal deformity. Spine 2005;30:S110-7.
167. Owens K, Glassman SD, Howard JM, Djurasovic M, Witten JL, Carreon LY. Perioperative complications with rhBMP-2 in transforaminal lumbar interbody fusion. Eur Spine J 2011;20:612-7.
168. Hamilton DK, Smith JS, Reames DL, Williams BJ, Chernavvsky DR, Shaffrey CI. Safety, Efficacy, and Dosing of Recombinant Human Bone Morphogenetic Protein-2 for Posterior Cervical and Cervicothoracic Instrumented Fusion With a Minimum 2-Year Follow-up. Neurosurgery 2011;69:103-11.
169. Hiremath GK, Steinmetz MP, Krishnaney AA. Is it safe to use recombinant human bone morphogenetic protein in posterior cervical fusion? Spine (Phila Pa 1976) 2009;34:885-9.
170. Stambough JL, Clouse EK, Stambough JB. Instrumented one and two level posterolateral fusions with recombinant human bone morphogenetic protein-2 and allograft: a computed tomography study. Spine (Phila Pa 1976) 2010;35:124-9.
171. Shen HX, Buchowski JM, Yeom JS, Liu G, Lin N, Riew KD. Pseudarthrosis in multilevel anterior cervical fusion with rhBMP-2 and allograft: analysis of one hundred twenty-seven cases with minimum two-year follow-up. Spine (Phila Pa 1976) 2010;35:747-53.

172. Stachniak JB, Diebner JD, Brunk ES, Speed SM. Analysis of prevertebral soft-tissue swelling and dysphagia in multilevel anterior cervical discectomy and fusion with recombinant human bone morphogenetic protein-2 in patients at risk for pseudarthrosis. J Neurosurg Spine 2011;14:244-9.
173. Leach J, Bittar RG. BMP-7 (OP-1) safety in anterior cervical fusion surgery. J Clin Neurosci 2009;16:1417-20.
174. Perri B, Cooper M, Lauryssen C, Anand N. Adverse swelling associated with use of rh-BMP-2 in anterior cervical discectomy and fusion: a case study. Spine J 2007;7:235-9.
175. Mindea SA, Shih P, Song JK. Recombinant human bone morphogenetic protein-2-induced radiculitis in elective minimally invasive transforaminal lumbar interbody fusions: a series review. Spine (Phila Pa 1976) 2009;34:1480-4; discussion 5.
176. Govender PV, Rampersaud YR, Rickards L, Fehlings MG. Use of osteogenic protein-1 in spinal fusion: literature review and preliminary results in a prospective series of high-risk cases. Neurosurg Focus 2002;13:e4.
177. Mines D, Gu Y, Kou TD, Cooper GS. Recombinant human bone morphogenetic protein-2 and pancreatic cancer: a retrospective cohort study. Pharmacoepidemiol Drug Saf 2011;20:111-8.
178. Steib JP, Bouchaib J, Walter A, Schuller S, Charles YP. Could an osteoinductor result in degeneration of a neurofibroma in NF1? Eur Spine J 2010;19 Suppl 2:S220-5.
179. Latzman JM, Kong L, Liu C, Samadani U. Administration of human recombinant bone morphogenetic protein-2 for spine fusion may be associated with transient postoperative renal insufficiency. Spine (Phila Pa 1976) 2010;35:E231-7.
180. Moshel YA, Hernandez EI, Kong L, Liu C, Samadani U. Acute renal insufficiency, supraventricular tachycardia, and confusion after recombinant human bone morphogenetic protein-2 implantation for lumbosacral spine fusion. J Neurosurg Spine 2008;8:589-93.
181. Vaccaro AR, Patel T, Fischgrund J, et al. A 2-year follow-up pilot study evaluating the safety and efficacy of op-1 putty (rhbmp-7) as an adjunct to iliac crest autograft in posterolateral lumbar fusions. Eur Spine J 2005;14:623-9.
182. Carreon LY, Glassman SD, Djurasovic M, et al. RhBMP-2 versus iliac crest bone graft for lumbar spine fusion in patients over 60 years of age: a cost-utility study. Spine (Phila Pa 1976) 2009;34:238-43.