Vertebral compression fractures

John E. O'Toole, M.D., and Vincent C. Traynelis, M.D.

Department of Neurosurgery, Rush University Medical Center, Chicago, Illinois

If we begin with certainties, we shall end in doubts; if we begin with doubts, and are patient, we shall end in certainties.

Francis Bacon, De Augmentis Scientarum (1605)

In our worship of certainty we must distinguish between the sound certainty and the sham, between what is gold and what is tinsel…

Benjamin N. Cardozo, The Growth of the Law (1924)

All that glitters is not gold.

Anonymous

Vertebral augmentation has come under intense scrutiny over the last year, following the publication of highly publicized trials and numerous editorials focused on the efficacy of percutaneous vertebroplasty (PV) and kyphoplasty (KP) for relief of pain due to vertebral compression fractures (VCFs).2,5,10–12,14,19,24–26,33,35,36 The lay press has concentrated almost solely21,32 on two randomized controlled trials (RCTs) published simultaneously in August 2009 in the New England Journal of Medicine (NEJM) that found no difference between PV and a sham anesthetic injection for pain relief in patients with osteoporotic VCFs.11,24 Despite this narrow emphasis and the negative impact it has had on the utilization of these procedures,27 there are, in fact, a number of other RCTs with high levels of evidence that address this very question with far more complex results than these two trials suggest.17,26,33,35

In this month’s issue, Farrhoki et al.17 add further data to the conflicted evidence base for PV and KP. The authors’ primary aim was to compare short- and long-term effects of PV on pain and quality of life (QOL), with secondary assessment of the incidence of new VCFs over 2 years as well as vertebral height and sagittal deformity correction. They did so by performing an RCT comparing PV to optimal medical therapy (OMT) in patients with osteoporotic VCFs. Inclusion criteria required confirmation of symptomatic VCF by physical examination findings of tenderness over the suspected site, in addition to positive x-ray and MR imaging findings. In this single-center study, 82 patients were randomized (40 to the PV and 42 to the OMT group) and followed up for 3 years. Statistically and clinically significant dominance for all primary and secondary outcome measures was found in favor of PV at all time points for at least 1 year and up to 3 years on some measures. The incidence of new VCFs was significantly higher in the OMT group than in the PV group. Limitations of the study include its single-center nature, the open-label design, unclear blinding of assessors, and the surprising extent of deformity correction documented post-PV. The authors should be acknowledged for a well-performed trial.

How Did We Get Here?

The state of the evidence base on vertebral augmentation prior to 2009 was perhaps best summarized by meta-analyses on PV and KP performed by Eck et al.16 in 2008 and on KP performed by Bouza et al.8 in 2006. Although the studies included in these analyses often provided low levels of evidence, an overall experience of more than 9000 patients with more than 15,000 fractures from dozens of different centers was summarized. Ultimately, both PV and KP demonstrated a consistent and immediate reduction in VAS pain scores of approximately 5 points. This well-recognized treatment effect is eerily consistent in the majority of RCTs that followed these reviews17,26,33,35 (and anecdotally identical in our own clinical practice as well).

In March 2009, the “FREE” study reported the results of an RCT comparing the effect of KP on osteoporotic and pathological VCFs versus nonoperative care.35 This multicenter, multinational, industry-sponsored trial enrolled 300 patients over 2 years, with more than 40% of eligible patients agreeing to participate. Inclusion was based primarily on MR imaging findings and clinical judgment, and the outcome measures included QOL, back function and mobility, back pain, narcotic analgesic use, and number of restricted activity days. Kyphoplasty proved superior to nonoperative care for all primary and secondary outcome measures between 6 and 12 months of follow-up. The groups appeared to have similar outcomes by the 12-month point. The VAS back pain scores improved immediately after KP by greater than 4 points. There was no difference in the overall frequency of adverse events or subsequent fractures between groups. Despite being the largest of all the RCTs performed, this study has been largely overlooked by both the lay press and those arguing about the RCTs for PV.
In 2009, the sham-controlled trials by Buchbinder et al.\textsuperscript{11} and Kallmes et al.\textsuperscript{24} produced an entirely different set of results than those seen previously (or since). Critics have rightly pointed out deficiencies that might threaten the validity of the studies.\textsuperscript{2,5,14,17,19,26,36} These concerns include the following: 1) difficulties in enrollment that resulted in a greater than 4-year enrollment period for 78\textsuperscript{11} and 131\textsuperscript{24} participants as well as a broadening of inclusion criteria and reduction in target study size after the study’s initiation; 2) the inclusion criteria did not use concordant physical examination findings, but relied primarily on radiographic assessment, which leads to the question of whether many of the treated fractures were actually asymptomatic; 3) the studies included both acute and chronic fractures; 4) a considerable number of patients on chronic narcotic pain medications were enrolled in both trials, possibly diluting any treatment effect; 5) the multicenter nature of the study by Buchbinder et al. has been challenged, since 68% of all procedures were performed at a single institution by one radiologist;\textsuperscript{11,14} 6) the procedures and evaluations were performed only by radiologists rather than clinically oriented spine specialists; 7) only unilateral approaches with small cement volumes were used during PV; 8) material used in the sham procedure in Kallmes et al.\textsuperscript{24} was not inert, and may have acted as a facet or periosteal block, with a therapeutic effect; 9) there was a high crossover ratio in Kallmes et al.\textsuperscript{24} (33% assigned to sham and 12% assigned to PV) that ultimately decreased the power to detect significant effect differences; and 10) most notably, neither the PV nor the sham groups enjoyed the magnitude of pain reduction seen in nearly all prior and subsequent PV/KP studies. Specifically, Buchbinder et al. found only an approximately 2-point VAS reduction in both groups at 1 week and at 1, 3, and 6 months, and Kallmes et al. found only an approximately 3-point VAS reduction in both groups at 3 days, 2 weeks, and 1 month. This overall diminution of a well-established treatment effect was never adequately explained by either set of authors. In fact, rather than reducing the effect, one would expect a placebo to have reproduced or even heightened the effect.

Farrokhni et al.\textsuperscript{17} point out the differences between their current study and those of Kallmes et al.\textsuperscript{24} and Buchbinder et al.\textsuperscript{11} Despite enrolling patients at just a single center, they were able to collect their desired study population in only 15 months, whereas the other studies required 4 years to enroll approximately the same number. Only 2% of eligible patients in Farrokhni et al. declined enrollment, versus 64%–70% in the sham-controlled trials.\textsuperscript{11,24} The inclusion criteria in Farrokhni et al. stipulated the need for physical examination evidence of tenderness over the putative fracture site concordant with x-ray and MR imaging findings (a routine practice pattern missing in the NEJM studies\textsuperscript{11-24}). The timing of enrollment after fracture onset (4 weeks to 1 year) was similar, but “painless” VCFs were excluded, thus avoiding patients with nonspecific chronic back pain that probably contaminat ed the sham-controlled trials’ populations and who may have actually had a treatment response to the anesthetic sham injection. Response to PV included an immediate 5-point reduction in VAS scores not seen in the control group or in either group in the sham-controlled studies.

In 2010, two additional RCTs comparing PV to nonoperative treatment have been published.\textsuperscript{26,33} Rousing et al.\textsuperscript{33} randomized 50 patients with acute osteoporotic VCFs at a single center in Denmark over the course of 7 years and found no difference in pain scores between groups at 3 and 12 months. However, the PV group showed a 6-point reduction in VAS within 24 hours of treatment, which was not seen in the control group. This difference was sustained at the 1-month follow-up point, but was lost after 3 months. This effect reduction may relate to the natural history of the disease, but may also signal lack of power in the study to detect differences over time, possibly exacerbated by the fact that the PV group had a statistically significantly lower baseline VAS score than the control group prior to enrollment.

Klazen et al.\textsuperscript{26} have published results of the Vertos II study examining PV versus nonoperative care in 202 randomized patients with acute osteoporotic VCFs at multiple centers over 3 years. Inclusion required concordance of physical examination findings and radiographic findings of acute VCF. This study provides an interesting view into the natural history of acute fractures; 50% of screened patients were ultimately deemed ineligible because their fracture pain resolved in the time period between initial screening and enrollment. The process of procuring participants in this study may, in fact, have screened out many of the patients with severe pain who ultimately “regress to the mean,” a phenomenon that some claim overestimates the treatment effects typically seen in PV/KP studies. The authors report a greater than 4-point immediate reduction in VAS scores in the PV group not seen in the controls. Clinically and statistically significant differences in pain levels were found in favor of PV over control for 12 months postenrollment.

How Do We Put the Evidence Base Into Practice?

Given that surgical RCTs are extremely difficult to perform, one may be willing to concede certain flaws in the internal validity of these studies. However, in generating practice recommendations one simply cannot ignore the serious threats to external validity. The process of extrapolating clinical research data to clinical practice relies on a critical evaluation of the generalizability of the results presented.\textsuperscript{6,18,22,30,31,34} Despite their status as the gold standard of clinical research, RCTs are often a form of “test tube” medicine. The restrictions and circumstances present in the study protocol and participating institutions rarely match the heterogeneity and characteristics of “real-life” practice.\textsuperscript{1,3,8,34} Generalizing results from RCTs requires a deep understanding of the tightly controlled conditions of the trial and the often unique characteristics of the patients enrolled. The results of single-center trials such as that of Farrokhni et al.\textsuperscript{17} (and in essence Buchbinder et al.\textsuperscript{11}) must be applied with caution because they may not be reproducible elsewhere.\textsuperscript{3,34} On the other hand, the high ratio of eligible to enrolled patients in Farrokhni et al. circumvents the “exclusion bias” seen in the NEJM studies\textsuperscript{11,24} that threatened their external validity and, in fact, makes the current study resonate more as “real-world.” Multi-
center trials with large patient enrollment mitigate these threats to external validity, but even in the best of circumstances, the characteristics of patients willing to enroll in RCTs are known to be different from those in the general population, often with respect to disease severity and treatment preferences. The practicing clinician, therefore, has to interpret the results of RCTs within the context of his or her own practice setting and population.

In the strictest interpretation, the generalization of Buchbinder et al. and Kallmes et al. would be that vertebroplasty is no better than sham surgery when all of the following criteria are met: 1) performed on VCFs up to 1 year postfracture; 2) patients have back pain that may or may not be related to a radiographically confirmed fracture and is not correlated with physical examination findings; 3) the selection and treatment of those patients is performed exclusively by interventional radiologists rather than spine specialists; and 4) the procedure is performed on the small subset of patients willing to participate in an RCT. To the extent to which these 4 conditions are contrary to the routine practice conditions of those caring for patients with VCFs, the practical application of inferences drawn from the studies becomes extremely limited.

From a design standpoint for a clinical trial, the routine use of nonblinded control groups without placebo in surgical trials is problematic. Biases on the part of both investigators and patients may alter the estimated effect of the intervention on outcomes, particularly when the outcome measure is something as subjective as pain. Treatment preferences on the part of patients, when unevenly distributed between groups, can lead to an overestimation of treatment effect. Similarly, the lack of strict methods in the control group can produce large intragroup differences that further “overpower” the study by weakening any treatment effects seen in the control group. In the present study, the details of the OMT as control strategy can be endlessly argued, and in this study included only medications, without specific use of physical therapy or bracing. However, data are lacking to support a specific OMT regimen that might have a decided effect on outcome after VCF. In fact, calcitonin (which was administered in the study performed by Farrokhi et al.) is one of the few nonoperative interventions with supportive evidence for use in acute fractures.

Blinding patients (if not providers and assessors) with placebo is a highly desirable way of avoiding bias in RCTs. Nevertheless, if we take the results of the sham-controlled (not necessarily placebo-controlled) PV trials at face value, then what do we offer our patients with painful VCFs? If both PV and sham interventions provided significant pain relief (albeit less than previously seen in most PV/KP studies), then one could logically conclude that patients with VCFs should be offered the sham procedure because the equivalent benefits lasted for up to 6 months (far longer than one would expect for a pure placebo effect). This absurd conclusion is an intrinsic fault of sham-controlled trials for surgical procedures.

So-called placebo-controlled surgical trials in which sham procedures are used would ideally include a concurrent nonsurgically treated group to differentiate the “true” placebo effect from both specific and nonspecific effects of the sham procedure. This is of particular concern when the sham procedure is not inert, or has a possible therapeutic effect. The case in Kallmes et al., in which the sham procedure included injection of local anesthetic around the facet and periosteum of the vertebral level of interest. In light of the concerns over whether the patients in these trials truly had fracture-related pain, this sham injection may have actually treated a subset of the patients who were suffering from degenerative facetogenic pain. The inability to distinguish true placebo effect from sham treatment–associated therapeutic effect can result in a Type II error by limiting the ability to detect a true difference between intervention and control. This problem is rarely accounted for in sample size calculations for sham trials, leaving many of them underpowered and prone to false-negative findings. A third, nonsurgically treated group can potentially provide the “control for the control” in these situations. Obviously, this increases the complexity of the trial and the demand for enrollment (although due to the secondary control, the total sample size need not be as much as 50% greater). Sluggish enrollment in Buchbinder et al. and Kallmes et al. could have been exacerbated further with the need for a third group. These considerations further highlight the questionable extrinsic value of studies in which 64%–70% of includable patients decline enrollment. Who precisely are we studying in these trials? Are they, in fact, the types of patients to whom many of us would not even offer surgery anyway? How can evidence-based practice decisions be made when the patient populations in RCTs do not reflect the populations we actually seek to treat?

The authors of the sham-controlled RCTs have flatly claimed that the populations in their studies are the same as those seen in clinical practice. However, as mentioned already, patients enrolling in RCTs are at best a selected subgroup and at worst a completely misrepresentative sample of the population at large with the disease of interest. These population differences often result in differences in how interventions perform when implemented in general practice than when subjected to rigorously controlled trials. Not infrequently, treatment effects are larger or more beneficial in practice than they are in the tight confines of an RCT. As an example, the nonsignificant trends toward a benefit of PV over sham treatment seen in Kallmes et al. might actually be a larger-magnitude effect in a “real-world” population.

The authors of the NEJM PV studies have sought to cut off debate on this entire topic by calling for an end to “…searching for spurious reasons to dismiss our results.” Legitimate concerns over the trustworthiness and generalizability of the data provided are not “spurious.” Moreover, absolutist statements by Buchbinder et al., such as “…it would be neither appropriate nor moral to offer this treatment in routine care,” have no place in the academic discourse regarding the application of RCT data to clinical practice. Such outright certainty in clinical medicine is rarely real, and these pronouncements simply serve to undermine the scientific process as well as misinform the public, payers, and others with a stake in the provision of health care here in the US and abroad.

The short-term results in most of these studies gener-
ate even more potential research questions regarding long-term consequences. Whether untreated VCFs lead chronically to progressive deformity, recurrent pain, and additional fractures remains, as yet, inadequately examined in the literature and awaits future study. What does seem to emerge from the clouds of conflicting data gathered so far is that there clearly are patients with acute to subacute fractures (or carefully selected nonhealing chronic fractures) for whom PV and KP can dramatically and rapidly reduce pain and improve function compared with routine nonoperative care. Most patients will arrive at a good level of function and low pain levels somewhere between 3 and 24 months regardless of treatment chosen, but this range remains uncomfortably large. There are significant physical, psychological, and economic costs associated with the immobilization and delayed healing that accompany the nonoperative care of VCFs. Furthermore, the amount of high-quality life “time” lost to pain, particularly for patients with metastatic cancer who have short life expectancies and pathological VCFs, makes justifying nonoperative care to patients difficult in daily practice. In the end, grounded in the practice-based evidence currently available, we believe the judicious use of vertebral augmentation by appropriately trained specialists is not only warranted but highly valuable. Studies like this one by Farrokhi et al.17 add further weight to this argument and should inspire more investigations into the appropriate selection of patients for this type of intervention.

References

Response

MARD REZA FARROKHI, M.D.

Shiraz Neurosciences Research Center, Shiraz, Iran

We greatly appreciate the careful, valuable scrutiny by Drs. O'Toole and Traynelis of vertebral augmentation in their analysis of practice-based evidence in the treatment of VCFs. Although the role of PV and KP for pain relief in acute and subacute VCFs has been covered well by many spine surgeons and interventionists, more well-designed RCTs are needed to document this treatment effect of PV. Performing an RCT of a surgical procedure is complex and problematic due to difficulties with medical ethics and local legal concerns; recruiting an identical control–sham procedure group; patient enrollment and verification; and appropriate concealment and blinding methods for patients, providers, and assessors. Despite these difficulties in planning and performing a surgical RCT according to Consolidated Standards for Reporting of Trials guidelines, we completed an RCT that compared PV with OMT in patients with osteoporotic VCFs. Our study compared short- and long-term effects of PV on the following outcomes: 1) primary outcomes including pain and QOL; and 2) secondary outcomes including the incidence of new adjacent fracture, restoration of vertebral height, and correction of the deformity.

We found a statistically significant improvement in pain in the PV group compared to the OMT group at all time points for 1 year up to 3 years. The patients’ QOL improved significantly in the PV group. We showed that PV can restore vertebral height and correct spinal deformity to some extent, and we also determined that the incidence of new VCFs was significantly higher in the OMT than in the PV group.

Although ours was a single-center study, the potential limitation this represents was partially offset by the fact that our hospital is the largest university referral center in southern Iran, which comprises 5 provinces with various cultures and traditions. All patients referred to us were randomly enrolled in the PV or OMT group at 15 months. A single neurosurgeon performed the PV in all 40 patients in the intervention group, and the management of treatment in all 42 patients in the OMT group was done by another specialist. As a result, the patients could be managed identically in each group, and we believe that this design prevented some of the bias that can occur in multicenter studies, in which treatment is provided by several specialists with different levels of experience and approaches to surgery.

We obtained informed consent from patients in the PV group according to local medical law, and each group was blinded to the treatment modality of the other group. Thus our trial was performed with a single-blind design. All providers and assessors were unaware of the study because their interactions with patients and staff formed part of their routine duties.

The sham-controlled trials by Buchbinder et al.3 and Kallmes et al.6 which were published in the NEJM in August 2009, found that relief of pain from VCF was similar in patients treated with PV and those treated with a sham procedure. We discuss these results in detail in our article. The shortcomings of these studies included difficulties with patient enrollment, treating patients with chronic fracture, the effect of the anesthetic agent used for painful vertebrae and facets, and the high patient crossover rate from the OMT to the PV group. These issues might compromise the validity of the two NEJM studies.1, 2, 4, 5, 7, 11 During 2010, several well-designed RCTs on PV have been reported.7–10 Those published by Klazen et al.7 (Versilos II study), Rousing et al.,8 and the “FREE” study9 reported that PV effectively reduced pain within 24 hours of treatment, and that the improvements were maintained for 3–12 months. Our trial documented a reduction in pain after PV that was maintained for 24 months.

The RCTs cited above provided high levels of evidence in support of the ability of PV to control pain for as long as 24 months. Furthermore, we agree with O’Toole and Traynelis that PV can reduce the high economic costs and the psychological issues associated with immobilization, delayed healing, and subsequent spinal deformity that accompany the nonsurgical treatment of VCFs.

We have learned valuable lessons from this RCT, and believe that our study and other well-designed RCTs have provided practice-based evidence that supports the benefits of PV. Finally, we again thank Drs. O’Toole and Traynelis for their acknowledgment of our trial. As they have noted, we recommend the judicious use of PV by appropriately trained specialists in well-selected patients, and are convinced that this treatment modality is not only warranted but highly effective.
References


Please include this information when citing this paper: published online March 4, 2011; DOI: 10.3171/2010.10.SPINE10622.