

## COMMENTARIES

### Between a Rock and a Hard Place: What is the Evidence for Vertebroplasty and Kyphoplasty?

Philip Sambrook

*University of Sydney, Sydney, Australia*

**Commentary on:** Buchbinder R, Osborne RH, Ebeling PR, Wark JD, Mitchell P, Wriedt C, Graves S, Staples MP, Murphy B. A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. *N Engl J Med.* 2009 Aug 6;361(6):557-68.

Kallmes DF, Comstock BA, Heagerty PJ, Turner JA, Wilson DJ, Diamond TH, Edwards R, Gray LA, Stout L, Owen S, Hollingworth W, Ghdoke B, Annesley-Williams DJ, Ralston SH, Jarvik JG. A randomized trial of vertebroplasty for osteoporotic spinal fractures. *N Engl J Med.* 2009 Aug 6;361(6):569-79.

**The highest-level evidence in clinical medicine comes from double-blind, randomized controlled trials (RCTs). Previously, the evidence for the efficacy of vertebroplasty has been based upon open label and poorly controlled trials that may have overestimated the treatment effect by failing to take into account the natural history of vertebral fractures and the placebo response to an invasive treatment. Two recent, double-blind, RCTs with sham control of the effect of vertebroplasty in acute spinal fractures have cast doubt on the efficacy of vertebroplasty. In light of the high-level evidence from these studies, kyphoplasty must also now be regarded as a procedure in need of a true placebo-controlled trial to prove its efficacy.**

We are all exhorted to practice evidence-based medicine. The highest-level evidence comes from RCTs, however, with some therapies in clinical medicine we are often reliant on evidence of lower quality because of the difficulty of doing proper RCTs. Moreover, in the past, we have seen a number of examples of how when a proper RCT was conducted, it challenged our previous conceptions of efficacy about a particular mode of treatment, such as the role of arthroscopy in knee osteoarthritis or facet injections in spinal osteoarthritis (1;2).

Vertebroplasty for spinal fractures is one such therapy that needs re-evaluation based on two recent RCTs, which cast doubt on its efficacy (3;4). It is also worthwhile considering reasons for the apparent lack of efficacy of these two trials when compared with the positive effect reported in a recent RCT of kyphoplasty by Wardlaw *et al.* (5), a similar albeit not identical therapeutic approach.

Prior to these two RCTs, the best available evidence of the efficacy of vertebroplasty arose from small open label trials that compared vertebroplasty with conservative treatment (6-9). As Buchbinder and colleagues rightly point out in their RCT (3), the lack of blinding and the lack of a true sham control in these earlier trials raises the concern that the observed benefits in these trials reflected a placebo response, and the same caveat might reasonably apply to the recent trial of balloon kyphoplasty (5).

In the RCT from Kallmes *et al.* (4), both the vertebroplasty and sham groups had substantial improvement in back-related disability and pain in the first three days after the procedure. In the Buchbinder *et al.* study (3), there were significant reductions in overall pain (the primary endpoint) and secondary measures such as the Roland Morris Disability Questionnaire (RDQ) in both the vertebroplasty and sham groups at each follow-up assessment but no

significant differences between groups. In the balloon kyphoplasty study by Wardlaw *et al.* (5), the primary outcome was the difference in change from baseline at one month in the short form SF-36 physical component summary score. This score improved by 7.2 points from 26.0 at baseline to 33.4 at one month in the kyphoplasty group and by 2.0 points from 25.5 to 27.4 in the nonsurgical group, which was statistically significant.

In the two vertebroplasty RCTs, considerable attempts were made to simulate the active procedure in the placebo group. In the study from Kallmes *et al.*, the radiologist infiltrated the skin and subcutaneous tissues as well as the periosteum of the pedicles of the target vertebra with local anesthetic before randomization. Verbal and physical cues such as pressure on the spine and exposing the methacrylate monomer to simulate the odor associated with mixing the cement was also performed in the placebo group. Similarly, in the study from Buchbinder *et al.*, patients assigned to the sham intervention underwent the same initial procedures as those in the vertebroplasty group. This included insertion of the needle until it rested on the lamina. Then the central sharp stylet was replaced with a blunt one and the vertebral body was gently tapped. As in the study from Kallmes *et al.*, for the placebo group the cement was prepared but not injected so that its smell permeated the procedure room. These efforts to simulate the active procedure in the placebo were thorough, so much so that in an accompanying editorial (10), it was suggested that the use of local anesthetic in the Kallmes *et al.* and Buchbinder *et al.* studies may in part be considered an active treatment. In contrast, in the kyphoplasty study, the control group was recruited concurrently but in an open label design and was randomized to receive nonsurgical care that could include analgesics, physiotherapy, rehabilitation and back braces. This represents a much easier recruitment strategy and 300 patients were recruited for this study.

In contrast, with the more stringent entry criteria of the vertebroplasty RCTs, both studies struggled to recruit the originally planned sample size. In the Kallmes *et al.* study, 250 patients were planned but the target sample size was subsequently reduced to 130 patients after a planned interim analysis of the first 90 patients. Nevertheless, with the reduced sample size, the study still had power of more than 80% to detect a three point difference in the primary outcome measure, namely the RDQ. In the Buchbinder *et al.* study, the primary endpoint was the overall pain score at three months, which required a modest sample size of 48. However, a two-year study had been planned and it was calculated that a sample size of 164 would be needed for this longer endpoint. As in the Kallmes *et al.* study, the Buchbinder *et al.* study was terminated early before reaching that sample size when it became evident this recruitment target could not be achieved in a reasonable time.

The Kallmes *et al.* study was different in design from the Buchbinder *et al.* study in that it allowed for crossover between groups after one month or later after the vertebroplasty if adequate pain relief was not achieved, although the exact criteria for inadequate pain relief were not specified in formal terms. Both groups had improvement in disability and pain scores by 3 days and after one month, but there were no significant differences in any primary or secondary endpoints. However, by three months, 12% of subjects in the vertebroplasty group and 43% in the control group had crossed over to the other group and this difference was statistically significant. Interestingly, patients in the control group, who subsequently crossed over, had shown improvement by 3 days after the control procedure, but this improvement had dissipated by the one month assessment. Despite the thorough attempts to conceal randomization, at 14 days, 63% of patients in the control group and 51% of patients in the vertebroplasty group correctly guessed which group they had been randomized to in the Kallmes *et al.* study. However even after they underwent the alternative intervention, patients who

were originally assigned to either the vertebroplasty group or the control group did not have the same level of improvement at three months as did patients who did not cross over.

The patients recruited into the three studies were remarkably homogenous. The mean ages were similar (72-78 years), as was the sex distribution (73-82% female). They are also similar in terms of their baseline use of opioid drugs, baseline EQ-5D and baseline RDQ. There were, however, some differences. In the Kallmes *et al.* and Buchbinder *et al.* studies, patients needed to have sustained their fractures within 12 months of enrollment whereas in the Wardlaw *et al.* study, fractures more than three months old were excluded. This translated into a mean duration of back pain of around 16-20 weeks and 9 weeks respectively for Kallmes *et al.* and Buchbinder *et al.*, compared to a mean age of fracture of just under 6 weeks in the Wardlaw *et al.* study.

Could this difference in the age of the vertebral fracture have affected the outcomes? The Kallmes *et al.* study, by allowing crossover after one month, assumed that the biggest effect of active treatment would be seen by four weeks. Indeed, in their study, most of the change in primary and secondary endpoints occurred within the first two weeks after vertebroplasty, real or sham. That is also consistent with the Buchbinder *et al.* study, where most of the change in endpoints occurred by one month. In contrast, in the Wardlaw *et al.* study, improvements in endpoints continued for up to three months, although the biggest change was seen in the first four weeks. Thus it could be argued that one explanation for the discrepancy is that the Wardlaw *et al.* study treated much fresher fractures (mean 5.6 weeks vs. 9 or 20 weeks).

Another possibility is a mechanical one. Although using similar approaches, there are differences between vertebroplasty and kyphoplasty. With vertebroplasty, cement is injected to stabilize the vertebral fracture and in this way provides pain relief. For

kyphoplasty, the objective is not only to stabilize the vertebra but also to 'restore the anatomy' of the fractured vertebra. How this would necessarily translate into improved pain remains unclear. The Wardlaw *et al.* study enrolled patients with Genant grade 2 or more severity in 70% of cases and grade 3 or more severity in 29% of cases. The severity of fracture grade is not stated in the Kallmes *et al.* study but in the Buchbinder *et al.* study, the Genant grade appears less severe with 24% being grade 1 and 47% being grade 2.

What are the weaknesses of these two vertebroplasty RCTs? Because of crossover to the other group, the intention to treat analysis at three months may have underestimated the true treatment effect in the Kallmes *et al.* study (10), however, most of the benefit was seen within two weeks. Also, because more patients than predicted were able to guess which treatment they received, there might have been a benefit in understanding the treatment effect in those who guessed the treatment accurately (10). This data is not available for the Buchbinder *et al.* study.

What can we conclude from these three RCTs about injecting cement into spinal fractures? In terms of the prior evidence, it is recognized that uncontrolled or poorly controlled trials will tend to overestimate the treatment effect by failing to take into account the natural history of the underlying condition, the tendency for regression to the mean and the placebo response to treatment, which is likely to be amplified when the treatment is invasive. Raised expectations from invasive intervention may explain the effect of a sham procedure. In addition, the use of local anesthetic down to the periosteum may have enhanced the placebo effect. It remains possible that vertebroplasty or kyphoplasty provides some benefit in the first few weeks after fracture. However, this needs to be tested in a properly blinded trial by the radiologists who currently perform this procedure. Accordingly, it must be concluded that high-level evidence like the Buchbinder *et al.* and Kallmes *et al.* RCTs do not currently support any benefit of vertebroplasty. It is also

reasonable to conclude that kyphoplasty remains a procedure in need of a true placebo-controlled trial to prove its efficacy.

**Conflict of Interest:** None reported.

**Peer Review:** This article has been peer-reviewed.

## References

1. Carrette S, Marcoux S, Truchon R, Grondin C, Gagnon J, Allard Y, Latulippe M. A controlled trial of corticosteroid injections into facet joints for chronic low back pain. *N Engl J Med.* 1991 Oct 3;325(14):1002-7.
2. Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM, Wray NP. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med.* 2002 Jul 11;347(2):81-8.
3. Buchbinder R, Osborne RH, Ebeling PR, Wark JD, Mitchell P, Wriedt C, Graves S, Staples MP, Murphy B. A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. *N Engl J Med.* 2009 Aug 6;361(6):557-68.
4. Kallmes DF, Comstock BA, Heagerty PJ, Turner JA, Wilson DJ, Diamond TH, Edwards R, Gray LA, Stout L, Owen S, Hollingsworth W, Ghdoke B, Annesley-Williams DJ, Ralston SH, Jarvik JG. A randomized trial of vertebroplasty for osteoporotic spinal fractures. *N Engl J Med.* 2009 Aug 6;361(6):569-79.
5. Wardlaw D, Cummings SR, Van Meirhaeghe J, Bastian L, Tillman JB, Ranstam J, Eastell R, Shabe P, Talmadge K, Boonen S. Efficacy and safety of balloon kyphoplasty compared with non-surgical care for vertebral compression fracture (FREE): a randomised controlled trial. *Lancet.* 2009 Mar 21;373(9668):1016-24.
6. Diamond TH, Champion B, Clark WA. Management of acute osteoporotic vertebral fractures: a nonrandomized trial comparing percutaneous vertebroplasty with conservative therapy. *Am J Med.* 2003 Mar;114(4):257-65.
7. Alvarez L, Alcaraz M, Pérez-Higueras A, Granizo JJ, de Miguel I, Rossi RE, Quiñones D. Percutaneous vertebroplasty: functional improvement in patients with osteoporotic compression fractures. *Spine (Phila Pa 1976).* 2006 May 1;31(10):1113-8.
8. Diamond TH, Bryant C, Browne L, Clark WA. Clinical outcomes after acute osteoporotic vertebral fractures: a 2-year non-randomised trial comparing percutaneous vertebroplasty with conservative therapy. *Med J Aust.* 2006 Feb 6;184(3):113-7.
9. Voormolen MH, Mali WP, Lohle PN, Fransen H, Lampmann LE, van der Graaf Y, Juttman JR, Janssens X, Verhaar HJ. Percutaneous vertebroplasty compared with optimal pain medication treatment: short-term clinical outcome of patients with subacute or chronic painful osteoporotic vertebral compression fractures. The VERTOS study. *AJNR Am J Neuroradiol.* 2007 Mar;28(3):555-60.
10. Weinstein JN. Balancing science and informed choice in decisions about vertebroplasty. *N Engl J Med.* 2009 Aug 6;361(6):619-21.