ORTHOPAEDIC SURGERY

Recombinant human bone morphogenetic protein-2 versus autogenous iliac crest bone graft for lumbar fusion: a meta-analysis of ten randomized controlled trials

Zhiguang Chen · Gen Ba · Tao Shen · Qin Fu

Abstract

Background Recombinant human bone morphogenetic protein-2 (rhBMP-2) as a substitute for iliac crest bone graft (ICBG) has been increasingly widely used in lumbar fusion. It has been proven non-inferior in fusion success and clinical outcomes when compared with ICBG. However, increasingly, some potentially uncommon and serious complications associated with the use of rhBMP-2 have been of great concern to surgeons. The purpose of this study was to determine whether rhBMP-2 could be considered an effective and, more importantly, a relatively safe substitute for ICBG in lumbar fusion.

Methods Randomized controlled trials that compared rhBMP-2 with ICBG for lumbar fusion were identified by computer and manual searching. The risk of bias and clinical relevance of the included studies were assessed. Publication bias was explored using funnel plot and statistical tests (Egger's test and Begg's test). Meta-analyses were performed using the Cochrane systematic review methods.

Results Ten randomized controlled trials (1,342 patients) met the inclusion criteria. Compared with ICBG, the use of rhBMP-2 significantly decreased the risk of fusion failure at all time intervals (6 months: $p < 0.0001$, RR = 0.55, 95 % CI = 0.42–0.72; 12 months: $p = 0.0003$, RR = 0.53, 95 % CI = 0.37–0.75; 24 months: $p < 0.00001$, RR = 0.31, 95 % CI = 0.21–0.46) and the rate of reoperation ($p = 0.0001$, RR = 0.52, 95 % CI = 0.37–0.72). There was no statistical difference in clinical improvement on the Oswestry Disability Index, although a favorable trend in the rhBMP-2 group was found ($p = 0.12$, RR = 0.73, 95 % CI = 0.49–1.08). Subgroup analyses stratified by the type of surgical procedure yielded similar results. Owing to the different data formats, meta-analysis on adverse events was not performed.

Conclusion RhBMP-2 was superior to the ICBG for achieving fusion success and avoiding reoperation. However, evidence from the Food and Drug Administration document and subsequent independent studies has demonstrated that original, industry-sponsored trials underestimated rhBMP-2-related adverse events. There are still security risks in the use of rhBMP-2.

Keywords Meta-analysis · Recombinant human bone morphogenetic protein-2 · Autogenous iliac crest bone graft · Complication · Lumbar fusion · Randomized controlled trial

Introduction

Lumbar fusion with autogenous iliac crest bone graft (ICBG) has been developed over several decades and has become the gold standard surgical procedure in the treatment of patients with symptomatic degenerative disc disease (DDD), spondylolisthesis, and other painful discogenic syndromes. The object of fusion is the attainment of rigid union of the degenerative and unstable motion segments to reduce or eliminate pain and decrease disability [1–3]. As is well known, there are several serious shortcomings in performing lumbar arthrodesis with ICBG, including donor-site morbidity and relatively high frequency of nonunion [4–6].

To prevent the morbidity associated with iliac crest harvesting and nonunion, graft substitutes such as local

Z. Chen · G. Ba · T. Shen · Q. Fu (✉)
Department of Orthopedic Surgery, Shengjing Hospital,
China Medical University, No. 36, Sanhao Street,
Heping District, Shenyang 110004, People's Republic of China
e-mail: qinfucmu@yahoo.cn

bone, demineralized bone matrix (DBM), and platelet gel have been advocated. However, there is not yet a substitute that has been proven superior to iliac crest graft [7–10]. Urist et al. [11] described bone morphogenetic protein (BMP) in 1965. For over 45 years thereafter, clinical applications of BMP have been developed. Currently, the widest clinical application of BMP is recombinant human bone morphogenetic protein-2 (rhBMP-2), an osteoinductive bone growth factor that is a member of the transforming growth factor-β superfamily [12, 13]. In previous clinical studies, rhBMP-2 delivered in a variety of carriers has been shown to have superior fusion rates to ICBG in lumbar fusion in different surgical procedures [14–18]. However, FDA-approved clinical use of rhBMP-2 has been limited since 2002 to anterior lumbar interbody fusion with proprietary lordotic tapered cages (LT cages, Medtronic Sofamor Danek, Memphis, TN), which means that rhBMP-2 use for other lumbar fusion procedures remains off-label. In addition, some potential complications associated with the use of rhBMP-2 in spinal fusion have been reported and highlighted by independent studies and the FDA. Given some of the limitations in these original controlled studies, which include variations in study design, sample size, and methods of outcome measurement, it is still uncertain whether rhBMP-2 can be regarded as a consummate alternative to ICBG in clinical use.

The purpose of this study is to evaluate the effectiveness and, more importantly, safety of rhBMP-2 compared with ICBG in lumbar fusion. We sought to evaluate all randomized controlled trials (RCTs) on this topic to determine whether rhBMP-2 can ultimately replace ICBG.

Materials and methods

Literature search

Relevant randomized controlled trials (RCTs) were identified up to February 2012 from PubMed, EMBASE, and the Cochrane Central Register of Controlled Trials. The references of the retrieved articles were also hand-searched. Key words that have been used for searching are: lumbar fusion, recombinant human bone morphogenetic protein-2, rhBMP-2, bone substitute. Two reviewers independently checked all titles, abstracts, and the full text of potentially eligible articles to decide which trials fit the inclusion criteria. Disagreement was resolved by discussion; if necessary, a third independent investigator made the decision.

Selection of studies

RCTs published in English were regarded eligible according to the following criteria: (1) participants were adults and underwent lumbar fusion for degenerative disease; (2) the intervention was lumbar fusion using rhBMP-2 as a substitute to ICBG in the treatment group; (3) the study reported at least one desirable outcome. Studies were excluded if (1) participants had an acute spinal fracture, infection, tumor, osteoporosis, or rheumatoid arthritis; (2) only the abstract was available.

Data extraction

The data including the following categories were extracted independently by the two reviewers: participant characteristics; sample size of each intervention group; follow-up rate and time (month); characteristics and details of each intervention; industry-sponsored information and financial interest. The primary outcome was fusion failure assessed by radiographs and computed tomography (CT) scans; the secondary outcomes were reoperation and clinical improvement of the Oswestry Disability Index (ODI) [19] (defined as a decrease of at least 15 % over the preoperative score). The adverse events reported in the studies were also recorded.

Risk of bias assessment

Risk of bias was independently assessed by the two reviewers according to the 12 criteria and the instructions recommended by the Cochrane Back Review Group (CBRG) [20].To each criterion, from three options, yes, unclear, or no is scored, where yes indicates the criterion has been met. The studies were rated as having a "low risk of bias" when at least 6 of the 12 CBRG criteria were met and the study had no serious flaws. Studies with serious flaws or those in which fewer than six of the criteria were met were rated as having a "high risk of bias". Disagreement was resolved by discussion.

Clinical relevance

The five questions recommended by the Cochrane Back Review Group [20] were used to assess the clinical relevance of the included studies. Positive (+) indicated that the clinical relevance item was met, negative (−) if the item was not met, and unclear (?) if the data were inadequate to answer the question.

Data analysis

For dichotomous outcome data (fusion failure rate, reoperation, clinical improvements in ODI), relative risk (RR) and 95 % confidence interval (95 % CI) were calculated. For continuous outcomes, weighted mean difference (WMD) and 95 % CI were calculated. In order to be more

likely to reveal a difference between the two interventions, fixed-effect model was used when there was no significantly statistical heterogeneity. The random-effects model acknowledges that the effect of treatment may not be identical from study to study, because of heterogeneity and typically results in a more conservative estimate [21]. Therefore, the statistical data were analyzed with the use of random-effects model when heterogeneity existed. We assessed the statistical heterogeneity using Q statistic and I^2 statistic [22], with significance at 0.1 and 50 %, respectively. Funnel plot and statistic tests (Egger's test [23] and Begg's test [24]) were used to explore potential publication bias [25]. To assess the stability in the overall result if publication bias existed, we corrected the summary results by the trim and fill method [26]. Given the different influence on the summary effect owing to different surgical procedures or total doses of rhBMP-2, subgroup analyses were conducted. To detect whether the combined RRs within subgroups had significant differences, tests of interaction developed by Altman et al. [27] were performed. We also performed sensitivity analyses to assess the changes in the overall result by omitting studies. Data analyses were performed using Review Manager 5.1 and STATA 12.0.

Results

From computer and manual searching, 1,053 publications were obtained. Based on titles and abstracts, 576 publications were immediately excluded. In the potentially relevant 477 publications, 464 were omitted according to the inclusion criteria. Full texts of the remaining 13 studies were reviewed by authors. Two of them were duplicate publications on the same set of participants [28, 29] and one study [30] only reported the data of the control group. Finally, ten RCTs [14–18, 31–35] involving 1,342 patients were eligible for our inclusion criteria (Fig. 1).

Basic demographics and follow-up information are shown in Table 1. Patients in nine RCTs [14–18, 31–33, 35] had symptomatic, single-level lumbar degenerative disc disease (DDD) with low back or leg pain, or both, and failed to respond to nonoperative treatment for at least 6 months. Additionally, seven RCTs [14, 15, 18, 31–33, 35] had patients with spondylolisthesis classified no higher than Meyerding grade 1[36]. Glassman et al. [34] enrolled patients over 60 years of age with single-/two-level lumbar DDD, stenosis, and adjacent level fusion. Patients underwent ALIF in three studies [16–18], and instrumented posterolateral fusion (iPLF) or posterolateral interbody fusion (PLIF) in seven studies [14, 15, 31–35]. The total doses of rhBMP-2 in five studies were fixed [14, 15, 31–33], from 12 to 40 mg. In four studies [16–18, 35], the total
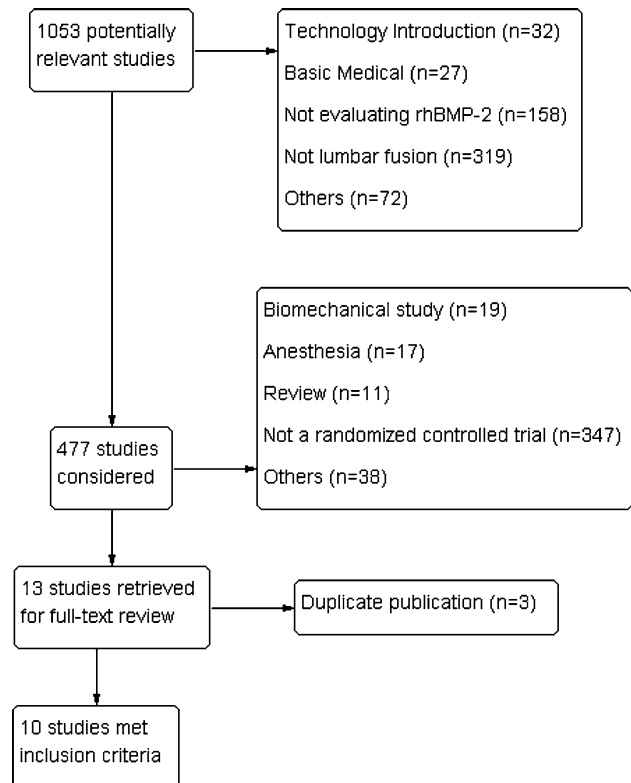

Fig. 1 A flowchart showed the study selection process

doses, from 1.95 to 12 mg, were varied by differently sized fusion devices. However, the concentration of rhBMP-2 remained constant within each study. In six studies [14–16, 31, 33, 35], all of the local bone removed as a result of the decompression was discarded; one study [34] used available local bone in all cases; the remaining three RCTs [17, 18, 32] did not give an explanation. Both the treatment and control groups used pedicle instrumentation in six studies, Cotrel-Dubousset Horizon pedicle screws and rods (CD Horizon Spinal System; Medtronic Sofamor Danek, Memphis, TN) were applied in four [15, 31–33], one study [14] used Texas Scottish Rite Hospital pedicle screw instrumentation (TSRH Spinal System; Medtronic Sofamor Danek, Memphis, TN), and one study [34] only described that all cases used the same screw/rod implant system. The details of the operative characteristics are presented in Table 2.

In one study [14], there were two treatment groups (rhBMP-2/TSRH group and rhBMP-2 only group) and one control group (ICBG/TSRH group). To avoid heterogeneity, rhBMP-2 only group was omitted from this meta-analysis.

The risk of bias assessment of the included studies is presented in Table 3. Given that the compliance of surgical treatment can be considered as irrelevant for interventions [20], Criteria 11 was scored as not applicable. Although the

Table 1 Basic characteristic of included studies

Study	Study design	Preoperative diagnosis	Sample size			Female (%)	Mean age (year)		Follow-up (month)	Follow-up rate (%)
			Treatment group	Control group	Total		Treatment group	Control group		
Boden et al. [14]	RCT[a]	Single-level lumbar DDD[b] Spondylolisthesis ≤ grade 1	11	5	16	68.8	57.6	52.9	24	100
Glassman et al. [32]	RCT	Single-level lumbar DDD Spondylolisthesis ≤ grade 1	38	36	74	59.5	53	53	12	97
Dimar et al. [15]	RCT	Single-level lumbar DDD Spondylolisthesis ≤ grade 1	NR[c]	NR	150	NR	50.9	52.7	24	65
Dimar 2nd et al. [31]	RCT	Single-level lumbar DDD Spondylolisthesis ≤ grade 1	239	224	463	56.2	53.2	52.3	24	89
Dawson et al. [33]	RCT	Single-level lumbar DDD Spondylolisthesis ≤ grade 1	25	21	46	58.7	55.9	56.9	24	87
Glassman et al. [34]	RCT	Single/multilevel lumbar DDD; spondylolisthesis; stenosis; adjacent level fusion	50	52	102	68.6	69.9	69.2	24	94.3
Haid et al. [35]	RCT	Single-level lumbar DDD Spondylolisthesis ≤ grade 1	34	33	67	52.2	46.3	46.1	24	94
Boden et al. [18]	RCT	Single-level lumbar DDD Spondylolisthesis ≤ grade 1	11	3	14	50	42.5	40.2	24	100
Burkus et al. [16]	RCT	Single-level lumbar DDD	143	136	279	47.7	43.3	42.3	24	91.7
Burkus et al. [17]	RCT	Single-level lumbar DDD	79	52	131	61.1	40.2	43.6	24	99.2

[a] *RCT* randomized controlled trial, [b] *DDD* degenerative disc disease, [c] *NR* not reported

study by Dimar et al. [15] met six criteria, this study was rated to be with "high risk of bias" because of its high dropout rate (35 %), which could be considered a serious flaw. Among the ten eligible RCTs, four trials [17, 18, 31, 33] reported adequate generation of the allocation sequence, and two trials [17, 31] reported allocation concealment. Only one trial [17] performed the care provider and outcome assessors blinding, and one trial [32] did not mention the blinding method. Finally, seven studies were rated to be with "low risk of bias" [14, 16–18, 31, 33, 35] and three studies were rated to be with "high risk of bias" [15, 32, 34].

Details on industry funds and financial interest are also presented in Table 3. Industry funds were received in eight studies [14, 15, 17, 18, 31–34] and one or more authors received financial interest in seven studies [14, 15, 18, 31–33, 35].

The clinical relevance of the included studies is presented in Table 4. In two trials [17, 32], complications as a main clinical outcome were not fully reported. In the study reported by Haid et al. [35], the likely treatment benefits were considered by reviewers not to be worth the potential harms. This was attributed mainly to the relatively low rate of satisfaction in the rhBMP-2 group (72.4 % vs. 80.0 %), as well as the significantly high rate of new bone formation

into the canal or neuroforamina (24 of 32 rhBMP-2 patients, 70.1 %; 4 of 31 ICBG patients, 12.9 %; $p < 0.0001$).

Fusion failure

Apart from plain or dynamic radiographs, all studies evaluated incorporated computed tomography (CT) scans, which may increase the accuracy of fusion assessment. Also, the time of postoperative rehabilitation may affect fusion rate. We pooled the relevant data at postoperative 6, 12 and 24 months. Seven studies reported data on fusion failure at 6 months [16–18, 31–33, 35], seven studies provided relevant data at 12 months [14, 16, 17, 31–33, 35], whereas, data at 24 months were obtained from nine studies [14–18, 31, 33–35]. The fusion failure rate at postoperative 6 months in the rhBMP-2 group was significantly lower than that of the ICBG group ($p < 0.0001$, RR = 0.55, 95 % CI = 0.42–0.72; Fig. 2). Significant differences were also found at 12 months ($p = 0.0003$, RR = 0.53, 95 % CI = 0.37–0.75) and 24 months ($p < 0.00001$, RR = 0.31, 95 % CI = 0.21–0.46; Fig. 3 and 4). Heterogeneity was absent or moderate during follow-up (6 months: $I^2 = 0$ %; 12 months: $I^2 = 46$ %; 24 month: $I^2 = 0$ %).

Table 2 Operative details of included studies

Study	Type of operation	No. of fused level	Local bone	Treatment group			Control group Bone graft	Fusion devices
				Total dose of RhBMP-2	Concentration of RhBMP-2	Osteoconductive bulking agent		
Boden et al. [14]	IPLF[a]	Single level	Discarded	40 mg	2 mg/ml	60 %HA–40 %TCP BCP granules (10 cm^3 per side)	ICBG	Pedicle instrumentation: TSRH
Glassman et al. [32]	iPLF	Single level	NR[d]	40 mg	2 mg/ml	HA-TCP CRM (10 cm^3 per side)	ICBG	Pedicle instrumentation: CD Horizon
Dimar et al. [15]	iPLF	Single level	Discarded	40 mg	2 mg/ml	HA-TCP CRM (10 cm^3 per side)	ICBG	Pedicle instrumentation: CD Horizon
Dimar 2nd et al. [31]	iPLF	Single level	Discarded	40 mg	2 mg/ml	15 %HA–85 % β-TCP CRM (10 cm^3 per side)	ICBG	Pedicle instrumentation: CD Horizon
Dawson et al. [33]	iPLF	Single level	Discarded	12 mg	1.5 mg/ml	15 %HA–85 %TCP ceramic granules (10 cm^3 per side)	ICBG	Pedicle instrumentation: CD Horizon
Glassman et al. [34]	iPLF	Single or multilevel	Used	NR	NR	–	ICBG	Pedicle instrumentation: NR
Haid et al. [35]	PLIF[b]	Single level	Discarded	4–8 mg	1.5 mg/ml	–	ICBG	Cylindrical threaded titanium fusion cages
Boden et al. [18]	ALIF[c]	Single level	NR	1.95–3.9 mg	1.5 mg/ml	–	ICBG	NOVUS LT tapered interbody fusion cage
Burkus et al. [16]	ALIF	Single level	Discarded	4.2–8.4 mg	1.5 mg/ml	–	ICBG	NOVUS LT tapered interbody fusion cage
Burkus et al. [17]	ALIF	Single level	NR	8.4–12 mg	1.5 mg/ml	–	ICBG	Threaded cortical bone dowels

[a] *iPLF* instrumental posterolateral fusion, [b] *PLIF* posterolateral interbody fusion, [c] *ALIF* anterior lumbar interbody fusion, [d] *NR* not reported

Reoperation

Reoperation rates were available from all included studies [14–18, 31–35]. The combined result showed a significantly lower rate of reoperation in the rhBMP-2 group in comparison to the ICBG group ($p = 0.0001$, RR = 0.52, 95 % CI = 0.37–0.72); no heterogeneity was found ($I^2 = 0$ %; Fig. 5).

Clinical improvement on ODI

Clinical improvements on ODI were summarized from four studies [16, 18, 33, 35]. Clinical improvement was defined as a decrease of at least 15 % from the preoperative ODI score. The summary RR and 95 % CI showed a strong favorable trend in the rhBMP-2 group, although no statistically significant difference was found ($p = 0.12$, RR = 0.73, 95 % CI = 0.49–1.08). There was no significant heterogeneity between trials ($I^2 = 0$ %; Fig. 6).

Clinical outcomes

Clinical outcomes at postoperative 24 months are shown in Table 5. Not surprisingly, the improvement in each clinical outcome at 24 months follow-up interval, as compared to preoperative scores, was statistically significant for both intervention groups. Eight studies [15–18, 31, 33–35] reported data on the mean improvement of the ODI score. Seven of them [15–18, 31, 33, 34] compared the mean improvement of the two intervention groups (p value) and none of them was significant. Two studies by Dimar 2nd et al. [31] and Burkus et al. [16] only depicted that the ODI scores were similar in both groups over all time intervals and did not report a specific p value. Four studies [15, 17, 33, 34] presented and compared the mean improvement of the SF-36 PCS, and only one study [17] showed a significant difference ($p = 0.015$). One study by Dimar 2nd et al. [31] did not provide the corresponding data, but demonstrated that the SF-36 PCS scores were similar in both groups over all time intervals. Eight studies [14–17, 31,

Table 3 Risk of bias assessment and conflicts of interest of included studies

Category of criterion	Boden et al. [14]	Glassman et al. [32]	Dimar et al. [15]	Dimar 2nd et al. [31]	Dawson et al. [33]	Glassman et al. [34]	Haid et al. [35]	Boden et al. [18]	Burkus et al. [16]	Burkus et al. [17]
1. The use of adequate randomization	Not sure	Not sure	Not sure	Yes	Yes	Not sure	Not sure	Yes	Not sure	Yes
2. The use of concealed allocation	No	No	No	Yes	No	No	No	No	No	Yes
3. The use of blinded patients	No	No	No	Not sure	Not sure	No	No	No	No	Not sure
4. The use of blinded care providers	No	No	No	Not sure	Not sure	No	No	No	No	Yes
5. The use of blinded outcome assessors	Yes	Not sure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6. The dropout rate was described and acceptable	Yes	Yes	No[b]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
7.The use of intention-to-treatment analysis	Yes	No	Yes	No	No	No	Not sure	Yes	No	Yes
8. Selective outcome reporting was avoided	Yes	Not sure	Yes	Yes	Yes	Not sure	Yes	Yes	Yes	Yes
9. The groups were similar at baseline	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
10. Co-interventions were avoided or similar	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
11. The compliance was acceptable	NA[a]	NA	NA	NA	NA	NA	NA	NA	NA	NA
12. The timing of the outcome assessment was similar in all groups	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Score and risk of bias	7/low	4/high	6/high	8/low	7/low	5/high	6/low	8/low	6/low	10/low
Industry funds	Yes	Yes	Yes	Yes	Yes	Yes	NR[c]	Yes	NR	Yes
Financial interest	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	NR	No

[a] *NA* not applicable, [b]dropout rate exceeding 30 %, [c]*NR* not reported

Table 4 Clinical relevance

Question	Boden et al. [14]	Glassman et al. [32]	Dimar et al. [15]	Dimar 2nd et al. [31]	Dawson et al. [33]	Glassman et al. [34]	Haid et al. [35]	Boden et al. [18]	Burkus et al. [16]	Burkus et al. [17]
1. Are the patients described in detail, so that you can decide whether they are comparable to those that you see in your practice?	+	−	+	+	+	+	+	+	+	+
2. Are the interventions and treatment settings described well enough so that you can provide the same for your patients?	+	+	+	+	+	?	+	+	+	+
3. Were all clinically relevant outcomes measured and reported?	+	−	+	+	+	+	+	+	+	−
4. Is the size of the effect clinically important?	+	?	+	+	+	+	+	+	+	+
5. Are the likely treatment benefits worth the potential harms?	+	?	+	+	+	+	−	+	+	+
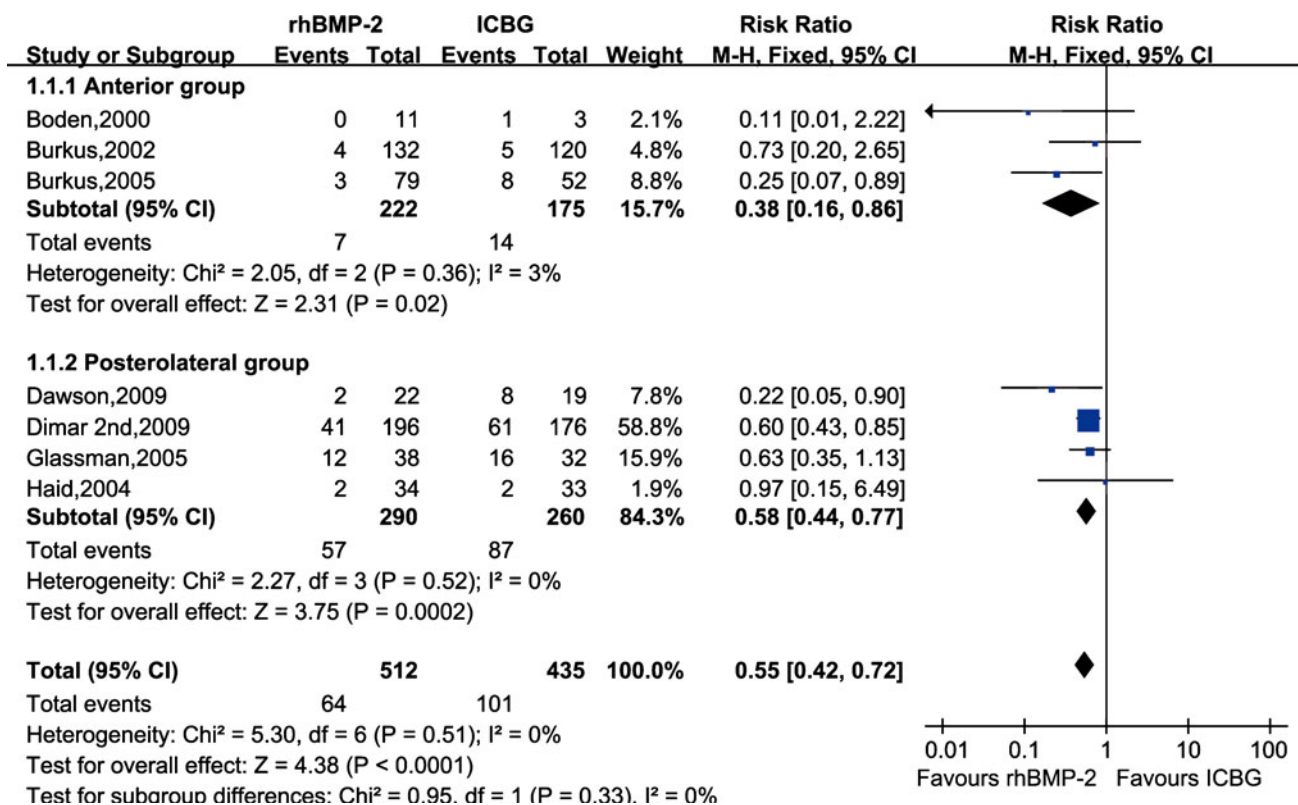
Fig. 2 Forest plot of fusion failure at 6 months

Fig. 3 Forest plot of fusion failure at 12 months

Study or Subgroup	rhBMP-2 Events	Total	ICBG Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI	Risk Ratio M-H, Fixed, 95% CI
3.1.1 Anterior group							
Boden,2000	0	11	1	3	2.4%	0.11 [0.01, 2.22]	
Burkus,2002	7	127	13	115	14.6%	0.49 [0.20, 1.18]	
Burkus,2005	1	79	12	52	15.4%	0.05 [0.01, 0.41]	
Subtotal (95% CI)		217		170	32.4%	0.25 [0.12, 0.52]	
Total events	8		26				
Heterogeneity: Chi² = 4.63, df = 2 (P = 0.10); I² = 57%							
Test for overall effect: Z = 3.72 (P = 0.0002)							
3.1.2 Posterolateral group							
Boden,2002	0	11	2	5	3.6%	0.10 [0.01, 1.77]	
Dawson,2009	1	20	6	18	6.7%	0.15 [0.02, 1.13]	
Dimar 2nd,2009	8	194	18	169	20.5%	0.39 [0.17, 0.87]	
Dimar,2006	5	53	12	45	13.8%	0.35 [0.13, 0.93]	
Glassman,2008	6	49	15	51	15.7%	0.42 [0.18, 0.99]	
Haid,2004	2	31	7	33	7.2%	0.30 [0.07, 1.35]	
Subtotal (95% CI)		358		321	67.6%	0.34 [0.22, 0.53]	
Total events	22		60				
Heterogeneity: Chi² = 1.67, df = 5 (P = 0.89); I² = 0%							
Test for overall effect: Z = 4.67 (P < 0.00001)							
Total (95% CI)		575		491	100.0%	0.31 [0.21, 0.46]	
Total events	30		86				
Heterogeneity: Chi² = 6.20, df = 8 (P = 0.63); I² = 0%							
Test for overall effect: Z = 5.96 (P < 0.00001)							
Test for subgroup differences: Chi² = 0.45, df = 1 (P = 0.50), I² = 0%							

0.005   0.1   1   10   200
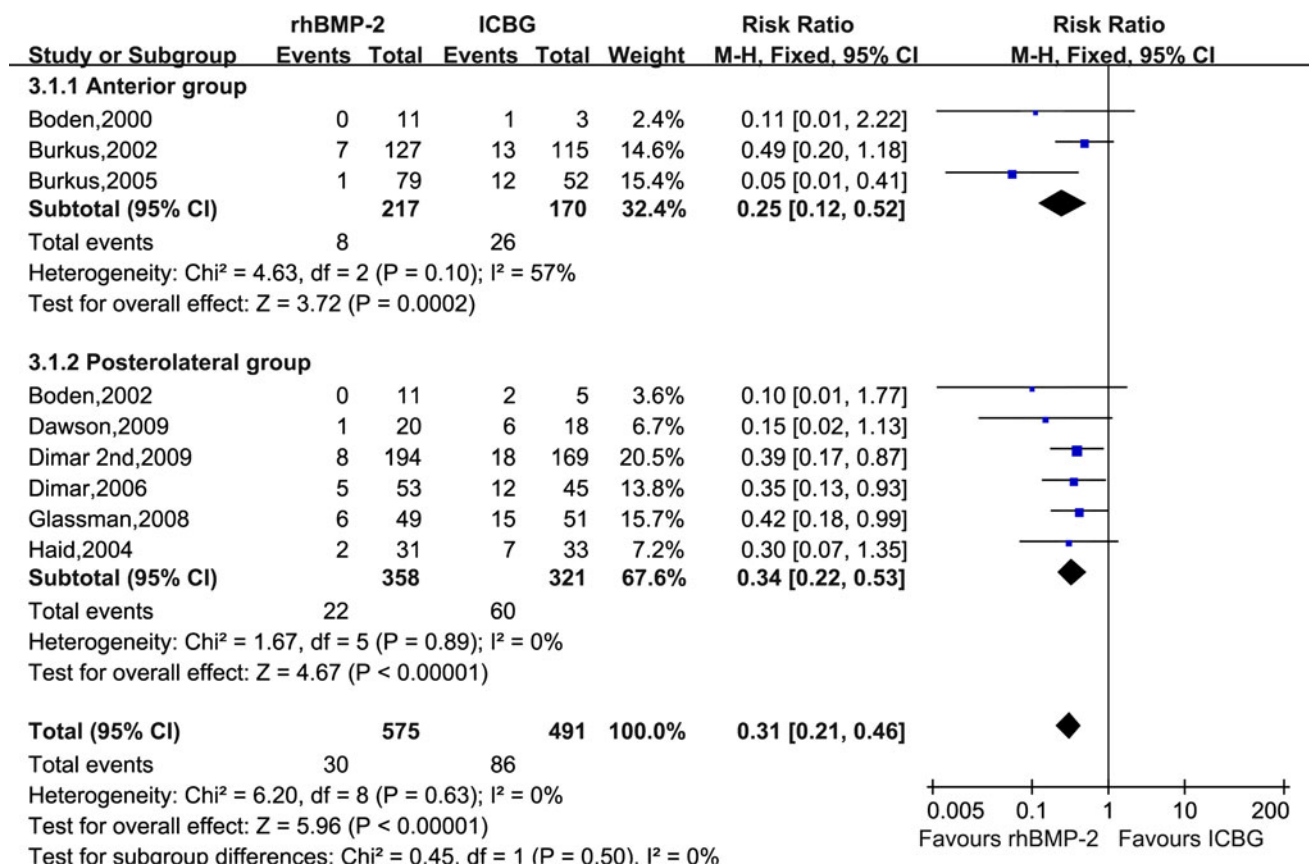Favours rhBMP-2   Favours ICBG

Fig. 4 Forest plot of fusion failure at 24 months

33–35] provided data on the mean improvement of the NRS back pain score. Of these studies, compared with ICBG, two RCTs [17, 35] showed a significant tendency toward better improvement with the use of rhBMP-2 ($p = 0.009$ and $p = 0.032$, respectively). In the study by Burkus et al. [17], patients treated with rhBMP-2 had a significantly improved NRS leg pain score compared to those treated by ICBG. But there were no significant differences in the remaining six studies which reported the corresponding data. [15, 16, 31, 33–35].

Subgroup analysis

In subgroup analyses, we divided all included studies into two groups, namely "posterolateral fusion group" and "anterior fusion group." At 6 months follow-up, the summary RRs on fusion failure were statistically significant (anterior subgroup: $p = 0.02$, RR = 0.38, 95% CI = 0.16–0.86; posterolateral subgroup: $p = 0.0002$, RR = 0.58, 95% CI = 0.44–0.77). No heterogeneity was observed within either subgroup ($I^2 = 3$% and $I^2 = 0$%, respectively; Fig. 2). At postoperative 12 months, there was significant difference between the two intervention groups on fusion failure in either the anterior subgroup ($p = 0.01$,

RR = 0.28, 95% CI = 0.11–0.74) or the posterolateral subgroup ($p = 0.007$, RR = 0.59, 95% CI = 0.41–0.87). However, heterogeneity was found in the posterolateral subgroup ($p = 0.09$, $I^2 = 51$%; Fig. 3). At 24 months follow-up, significant differences were found in both subgroups (anterior subgroup: $p = 0.0002$, RR = 0.25, 95% CI = 0.12–0.52; posterolateral subgroup: $p < 0.00001$, RR = 0.34, 95% CI = 0.22–0.53). Heterogeneity was present in the anterior subgroup ($p = 0.10$, $I^2 = 57$%; Fig. 4). The result of subgroup analysis on reoperation has demonstrated that, in both subgroups, the rate of reoperation in rhBMP-2 group was significantly lower than that of the ICBG group (anterior subgroup: $p = 0.02$, RR = 0.48, 95% CI = 0.26–0.89; posterolateral subgroup: $p = 0.002$, RR = 0.54, 95% CI = 0.36–0.79). However, heterogeneity was found in the anterior subgroup ($p = 0.12$, $I^2 = 52$%; Fig. 5). With regard to the ODI clinical improvement, the summary RRs stratified by the type of operative procedure were also not statistically significant (anterior subgroup: $p = 0.59$, RR = 0.86, 95% CI = 0.50–1.49; posterolateral subgroup: $p = 0.07$, RR = 0.58, 95% CI = 0.32–1.04), with no heterogeneity (Fig. 6).

We repeated subgroup analysis in accordance with the different total doses of rhBMP-2. One study reported by
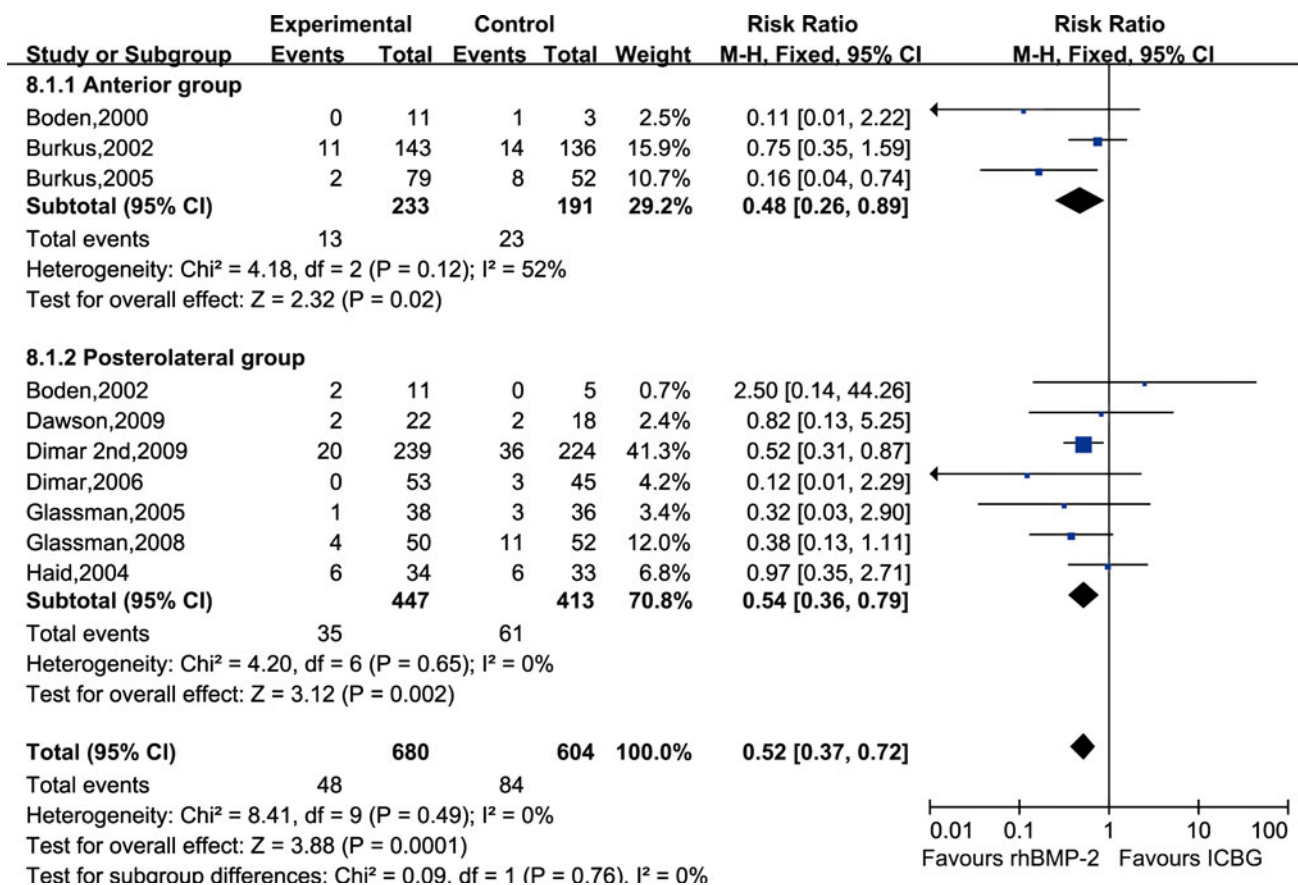
| Study or Subgroup | Experimental | | Control | | | Risk Ratio | Risk Ratio |
	Events	Total	Events	Total	Weight	M-H, Fixed, 95% CI	M-H, Fixed, 95% CI
8.1.1 Anterior group							
Boden,2000	0	11	1	3	2.5%	0.11 [0.01, 2.22]	
Burkus,2002	11	143	14	136	15.9%	0.75 [0.35, 1.59]	
Burkus,2005	2	79	8	52	10.7%	0.16 [0.04, 0.74]	
Subtotal (95% CI)		233		191	29.2%	0.48 [0.26, 0.89]	
Total events	13		23				
Heterogeneity: Chi² = 4.18, df = 2 (P = 0.12); I² = 52%							
Test for overall effect: Z = 2.32 (P = 0.02)							
8.1.2 Posterolateral group							
Boden,2002	2	11	0	5	0.7%	2.50 [0.14, 44.26]	
Dawson,2009	2	22	2	18	2.4%	0.82 [0.13, 5.25]	
Dimar 2nd,2009	20	239	36	224	41.3%	0.52 [0.31, 0.87]	
Dimar,2006	0	53	3	45	4.2%	0.12 [0.01, 2.29]	
Glassman,2005	1	38	3	36	3.4%	0.32 [0.03, 2.90]	
Glassman,2008	4	50	11	52	12.0%	0.38 [0.13, 1.11]	
Haid,2004	6	34	6	33	6.8%	0.97 [0.35, 2.71]	
Subtotal (95% CI)		447		413	70.8%	0.54 [0.36, 0.79]	
Total events	35		61				
Heterogeneity: Chi² = 4.20, df = 6 (P = 0.65); I² = 0%							
Test for overall effect: Z = 3.12 (P = 0.002)							
Total (95% CI)		680		604	100.0%	0.52 [0.37, 0.72]	
Total events	48		84				
Heterogeneity: Chi² = 8.41, df = 9 (P = 0.49); I² = 0%							
Test for overall effect: Z = 3.88 (P = 0.0001)							
Test for subgroup differences: Chi² = 0.09, df = 1 (P = 0.76), I² = 0%							

0.01 0.1 1 10 100
Favours rhBMP-2 Favours ICBG

Fig. 5 Forest plot of reoperation

| Study or Subgroup | rhBMP-2 | | ICBG | | | Risk Ratio | Risk Ratio |
	Events	Total	Events	Total	Weight	M-H, Fixed, 95% CI	M-H, Fixed, 95% CI
6.4.1 Anterior group							
Boden,2000	1	11	1	3	3.6%	0.27 [0.02, 3.19]	
Burkus,2002	20	127	20	115	48.2%	0.91 [0.51, 1.59]	
Subtotal (95% CI)		138		118	51.9%	0.86 [0.50, 1.49]	
Total events	21		21				
Heterogeneity: Chi² = 0.87, df = 1 (P = 0.35); I² = 0%							
Test for overall effect: Z = 0.53 (P = 0.59)							
6.4.2 Posterolateral / Posterior group							
Dawson,2009	2	23	6	20	14.8%	0.29 [0.07, 1.28]	
Haid,2004	10	31	15	33	33.4%	0.71 [0.38, 1.34]	
Subtotal (95% CI)		54		53	48.1%	0.58 [0.32, 1.04]	
Total events	12		21				
Heterogeneity: Chi² = 1.23, df = 1 (P = 0.27); I² = 19%							
Test for overall effect: Z = 1.83 (P = 0.07)							
Total (95% CI)		192		171	100.0%	0.73 [0.49, 1.08]	
Total events	33		42				
Heterogeneity: Chi² = 2.67, df = 3 (P = 0.45); I² = 0%							
Test for overall effect: Z = 1.57 (P = 0.12)							
Test for subgroup differences: Chi² = 0.93, df = 1 (P = 0.33), I² = 0%							

0.05 0.2 1 5 20
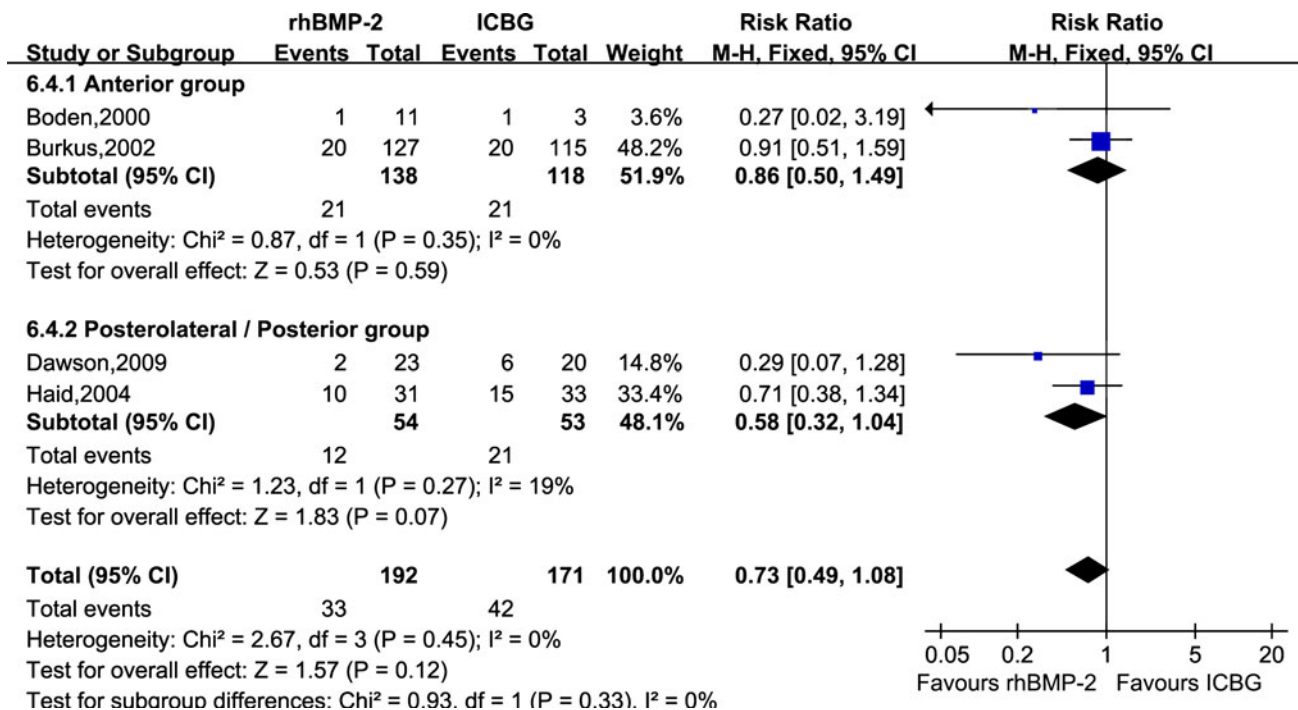Favours rhBMP-2 Favours ICBG

Fig. 6 Forest plot of the Oswestry Disability Index at 12–24 months

Table 5 Clinical outcomes

Study	The Oswestry Disability Index		The SF-36 physical component summary		The numeric rating scores for back pain		The numeric rating scores for leg pain	
	Mean improvement at 24 months RhBMP-2 group, ICBG group	p value[a]	Mean improvement at 24 months RhBMP-2 group, ICBG group	p value[a]	Mean improvement at 24 months RhBMP-2 group, ICBG group	p value[a]	Mean improvement at 24 months RhBMP-2 group, ICBG group	p value[a]
Boden et al. [14]	–	–	–	–	-5.1 ± 2.1; -6.2 ± 2.1	0.37	–	–
Glassman et al. [32]	–	–	–	–	–	–	–	–
Dimar et al. [15]	−24.5; −21.4	0.445	8.6; 10.7	0.378	−6.6; −7.4	0.685	−5.7; −5.2	0.567
Dimar 2nd et al. [31]	–	>0.05	–	>0.05	−8.5; −8.0	0.145	−7.8; −7.3	0.214
Dawson et al. [33]	−28.2; −23.0	0.953	13; 9.9	0.927	−9.6; −7.2	0.664	−9.3; −7.2	0.892
Glassman et al. [34]	-15.8 ± 17.7; -13.0 ± 15.5[b]	0.40	6.6 ± 9.3; 7.5 ± 8.4[b]	0.95	-3.5 ± 3.3; -3.4 ± 3.5[b]	0.88	-4.0 ± 3.4; -3.3 ± 3.9[b]	0.34
Haid et al. [35]	−29.6; −24.9	–	–	–	−9; −4.5	0.009	−7.7; −6.5	>0.05
Boden et al. [18]	−25; −15	0.12	–	–	–	–	–	–
Burkus et al. [16]	−29; −29.5	>0.05	–	–	−8.4; −8.1	–	−6.5; −5.9	>0.05
Burkus et al. [17]	−33.4; −27.0	0.119	15.7; 11.6	0.015	−8.6; −7.1	0.032	−6.8; −4.9	0.011

[a] p value > 0.05 indicates that the difference between the two intervention groups was not statistically significant

[b] Mean improvement ± standard deviation/standard error

Glassman et al. [34] was omitted because the total dose of rhBMP-2 was not provided. Of the remaining nine studies, four studies [14, 15, 31, 32] and five studies [16–18, 33, 35] were allocated to the "high-dose subgroup (40 mg)" and "low-dose subgroup (<40 mg)", respectively. On the primary outcome, both subgroups showed a significant reduction on the fusion failure rate at postoperative 12–24 months (high-dose subgroup: $p < 0.0001$, RR = 0.34, 95 % CI = 0.20–0.57; low-dose subgroup: $p < 0.00001$, RR = 0.25, 95 % CI = 0.13–0.46) and no heterogeneity was found (Fig. 7).

Adverse events

Eight studies reported data on adverse events [14–16, 18, 31, 33–35]. All the authors of the included studies declared that no unanticipated rhBMP-2-related adverse events occurred in either intervention group. Owing to the different data formats, meta-analysis was not performed. The details are presented in Table 6. Glassman et al. [34] reported that multiple complications were only observed in six ICBG patients, but in none of the rhBMP-2 patients. A rate of retrograde ejaculation (RE) of 4.1 % (6 of 146)

was reported by Burkus et al. [16] in the entire cohort who underwent ALIF, without comparison between the two intervention groups.

Sensitivity analysis

To evaluate whether the studies rated to be with high risk of bias significantly affected our results, we performed sensitivity analysis by excluding these studies [15, 32, 34]. After excluding these studies, the summary RR of fusion failure at 12 months was 0.57 (95 % CI = 0.39–0.82) and at 24 months was 0.28 (95 % CI = 0.17–0.45); neither was significantly different from previous RRs.

Test of interaction

On the fusion failure rate, RRs in the respective subgroups at each postoperative interval were compared and no significant differences were found at 6 months follow-up ($Z = -0.93$, $p = 0.176$, 95 % CI = 0.27–1.58), at 12 months follow-up ($Z = -1.43$, $p = 0.076$, 95 % CI = 0.17–1.32), or at 24 months follow-up ($Z = -0.7$, $p = 0.242$, 95 % CI = 0.31–1.73).
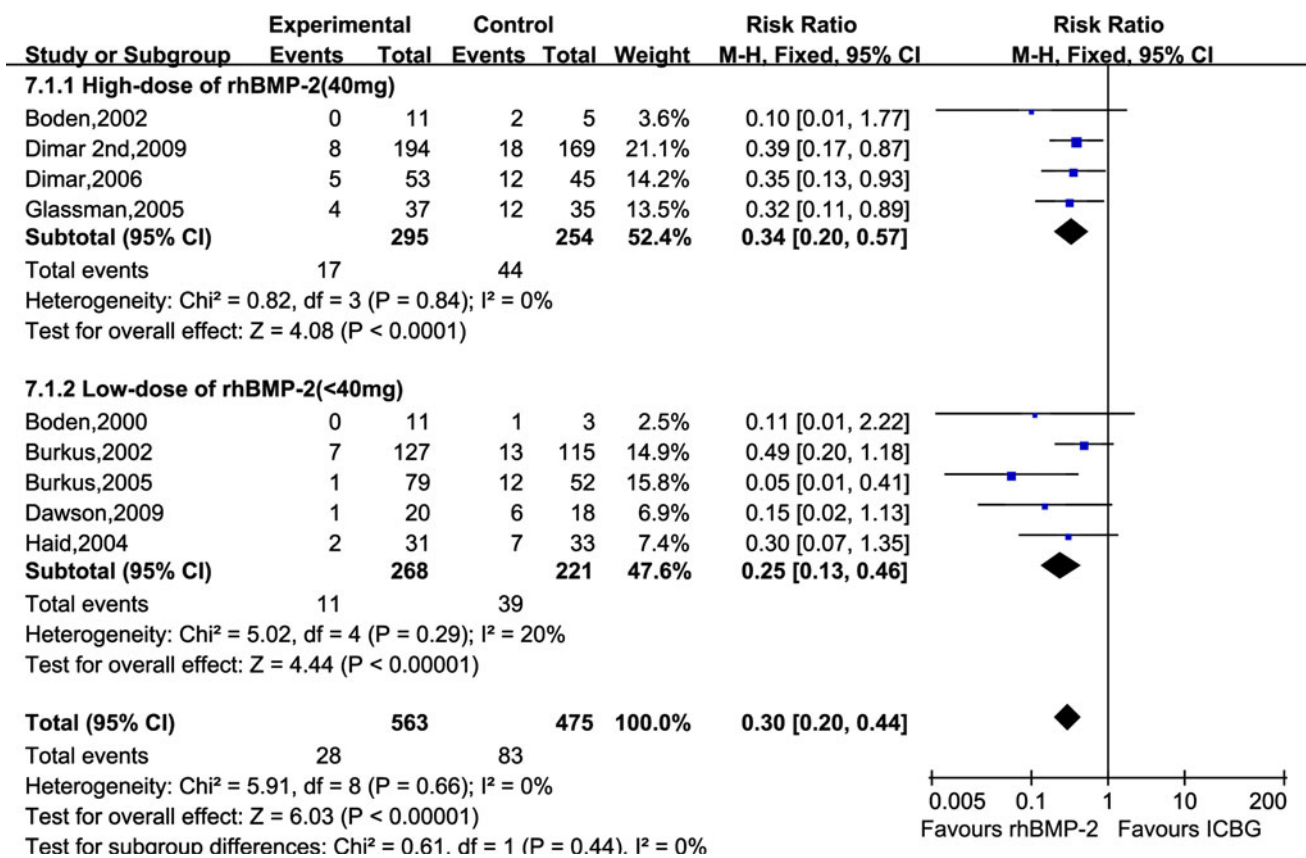
Study or Subgroup	Experimental Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI	Risk Ratio M-H, Fixed, 95% CI
7.1.1 High-dose of rhBMP-2(40mg)							
Boden,2002	0	11	2	5	3.6%	0.10 [0.01, 1.77]	
Dimar 2nd,2009	8	194	18	169	21.1%	0.39 [0.17, 0.87]	
Dimar,2006	5	53	12	45	14.2%	0.35 [0.13, 0.93]	
Glassman,2005	4	37	12	35	13.5%	0.32 [0.11, 0.89]	
Subtotal (95% CI)		295		254	52.4%	0.34 [0.20, 0.57]	
Total events	17		44				
Heterogeneity: Chi² = 0.82, df = 3 (P = 0.84); I² = 0%							
Test for overall effect: Z = 4.08 (P < 0.0001)							
7.1.2 Low-dose of rhBMP-2(<40mg)							
Boden,2000	0	11	1	3	2.5%	0.11 [0.01, 2.22]	
Burkus,2002	7	127	13	115	14.9%	0.49 [0.20, 1.18]	
Burkus,2005	1	79	12	52	15.8%	0.05 [0.01, 0.41]	
Dawson,2009	1	20	6	18	6.9%	0.15 [0.02, 1.13]	
Haid,2004	2	31	7	33	7.4%	0.30 [0.07, 1.35]	
Subtotal (95% CI)		268		221	47.6%	0.25 [0.13, 0.46]	
Total events	11		39				
Heterogeneity: Chi² = 5.02, df = 4 (P = 0.29); I² = 20%							
Test for overall effect: Z = 4.44 (P < 0.00001)							
Total (95% CI)		563		475	100.0%	0.30 [0.20, 0.44]	
Total events	28		83				
Heterogeneity: Chi² = 5.91, df = 8 (P = 0.66); I² = 0%							
Test for overall effect: Z = 6.03 (P < 0.00001)							
Test for subgroup differences: Chi² = 0.61, df = 1 (P = 0.44), I² = 0%							

0.005　0.1　1　10　200
Favours rhBMP-2　Favours ICBG

Fig. 7 Forest plot of the subgroup analysis on different total doses of rhBMP-2

Publication bias

The funnel plot of fusion failure at 24 month is presented in Fig. 8. Evidences of publication bias were found in both Egger's test ($p = 0.003 < 0.1$) and Begg's test ($p = 0.048 < 0.05$). After correcting for publication bias by the trim and fill method, the summary RR in either fixed-effect model or in random-effect model was 0.340 (95 % CI = 0.229–0.505), which was not materially different from the uncorrected result (0.31, 95 % CI = 0.21–0.46).

Discussion

Ten RCTs involving 1,342 patients that compared rhBMP-2 with ICBG for lumbar fusion were identified in this meta-analysis. Compared with ICBG, the use of rhBMP-2 significantly reduces the risk of fusion failure and reoperation at all time intervals. Subgroup analyses stratified by the type of surgical procedure yielded similar results. A greater ODI clinical improvement was found in the rhBMP-2 group when compared with that in the ICBG group, but there was no significance. The results from the test of

interaction demonstrated that the use of rhBMP-2 in each approach obtained higher fusion rates than with ICBG, but there was no significant evidence to support a different treatment effect in the different approaches. In clinical practice, the most important reason why patients consider fusion surgery is for pain relief and return of function. In this meta-analysis, all included studies, which reported clinical outcomes (the ODI score, the SF-36 PCS score, or the NRS back pain and leg pain score), showed a significant improvement in the rhBMP-2 group when compared with the preoperative scores. However, significant differences between the rhBMP-2 group and the ICBG group on clinical outcomes were only found in the SF-36 PCS score reported by Burkus et al. [17], in the NRS back pain score reported by Burkus et al. [17] and Haid et al. [35], and in the NRS leg pain score reported by Burkus et al. [17].

In contrast to the anterior approach, variable factors in the posterolateral approach, such as extensive soft tissue stripping, limited bony surface area, and compression due to tension of the posterior muscles, produce a challenging environment for attaining robust fusion [32, 37].Valuable experience from previous nonhuman primate trials in posterolateral spinal arthrodesis [38–40] has demonstrated that a stand-alone collagen sponge carrier was insufficient

Table 6 Category of adverse events

Study	Treatment group	Control group
Boden et al. [14]	Leg pain (1), epidural hematoma (1)	–
Dimar et al [15]	Gastrointestinal (9), traumas (14), cardio/vascular (9), urogenital (6), dural tears (3)	Gastrointestinal (10), traumas (9), cardio/vascular (6), urogenital (6), duraltears (5), malpositioned implants (3), surgical infection (1), vertebral fracture (1)
Dimar 2nd et al. [31]	Operative adverse event (20)	Operative adverse event (20)
Dawson et al. [33]	Incidental durotomy (1), wound infection (1)	Incidental durotomy (1), wound infection (2)
Glassman et al. [34]	Cardiac (1), wound infection (1), line-related sepsis (1), gastrointestinal (2), urinary tract infection (1), shingles (1)	Cardiac (7), wound infection (4), back or leg pain (3), gastrointestinal (3), urinary tract infection (1), neurologic deficit (1)
Haid et al. [35]	Dural tears (3), neurological complication (13)	Deep venous thrombosis (1), dural tears (2), donor-site pain (1), hematoma (1), others (16)
Boden et al. [18]	Postop ileus (1), wound dehiscence (1), back pain (1)	Postop ileus (1), urinary retention (1)
Burkus et al. [16][a]	Intraoperative vascular events (6)	Deep venous thrombosis (2), intraoperative vascular events (5), injury to the lateral femoral cutaneous nerve (3), fracture of iliac crest (2), donor sit infection (1), hematoma (1)

The number in parentheses indicates the frequency of adverse events

[a] Six male patients complained of retrograde ejaculation after surgery in the entire cohort
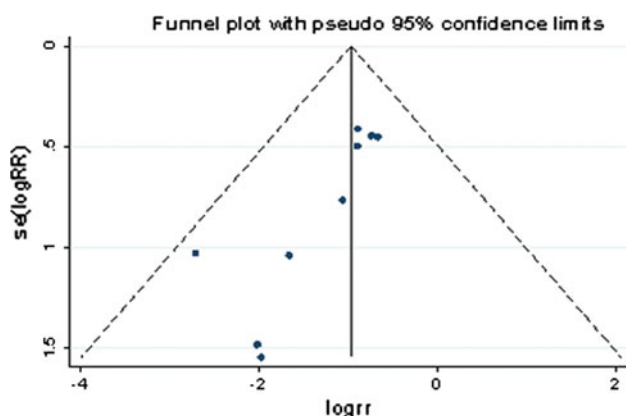
Fig. 8 Funnel plot on fusion failure

to obtain solid fusion. Therefore, for this approach, surgeons should use specific doses of rhBMP-2 combined with specially designed carriers. RhBMP-2 stimulates bone formation and healing by promoting the differentiation of primitive mesenchymal cells into osteoblasts by a cascade mechanism [41, 42]. Moreover, this effect requires a relatively stable environment. Osteoconductive bulking agents, such as biphasic calcium phosphate (BCP) granules and HA-TCP compression-resistant matrix (CRM), have been proven to resist the compression force of the posterior musculature and to ensure close contact between rhBMP-2 and the surface of the cancellous bone [43, 44]. Therefore, rhBMP-2 with osteoconductive materials can maximize the

effect of osteoinduction. In this meta-analysis, studies that used osteoconductive materials in posterolateral lumbar fusion had successful fusion rates at 2 years follow-up (88–100 %), which is consistent with the results of previous animal trials [38, 43–45].

The purpose of this meta-analysis was to evaluate the effectiveness and, more importantly, safety of rhBMP-2 compared with ICBG in lumbar fusion. Unfortunately, due to the lack of data and the inconsistent data formats, a quantitative analysis of complications associated with rhBMP-2 use cannot be conducted. Despite that none of the authors of the studies included in our analysis directly attributed complications to the use of rhBMP-2 and its potential advantages, the use of rhBMP-2 has been associated with various potential complications.

Haid et al. [35] identified unintended bone formation outside of the disc space by thin-cut CT scans in 24 of 32 rhBMP-2 patients (75 %) who underwent single-level PLIF as compared to 4 of 31(13 %) control patients. Although the authors stated that these findings were not associated with adverse outcomes, a difference between incidences with highly statistical significance ($p < 0.0001$) was revealed. To our knowledge, the formation of heterotopic bone may occur when rhBMP-2 spills from carriers into unwanted sites [46]. But these complications were not included in the author's comments on adverse events related to rhBMP-2. In a case report, five patients with ectopic bone formation in the spinal canal and neurological

complaints after either PLIF or TLIF using rhBMP-2 were identified by Wong et al. [47]. Three patients underwent an extensive revision surgery. Similarly, Chen et al. [48] presented four cases (4/147, 3 %) of bone formation in and around the neural foramen following the use of rhBMP-2 in TLIF. These patients reported neurological complaints and three of them required decompression. In view of this, the possible complication should be recognized when rhBMP-2 is used.

Bony resorption or osteolysis which presumably results from excessive osteoclastic activity [46, 49] is another potential complication, although resorption or osteolysis may represent a normal remodeling process in a progression to fusion [46]. As we know, severe resorption or osteolysis can also cause problems with graft subsidence and result in mechanical failure. In a 6-year follow-up study on 279 patients originally reported in 2002 [16], Burkus et al. [50] revealed seven adverse subsidence events (5.4 %) and four of these required additional surgery. Interestingly, all these adverse events were reported to have occurred within the first 2 years [50], but were not reported in the original 2-year follow-up industry-sponsored publication [16]. In a nonindustry-supported prospective study, Vaidya et al. [51] reported significant lucency and subsidence in 70 % (14 of 20) of ALIF levels with rhBMP-2 (27 % had >10 % subsidence). However, the mean subsidence without rhBMP-2 was 6 %.

Retrograde ejaculation (RE), which is secondary to trauma to the superior hypogastric plexus in the retroperitoneal space [52], also arouses our attention. In the RCT of ALIF comparing rhBMP-2 against ICBG with the LT-cage, Burkus et al. [16] reported a rate of RE of 4.1 % (6 of 146) in the entire cohort, without comparison between the two intervention groups. However, reanalyzing relevant data from the 2002 Food and Drug Administration document on the same cohort [53], Smoljanovic et al. [54] reported a high rate of RE associated with rhBMP-2 (7.9 %, rhBMP-2 group vs. 1.4 %, ICBG group; Fisher exact $p = 0.05$). Similarly, Carragee et al. [55] conducted a retrospective analysis on 243 patients who underwent ALIF with or without rhBMP-2; five RE events (7.2 %) were reported in the rhBMP-2 group and one (0.6 %) in the ICBG group (Fisher exact $p = 0.0025$). Although the incidence of RE might vary by approach, technical level and comorbidities (such as diabetes), findings with statistical significance indicated a strong association of rhBMP-2 with RE.

The risk of malignancy with the use of high-dose (40 mg) rhBMP-2 in iPLF has already been highlighted by the FDA [56]. Dimar 2nd et al. [31] reported on 463 patients undergoing iPLF using a dose and concentration of rhBMP-2 (40 mg, 2 mg/ml), which was significantly higher than the dose and concentration in clinically available rhBMP-2/ACS (12 mg, 1.5 mg/ml). The authors

found no difference in complication rates between the two intervention groups, and reported no BMP-related complications. In a retrospective case series of 1,037 cases (including the participants in the study by Dimar 2nd et al. [31]) in 2011, relevant statistical data from three studies [57–59] in which very large doses of rhBMP-2 (36–320 mg, mean: 91.2 mg) were used in spinal fusion were quoted by Glassman et al. [60] to support a conclusion that the use of high-dose rhBMP-2 did not increase the incidence of associated complications. However, on the same cohort in the study reported by Dimar 2nd et al. [31], the FDA document in 2010 reported a 3.8 % rate (9/239) incidence of new malignancy with the use of high-dose (40 mg) rhBMP-2 compared with a 0.9 % rate (2/224) in controls (NNH < 33), meaning that there may be a real association with approximately 90–95 % probability [61]. Unfortunately, this finding was not mentioned by the original authors [31, 60].

One of the major concerns for surgeons is that the use of a new technology should bring benefits to patients while minimizing the damage. Our results of this meta-analysis demonstrated that the use of rhBMP-2 could significantly decrease the risk of fusion failure and reoperation, as well as achieve significant improvement in various clinical outcomes when compared with the preoperative score. Owing to the inconsistent statistics and conclusions on the rhBMP-2 morbidity, the presentation of rhBMP-2 complication in the original industry-sponsored trials did not seem to fully reflect the data available from the FDA document and subsequent independent studies. In a systematic review focusing on the safety of rhBMP-2, Carragee et al. [61] concluded that original, industry-sponsored trials underestimated rhBMP-2-related adverse events. Evidence from multiple independent studies [47, 48, 51, 54, 55] indicated that there appears to have been an increased risk of uncommon and serious complications with the use of rhBMP-2 in lumbar fusion. Therefore, in sum, it is difficult for us to determine the true nature and frequency of complication associated with rhBMP-2.

There are several limitations in this meta-analysis. First, only ten RCTs were identified and were limited to English as publication language. As a result, publication bias exists. Although the treatment's effect may be overestimated [23], the results after correcting for this bias were not materially different from the uncorrected results. Second, three of ten included studies were rated to be with "high risk of bias". Furthermore, almost all included studies reported conflicts of interest except the one in 2002 of Burkus et al. [16]. As a result, to some extent, the presence of conflicts of interest and the potential possibility for reporting bias may affect the authenticity of results. However, this deviation can hardly be distinguished by the risk of bias assessment form recommended by the CBRG [20]. As a consequence of

these factors, additional high-quality independent RCTs with long-term outcomes and concerns about complication are needed to further evaluate the efficacy and safety of rhBMP-2.

Conclusions

Compared with ICBG, the use of rhBMP-2 can significantly increase fusion success rate and simultaneously decrease the risk of reoperation. However, compared with those original publications, the subsequent studies and the FDA documents have presented a completely different pattern of complications associated with rhBMP-2 use. The complications seem to be more common and serious than reported in these included studies. Therefore, for surgeons, the advantages of rhBMP-2 as an alternative to ICBG in lumbar surgery must be cautiously weighed against the disadvantages. Furthermore, we must note that great efforts should be made to further define the relationship between rhBMP-2 use and potential complications.

References

1. Kaufman HH, Jones E (1989) The principles of bony spinal fusion. Neurosurgery 24(2):264–270. doi:10.1097/00006123-198902000-00018
2. Fujiya M, Saita M, Kaneda K, Abumi K (1990) Clinical study on stability of combined distraction and compression rod instrumentation with posterolateral fusion for unstable degenerative spondylolisthesis. Spine 15(11):1216–1222. doi:10.1097/00007632-199011010-00023
3. Blumenthal SL, Baker J, Dossett A, Selby DK (1988) The role of anterior lumbar fusion for internal disc disruption. Spine 13(5):566–569. doi:10.1097/00007632-198805000-00023
4. Arrington ED, Smith WJ, Chambers HG, Bucknell AL, Davino NA (1996) Complications of iliac crest bone graft harvesting. Clin Orthop Relat Res 329:300–309. doi:10.1097/00003086-199608000-00037
5. Younger EM, Chapman MW (1989) Morbidity at bone graft donor sites. J Orthop Trauma 3(3):192–195. doi:10.1097/00005131-198909000-00002
6. Banwart JC, Asher MA, Hassanein RS (1995) Iliac crest bone graft harvest donor site morbidity. A statistical evaluation. Spine 20(9):1055–1060. doi:10.1097/00007632-199505000-00012
7. Vaccaro AR, Chiba K, Heller JG, Patel T, Thalgott JS, Truumees E, Fischgrund JS, Craig MR, Berta SC, Wang JC (2002) Bone grafting alternatives in spinal surgery. Spine J 2(3):206–215. doi:10.1016/S1529-9430(02)00180-8
8. Boden SD, Schimandle JH (1995) Biologic enhancement of spinal fusion. Spine 20(24 Suppl):113S–123S. doi:10.1097/00007632-199512151-00008
9. Weiner BK, Walker M (2003) Efficacy of autologous growth factors in lumbar intertransverse fusions. Spine 28(17):1968–1970. doi:10.1097/01.BRS.0000083141.02027.48 (discussion 1971)
10. Whang PG, Wang JC (2003) Bone graft substitutes for spinal fusion. Spine J 3(2):155–165. doi:10.1016/S1529-9430(02)00539-9
11. Urist MR (1965) Bone: formation by autoinduction. Science 150(3698):893–899. doi:10.1126/science.150.3698.893
12. Urist MR, Strates BS (1971) Bone morphogenetic protein. J Dent Res 50(6):1392–1406. doi:10.1177/00220345710500060601
13. Sampath TK, Coughlin JE, Whetstone RM, Banach D, Corbett C, Ridge RJ, Ozkaynak E, Oppermann H, Rueger DC (1990) Bovine osteogenic protein is composed of dimers of OP-1 and BMP-2A, two members of the transforming growth factor-beta superfamily. J Biol Chem 265(22):13198–13205
14. Boden SD, Kang J, Sandhu H, Heller JG (2002) 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 27(23):2662–2673. doi:10.1097/00007632-200212010-00005
15. Dimar JR, Glassman SD, Burkus KJ, Carreon LY (2006) Clinical outcomes and fusion success at 2 years of single-level instrumented posterolateral fusions with recombinant human bone morphogenetic protein-2/compression resistant matrix versus iliac crest bone graft. Spine 31(22):2534–2539. doi:10.1097/01.brs.0000240715.78657.81 (discussion 2540)
16. Burkus JK, Gornet MF, Dickman CA, Zdeblick TA (2002) Anterior lumbar interbody fusion using rhBMP-2 with tapered interbody cages. J Spinal Disord Tech 15(5):337–349. doi:10.1097/00024720-200210000-00001
17. Burkus JK, Sandhu HS, Gornet MF, Longley MC (2005) Use of rhBMP-2 in combination with structural cortical allografts: clinical and radiographic outcomes in anterior lumbar spinal surgery. J Bone Joint Surg Am 87(6):1205–1212. doi:10.2106/JBJS.D.02532
18. Boden SD, Zdeblick TA, Sandhu HS, Heim SE (2000) The use of rhBMP-2 in interbody fusion cages. Definitive evidence of osteoinduction in humans: a preliminary report. Spine 25(3):376–381. doi:10.1097/00007632-200002010-00020
19. Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry Low Back Pain Disability Questionnaire. Physiotherapy 66(8):271–273
20. Furlan AD, Pennick V, Bombardier C, van Tulder M (2009) 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine 34(18):1929–1941. doi:10.1097/BRS.0b013e3181b1c99f
21. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Control Clin Trials 7(3):177–188. doi:10.1016/0197-2456(86)90046-2
22. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327(7414):557–560. doi:10.1136/bmj.327.7414.557
23. Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. BMJ 315(7109):629–634. doi:10.1136/bmj.315.7109.629
24. Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. Biometrics 50(4):1088–1101. doi:10.2307/2533446
25. Egger M, Smith GD (1998) Bias in location and selection of studies. BMJ 316(7124):61–66. doi:10.1136/bmj.316.7124.61
26. Duval S, Tweedie R (2000) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in

meta-analysis. Biometrics 56(2):455–463. doi:10.1111/j.0006-341X.2000.00455.x

27. Altman DG, Bland JM (2003) Interaction revisited: the difference between two estimates. BMJ 326(7382):219. doi:10.1136/bmj.326.7382.219

28. Burkus JK, Dorchak JD, Sanders DL (2003) Radiographic assessment of interbody fusion using recombinant human bone morphogenetic protein type 2. Spine 28(4):372–377. doi:10.1097/01.BRS.0000048469.45035.B9

29. Burkus JK, Transfeldt EE, Kitchel SH, Watkins RG, Balderston RA (2002) Clinical and radiographic outcomes of anterior lumbar interbody fusion using recombinant human bone morphogenetic protein-2. Spine 27(21):2396–2408. doi:10.1097/00007632-200211010-00015

30. Dimar JR 2nd, Glassman SD, Burkus JK, Pryor PW, Hardacker JW, Carreon LY (2009) Two-year fusion and clinical outcomes in 224 patients treated with a single-level instrumented posterolateral fusion with iliac crest bone graft. Spine J 9(11):880–885. doi:10.1016/j.spinee.2009.03.013

31. Dimar JR 2nd, Glassman SD, Burkus JK, Pryor PW, Hardacker JW, Carreon LY (2009) Clinical and radiographic analysis of an optimized rhBMP-2 formulation as an autograft replacement in posterolateral lumbar spine arthrodesis. Journal Bone Joint Surg Am 91(6):1377–1386. doi:10.2106/JBJS.H.00200

32. Glassman SD, Dimar JR, Carreon LY, Campbell MJ, Puno RM, Johnson JR (2005) Initial fusion rates with recombinant human bone morphogenetic protein-2/compression resistant matrix and a hydroxyapatite and tricalcium phosphate/collagen carrier in posterolateral spinal fusion. Spine 30(15):1694–1698. doi:10.1097/01.brs.0000172157.39513.80

33. Dawson E, Bae HW, Burkus JK, Stambough JL, Glassman SD (2009) 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 91(7):1604–1613. doi:10.2106/JBJS.G.01157

34. Glassman SD, Carreon LY, Djurasovic M, Campbell MJ, Puno RM, Johnson JR, Dimar JR (2008) RhBMP-2 versus iliac crest bone graft for lumbar spine fusion: a randomized, controlled trial in patients over 60 years of age. Spine 33(26):2843–2849. doi:10.1097/BRS.0b013e318190705d

35. Haid RW Jr, Branch CL Jr, Alexander JT, Burkus JK (2004) Posterior lumbar interbody fusion using recombinant human bone morphogenetic protein type 2 with cylindrical interbody cages. Spine J 4(5):527–538. doi:10.1016/j.spinee.2004.03.025 (discussion 538–529)

36. Meyerding HW (1956) Spondylolisthesis; surgical fusion of lumbosacral portion of spinal column and interarticular facets: use of autogenous bone grafts for relief of disabling backache. Journal Int Coll Surg 26(5 Part 1):566–591

37. McKay B, Sandhu HS (2002) Use of recombinant human bone morphogenetic protein-2 in spinal fusion applications. Spine 27(16 Suppl 1):S66–S85. doi:10.1097/00007632-200208151-00014

38. Boden SD, Martin GJ Jr, Morone MA, Ugbo JL, Moskovitz PA (1999) Posterolateral lumbar intertransverse process spine arthrodesis with recombinant human bone morphogenetic protein 2/hydroxyapatite-tricalcium phosphate after laminectomy in the nonhuman primate. Spine 24(12):1179–1185. doi:10.1097/00007632-199906150-00002

39. Martin GJ Jr, Boden SD, Marone MA, Moskovitz PA (1999) Posterolateral intertransverse process spinal arthrodesis with rhBMP-2 in a nonhuman primate: important lessons learned regarding dose, carrier, and safety. J Spinal Disord 12(3):179–186. doi:10.1097/00002517-199906000-00001

40. Boden SD, Schimandle JH, Hutton WC (1995) 1995 Volvo Award in basic sciences. The use of an osteoinductive growth factor for lumbar spinal fusion. Part II: study of dose, carrier, and species. Spine 20(24):2633–2644. doi:10.1097/00007632-1995 12150-00004

41. Wozney JM (1992) The bone morphogenetic protein family and osteogenesis. Mol Reprod Dev 32(2):160–167. doi:10.1002/mrd.1080320212

42. Yoon ST, Park JS, Kim KS, Li J, Attallah-Wasif ES, Hutton WC, Boden SD (2004) ISSLS prize winner: lMP-1 upregulates intervertebral disc cell production of proteoglycans and BMPs in vitro and in vivo. Spine 29(23):2603–2611. doi:10.1097/01.brs.00001 46103.94600.85

43. Akamaru T, Suh D, Boden SD, Kim HS, Minamide A, Louis-Ugbo J (2003) Simple carrier matrix modifications can enhance delivery of recombinant human bone morphogenetic protein-2 for posterolateral spine fusion. Spine 28(5):429–434. doi:10.1097/01.BRS.0000048644.91330.14

44. Suh DY, Boden SD, Louis-Ugbo J, Mayr M, Murakami H, Kim HS, Minamide A, Hutton WC (2002) Delivery of recombinant human bone morphogenetic protein-2 using a compression-resistant matrix in posterolateral spine fusion in the rabbit and in the non-human primate. Spine 27(4):353–360. doi:10.1097/00007632-200202150-00006

45. Barnes B, Boden SD, Louis-Ugbo J, Tomak PR, Park JS, Park MS, Minamide A (2005) Lower dose of rhBMP-2 achieves spine fusion when combined with an osteoconductive bulking agent in non-human primates. Spine 30(10):1127–1133. doi:10.1097/01.brs.0000162623.48058.8c

46. Benglis D, Wang MY, Levi AD (2008) A comprehensive review of the safety profile of bone morphogenetic protein in spine surgery. Neurosurgery 62(5 Suppl 2):ONS423–ONS431. doi:10.1227/01.neu.0000326030.24220.d8 (discussion ONS431)

47. Wong DA, Kumar A, Jatana S, Ghiselli G, Wong K (2008) 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 8(6):1011–1018. doi:10.1016/j.spinee.2007.06.014

48. Chen NF, Smith ZA, Stiner E, Armin S, Sheikh H, Khoo LT (2010) Symptomatic ectopic bone formation after off-label use of recombinant human bone morphogenetic protein-2 in transforaminal lumbar interbody fusion. J Neurosurg Spine 12(1):40–46. doi:10.3171/2009.4.SPINE0876

49. Kanatani M, Sugimoto T, Kaji H, Kobayashi T, Nishiyama K, Fukase M, Kumegawa M, Chihara K (1995) Stimulatory effect of bone morphogenetic protein-2 on osteoclast-like cell formation and bone-resorbing activity. J Bone Miner res 10(11):1681–1690. doi:10.1002/jbmr.5650101110

50. Burkus JK, Gornet MF, Schuler TC, Kleeman TJ, Zdeblick TA (2009) 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 91(5):1181–1189. doi:10.2106/JBJS.G.01485

51. Vaidya R, Weir R, Sethi A, Meisterling S, Hakeos W, Wybo CD (2007) Interbody fusion with allograft and rhBMP-2 leads to consistent fusion but early subsidence. J Bone Joint Surg Br 89(3):342–345. doi:10.1302/0301-620X.89B3.18270

52. Lu S, Xu YQ, Chang S, Zhang YZ, Shi JH, Ding ZH, Li ZH, Zhong SZ (2009) Clinical anatomy study of autonomic nerve with respective to the anterior approach lumbar surgery. Surg Radiol Anat 31(6):425–430. doi:10.1007/s00276-009-0461-7

53. United States Food and Drug Administration, Department of Health and Human Services, C-enter for Devices and Radiological Health. (2011). InFUSE bone Graft/LT-CAGE? Lumbar tapered fusion Devices—P000058. Available at: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cftopic/pma/pma.cfm?num5P000058. Accessed 20 January 2011

54. Smoljanovic T, Siric F, Bojanic I (2010) 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 92(15):2614–2615 (author reply 2615–2616)

55. Carragee EJ, Mitsunaga KA, Hurwitz EL, Scuderi GJ (2011) Retrograde ejaculation after anterior lumbar interbody fusion using rhBMP-2: a cohort controlled study. Spine J 11(6):511–516. doi:10.1016/j.spinee.2011.02.013

56. Food and Drug Adminstration. (2011). Food and Drug Administration Executive Summary for P050036 Medtronic's AMPLIFY_ rhBMP-2 matrix orthopaedic and rehabilitation devices advisory panel. Available at: http://www.fda.gov/downloads/Advisory Committees/Committees MeetingMaterials/MedicalDevices/ MedicalDevicesAdvisoryCommittee/OrthopaedicandRehabilitati-on DevicesPanel/UCM220079.pdf. Accessed July 9, 2011

57. Luhmann SJ, Bridwell KH, Cheng I, Imamura T, Lenke LG, Schootman M (2005) Use of bone morphogenetic protein-2 for adult spinal deformity. Spine 30(17 Suppl):S110–S117. doi: 10.1097/01.brs.0000175184.27407.6a

58. Maeda T, Buchowski JM, Kim YJ, Mishiro T, Bridwell KH (2009) Long adult spinal deformity fusion to the sacrum using rhBMP-2 versus autogenous iliac crest bone graft. Spine 34(20):2205–2212. doi:10.1097/BRS.0b013e3181b0485c

59. Mulconrey DS, Bridwell KH, Flynn J, Cronen GA, Rose PS (2008) Bone morphogenetic protein (RhBMP-2) as a substitute for iliac crest bone graft in multilevel adult spinal deformity surgery: minimum two-year evaluation of fusion. Spine 33(20):2153–2159. doi:10.1097/BRS.0b013e31817bd91e

60. Glassman SD, Howard J, Dimar J, Sweet A, Wilson G, Carreon L (2011) Complications with recombinant human bone morphogenic protein-2 in posterolateral spine fusion: a consecutive series of 1037 cases. Spine 36(22):1849–1854. doi:10.1097/BRS. 0b013e3181d133d0

61. Carragee EJ, Hurwitz EL, Weiner BK (2011) A critical review of recombinant human bone morphogenetic protein-2 trials in spinal surgery: emerging safety concerns and lessons learned. Spine J 11(6):471–491. doi:10.1016/j.spinee.2011.04.023