

Evidence-Based Recommendations for Spine Surgery

Alexander R. Vaccaro, MD, PhD; Charles G. Fisher, MD; Peter G. Whang, MD; Alpesh A. Patel, MD; Srinivas K. Prasad, MD; Peter D. Angevine, MD, MPH; Kishore Mulpari, MBBS, MHSc; Ken C. Thomas, MD, MHSc

Imaging strategies for low-back pain: systematic review and meta-analysis.

Chou R, Fu R, Carrino JA, et al. *Lancet* 2009;373:463–72.

STUDY SUMMARY

Lumbar spinal imaging is commonly utilized in the evaluation of low back pain. The rationale for imaging, either plain radiographs, computed tomography (CT), or magnetic resonance imaging (MRI), is primarily based on identifying anatomical sources of pain. Unfortunately, the correlation between findings on imaging and clinical symptoms can be limited. A number of studies have been performed to elucidate the value of spinal imaging in the setting of acute back pain. The definition of “value”, however, varies from study to study. Investigations have focused separately on diagnostic infor-

mation, treatment interventions, patient outcomes, or patient satisfaction. Additionally, the inclusion and exclusion criteria for these studies have not been uniform. Specifically the definition of “red flags” for serious disease (fevers, weight loss, neurological deficits, etc.) are subjective and, in some instances, not defined. Nonetheless the clinical question remains: is immediate routine lumbar spine imaging more effective than usual care without imaging in patients with low back pain and no suggestion of “red flags.” Chou et al. used methods of systematic review and meta-analysis to address this question.

METHODOLOGICAL REVIEW

The authors conducted a systematic review and meta-analysis with sound and reproducible methodology. The methods used to select articles were well described with clear inclusion and exclusion criteria. Relevant outcome measures formed part of the inclusion criteria. Randomized controlled trials that compared immediate lumbar imaging to routine care for low back pain patients without indication of serious underlying conditions were considered. Trials were included if their outcomes reported on pain, function, mental health, quality of life, patient satisfaction and overall patient reported improvement. Two reviewers independently assessed titles and abstract for study inclusion. Two reviewers abstracted data from included papers and ranked each paper as higher or lower quality according to accepted and pre-specified criteria. Two primary outcomes were identified: pain and function. Secondary outcomes included mental status, quality of life, patient satisfaction and overall improvement. Outcomes were categorized as short-term (<3 months), long-term (3–6 months) and extended (>1 year). Given that different trials used different scales for measurement of specific outcomes, outcome scores were standardized to allow for pooling. The authors defined clinically important differences in outcome measures a priori, diminishing the likelihood of bias. Point estimates and their confidence intervals were reported using a random effects model. Meta-regression was performed for pain and function using duration of pain, imaging technique and trial quality as independent variables.

Peter G. Whang, MD, Department of Orthopaedics and Rehabilitation, Yale University School of Medicine, New Haven, CT, USA

Alpesh A. Patel, MD, Departments of Orthopaedic Surgery and Neurosurgery, University of Utah, Salt Lake City, UT, USA

Alexander R. Vaccaro, MD, PhD, Srinivas K. Prasad, MD, Departments of Orthopaedic Surgery and Neurological Surgery, Thomas Jefferson University and The Rothman Institute, Philadelphia, PA, USA

Peter D. Angevine, MD, MPH, Department of Neurological Surgery, Columbia University College of Physicians and Surgeons, New York, NY, USA

Kishore Mulpari, MBBS, MHSc, British Columbia Childrens Hospital, Vancouver, BC, Canada

Ken C. Thomas, MD, MHSc, Division of Spine, Department of Orthopaedics, University of British Columbia, Vancouver, BC, Canada

Charles G. Fisher, MD, Departments of Orthopaedic Surgery and Neurological Surgery, Thomas Jefferson University and The Rothman Institute, Philadelphia, PA

Corresponding Author: Alexander R. Vaccaro, Department of Orthopaedic Surgery and the Rothman Institute, Thomas Jefferson University, 925 Chestnut Street, 5th Floor, Philadelphia, PA 19107, USA. Phone: (215) 955-3458; E-mail: alexvaccaro3@aol.com

CLINICAL INTERPRETATION

Six randomized controlled trials capturing 1804 patients were included; 5 of the 6 studies were evaluated as higher quality. The results of the meta-analysis demonstrate that immediate spinal imaging did not affect primary clinical outcomes at either short- or long-term durations. The authors' state that, based upon these results, "clinicians should refrain from routine, immediate lumbar imaging in patients with low-back pain and without features of a serious underlying condition." While this meta-analysis demonstrates the potential shortcomings of routine lumbar imaging in back pain, the limitations of the study make generalization of the results somewhat guarded.

The meta-analysis performed by Chou et al. is, like all studies of similar methodology, inherently limited by the quality of the primary sources. As the authors note, exclusion criteria of the included trials varied with little consistency based upon age, patient history, patient symptoms, or exam findings. Most notably, the presence of neurological symptoms (e.g. leg pain, weakness, numbness, etc.) was not consistently addressed across all studies. Additionally, the type of lumbar imaging (plain radiographs, CT, MRI) investigated was not consistent with four trials utilizing plain radiographs and two using CT and/or MRI. Lastly, as noted by the authors, the

reported outcomes provide only a limited perspective on the value of imaging. The diagnostic value of these studies as well as their effects on medical decision-making, changes in treatment, and patient satisfaction are not adequately addressed. The authors also state, without referenced evidence, that profit motive is a driving factor behind the use of lumbar spinal imaging suggests a potential for pre-existing bias against spinal imaging. Given the heterogeneity in patient populations, inclusion/exclusion criteria, investigational treatment (imaging modality), and limitations in measured outcomes, generalization of the results is limited.

RECOMMENDATION ON IMPACT TO CLINICAL PRACTICE

The fundamental question of whether "immediate routine lumbar spine imaging is more effective than usual care without imaging in patients with low back pain and no suggestion of red flags" is important not only from a clinical but societal perspective. The exhaustive number of factors that influence the answer to this question make research in this area extremely difficult. In consideration of this study, which probably represents the best available evidence to date, and clinical experience around costs, burdens and risks we would make a weak recommendation to not carry out immediate routine lumbar spine imaging in patients with low back pain and no "red flags".

Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. Four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts.

Weinstein JN, Lurie JD, Tosteson TD, et al. *J Bone Joint Surg Am* 2009;91:1295–304.

STUDY SUMMARY

The SPORT studies have provided high quality, prospective randomized and cohort data on the treatment of various spinal conditions. In 2007, Weinstein et al. described the short-term (min 2 year) clinical outcomes among a group of patients with degenerative spondylolisthesis.^{iv} The study design allowed for the 892 eligible patients to consent for randomization (304 patients) or observational (303 patients) cohort of patients undergoing either surgical versus nonsurgical treatment. 285 (32%) patients declined to participate, thus bringing the total to 558(66%) who could not be randomized.

The high rate of crossover among randomized patients, nearly 40%, compromised the intent-to-treat analysis. The as-treated analysis of both cohorts demonstrated statistically significant improvements in the operative versus the non-operative group at 2-year follow-up. Questions about the long-term efficacy of surgical treatment remained as deterioration of surgical results has been reported previously.^{v,vi} Late complications from surgical treatment as well as secondary operations may not be captured within the short term and may negatively influence late results. Longer-term results may allow for identification of patients who would be optimally treated with non-surgical management. To address this, Weinstein et al. reported on the same group of patients initially described in 2007.^{vii} The authors attempted to identify any differences in pain and functional outcome in patients with degenerative spondylolisthesis 4 years after they were either treated nonoperatively or with surgery.

METHODOLOGICAL REVIEW

The authors have performed a methodologically sound randomized controlled trial. Setting, patient population, interventions and outcomes were stated very clearly. The

Spine Patient Outcomes Research Trial (SPORT) was conducted at 13 medical centers with multidisciplinary spine practices in 11 states in United States strengthening generalizability. The SPORT included both a randomized cohort and a concurrent observational cohort of patients who declined randomization, reflecting patient preference. All patients who had neurogenic claudication or radicular leg pain with associated neurological signs, spinal stenosis seen on cross-sectional imaging, degenerative spondylolisthesis seen on standing lateral radiographs, symptoms that had persisted for at least twelve weeks, and physician confirmation that they were a surgical candidate were eligible for inclusion. Patients with adjacent levels of stenosis were eligible but those with spondylolysis and isthmic spondylolisthesis were not.

The protocol surgery consisted of a standard posterior decompressive laminectomy with or without bilateral single level fusion (autogenous iliac crest bone-grafting with or without posterior pedicle screw instrumentation). The non-operative protocol was “usual recommended care,” which at least includes active physical therapy, education and counseling with instructions regarding home exercise, and non-steroidal anti-inflammatory drugs if the patient could tolerate them.

The primary outcomes were the Short Form-36 (SF36), bodily pain and physical function scores, and the American Academy of Orthopaedic Surgeons MODEMS (Musculoskeletal Outcomes Data Evaluation and Management System) version of the Oswestry Disability Index measured at 6 weeks, 3 months, 6 months, and yearly up to 4 years.

Although the details of a power analysis and sample size calculation were not stated in this publication, it was done in their previous publication. One of the problems with any surgical trial is patient recruitment. The authors gave a good description of patient selection, randomization, crossover and follow-up. Only 34% of eligible patients agreed to randomization with the characteristics of the remaining prospective cohort and declining patients carefully analyzed and compared.

The authors should be commended for this essential methodological principle. As this was an effectiveness trial, they performed an intention to treat analysis which biases towards the null hypothesis. They also performed an “as treated analysis” (sensitivity analysis) because there was crossover of 54% and 33% from the randomized and observational cohorts, respectively reflecting the impact of patient preference on surgical trials.

CLINICAL INTERPRETATION

In the as treated analysis, combining the randomized and observational cohorts of patients with spinal stenosis secondary to degenerative spondylolisthesis, those managed surgically were found to have significantly greater improvement in scores for pain, function, satisfaction, and self-rated progress over 4 years compared with patients treated nonoperatively. The authors reported sustained improvement in outcomes among patients treated surgically in an as-treated analysis at both 3 and 4 year follow-up.

The high rate of crossover, which increased up to 54% in the non-operative randomized group, confounded the results with no significant differences noted in the intent-to-treat analysis. The as-treated analysis combining the randomized and observational cohorts that adjusted for potential confounders demonstrated that the clinically relevant advantages of surgery that had been previously reported at 2 years were maintained at 4 years, with treatment effects of 15.3 (95% confidence interval, 11 to 19.7) for bodily pain, 18.9 (95% confidence interval, 14.8 to 23) for physical function, and 214.3 (95% confidence interval, 217.5 to 211.1) for the Oswestry Disability Index. The authors reported a re-operation rate of 15% at 4 years among all surgically treated patients with a re-stenosis rate of 5%.

The study, like its predecessor, is most significantly limited by low recruitment and nonadherence to the randomized treatment arm with crossover reported to be 46% in the surgical group and 54% in the non-surgical group at 4 years. The study is somewhat limited by the heterogeneity of the treatments provided with no standardization of operative or non-operative care; however the tradeoff of providing broader generalization wins out. Despite these limitations, this study provides strong intermediate-term (3 to 4 year) evidence that surgical management for degenerative spondylolisthesis provides superior results to non-operative care and results consistent with previous studies.

RECOMMENDATION TO IMPACT ON CLINICAL PRACTICE

A weak recommendation can be made for the surgical treatment of degenerative spondylolisthesis in the setting of high quality literature such that a majority of surgeons and patients would choose surgery because of superior results but some would not based on personal preferences or clinical circumstances.

Posterior short-segment fixation with or without fusion for thoracolumbar burst fractures. A five to seven-year prospective randomized study.

Dai LY, Jiang LS, Jiang SD. *J Bone Joint Surg Am* 2009 May;91(5):1033–41.

STUDY SUMMARY

Thoracolumbar burst fractures represent 10–20% of all spinal fractures and are highly variable in their presentation. The appropriate management of these fractures is controversial and ranges from non-operative management to combined anterior-posterior procedures and is typically individualized according to a number of factors including fracture morphology, neurology, surgeon and patient preference, coincident injuries and more. Numerous thoracolumbar injury classification systems have been developed to systematize this heterogeneous group of fractures in an effort to facilitate communication and develop management recommendations. Sethi et al.^{viii} recently reviewed the evolution of thoracolumbar injury classification systems from inception through the most recent iteration, TLISS/TLICS.^x This evolution in classification systems has been fueled by progressive understanding of the importance of biomechanical stability, injury mechanism and neurological condition. Despite this, there remains significant variability in management recommendations for these protean injuries.

One of the principal determinations made in the evaluation of thoracolumbar burst fractures is whether operative intervention is warranted at all. While there is a wealth of literature exploring this question, this remains a central controversy within thoracolumbar injury management models. Although Thomas et al. embarked on a systematic review of outcomes in non-operative vs. operative management of patients with thoracolumbar fractures who are neurologically intact, they were unable to abstract meaningful conclusions, at least partially because of heterogeneous injury classification and management criteria.^{xi} The complex landscape of thoracolumbar injury management is ripe with opportunity for meaningful clinical research, but will depend heavily on the adoption of a more universal classification system.

Within this broad landscape, Dai et al. sought to address the important – if narrow – relevance of fusion in the setting of posterior short-segment fixation for thoracolumbar burst fractures Denis type-B between T11 and L2.^{xii} The authors invoked the Denis classification system^{xiii} and the Load Sharing Classification system proposed by McCormack in 1994.^{xiv} This investigation represents a prospective, randomized study comparing short-segment fixation with or without autologous ICBG

in patients in burst fractures with load-sharing scores ≤ 6 . Addressing available evidence suggesting that fusion better preserves kyphosis correction, and emerging short-term evidence that fusion may not be essential,^{?,??} this study presents valuable evidence examining the importance of fusion within this narrow clinical context.

METHODOLOGICAL REVIEW

The authors identified 91 possible patients for enrollment in this study from their Level 1 Trauma Center over a two-year period. Nine patients were excluded due to co-morbid conditions. Nine patients refused surgery. This left 73 patients between the ages of 18 and 60, fifty-six of which were male. Inclusion criteria stipulated that only burst fractures between T11 and L2 with isolated fracture of the superior endplate and a load sharing score of less than or equal to 6 be included.

Simple randomization was used to place 36 and 37 patients in the non-fusion and fusion groups respectively. The technique of posterior short segment fixation was described. Surgical treatment was similar in both groups with the exception of the harvest of autogenous iliac crest bone graft and its application to the posterolateral aspect of the spine in the fusion group. Twenty-five patients with neurologic deficits were equally divided between treatment groups. Only one patient underwent decompression in the form of a hemilaminectomy.

The postoperative management was the same between groups. Patients were kept in bed for 3 days following surgery. A brace was not prescribed. Follow-up was at 1, 3, 6, 12 months and then annually thereafter to a minimum of 5 years. Both functional and radiographic outcomes were evaluated by independent observers. Fusion status was evaluated using plain x-rays according to a pre-defined set of criteria. When fusion was in question computerized tomography was assessed.

Simple parametric and non-parametric statistics provide the basis for hypothesis testing. A priori power calculations are described for the primary outcomes loss of kyphosis correction and visual analog scale for back pain; although the origin of 3 degrees as a clinically important loss of kyphosis correction was not provided and this parameter is probably not valid or reliable. Other outcomes assessed in this study included neurologic improvement (as measured by Frankel grade and ASIA motor score) as well as visual analog scale for pain over the bone graft site, SF-36 quality of life measure and fusion status.

With respect the study's primary outcome measure, no significant differences were found in the loss of kyphotic correction or the VAS for back pain, with minimum 5 year follow-up. Although no significant difference was identified in long-term back pain VAS, 25 of 37

patients in the fusion group continued to report non-trivial donor-site pain even at last follow-up. Moreover, the authors report significantly more operative time and blood loss in the fusion group. No significant differences were reported in the peri and early post-operative course with respect to surgical timing, return to weight-bearing status or complication rates. No significant differences are reported in functional outcome between the two treatment groups. Notably, the authors report that no hardware malfunction was identified and that no hardware revision or removal was warranted throughout the study period. Finally, the authors report that all 37 patients in the fusion group had solid fusion as measured by X-Ray, with CT use as necessary. Based on their findings, they concluded that the addition of a fusion to short segment posterior fixation for non-comminuted thoracolumbar burst fractures does not confer added patient benefit, as measured by kyphotic deformity or VAS for back pain at minimum 5 year follow-up.

CLINICAL INTERPRETATION

The authors should be congratulated for completing this randomized controlled trial in which all consecutive patients were accounted for at enrollment and in which they had 100% follow-up at each time point over 5 to 7 years. A power calculation was performed which proved to be imperative given the null result.

While this study was well designed and executed, Dai et al. do not mention blinding of the clinical assessment. The radiographic analysis was performed by independent evaluators but not blinded to treatment group. Indeed, it would be interesting for the reader to know how many patients in the non-fusion group fused their fractures, radiographically.

The generalizability of this study's findings to all thoracolumbar bursts is somewhat limited. Patients with Denis type B burst fractures (superior endplate fractures) with a load sharing score of 6 or less were included. The more comminuted thoracolumbar burst fractures were excluded as the load sharing classification would suggest they are better treated anteriorly but this is controversial. Neurological function was not a critical consideration in this study, with only one patient in each group getting any neurological decompression (unilateral laminectomy in both). As the authors acknowledge, many of these fractures may have been managed non-operatively altogether but this is not the question being addressed in this study. For many clinicians, the presence of a neurological deficit with a thoracolumbar burst fracture suggests the need for a possible anterior or posterolateral decompression. The surgical construct under scrutiny was not studied in the face of decompression which limits the generalizability of this investigation to this scenario.

RECOMMENDATION ON IMPACT TO CLINICAL PRACTICE

The term "thoracolumbar injury" remains a widely variable entity with equally variable management options. This investigation introduces Level I evidence with relatively few limitations for a narrow segment within this universe, but raises interesting questions about the relevance of posterior lateral fusion in the management of thoracolumbar burst fractures. While this study was well designed and executed, the generalizability of their findings beyond Denis B type burst fractures is limited. Thus, we propose a weak recommendation not to include posterolateral fusion for the operative treatment of Denis Type B Thoracolumbar burst fractures.

Percutaneous Vertebroplasty Compared to Conservative Treatment in Patients With Painful Acute or Subacute Osteoporotic Vertebral Fractures: Three-Months Follow-up in a Clinical Randomized Study.

Rousing R, Andersen MO, Jespersen SM, et al. *Spine* 2009;34:1349–54.

STUDY SUMMARY

Osteoporotic vertebral body compression fractures (VCFs) are routinely treated with vertebral body augmentation techniques such as percutaneous vertebroplasty (PVP) and kyphoplasty. While these interventions

often confer significant clinical benefit, there is a growing body of literature seeking to establish the comparative effectiveness of vertebral augmentation within the broader spectrum of VCF management, most notably using medical management as the benchmark treatment. The VERTOS study represents the first prospective randomized controlled trial comparing PVP to medical management.^{xvii} In a study population of 34 patients who failed 6-weeks to 6-months of conservative treatment, Voormolen et al. revealed that PVP was associated with significantly greater pain reduction, decreased analgesic use and improvement in disability at the 1-day and 2week time points. This study was limited in that crossover was permitted at 2 weeks and 14 of 16 patients randomized to medical management elected to have PVP performed after 2 weeks; this impaired any further analysis and the study was prematurely aborted accordingly.

There have been 3 Level II studies comparing PVP with conservative management, all of which show a short-term benefit with PVP that diminishes within 12 months.^{xviii,xix,xx} Rousing et al. have conducted a prospective, randomized, controlled trial comparing PVP with conservative management for the treatment of acute (<2wks) and subacute (2–8wks) osteoporotic compression fractures.^{xxi} This represents only the second Level I study examining PVP versus conservative treatment.

METHODOLOGICAL REVIEW

The study was performed at a university hospital in Odense, Denmark from January 2001 to January 2008. 50 subjects were included in the study, 26 were allocated to PVP group and 24 to conservative treatment group. The criteria for inclusion were intractable pain due to either acute (40 patients) or subacute (10 patients) osteoporotic fractures, inability of the patient to perform activities of daily living due to the pain, and sufficient cognitive function to complete the study. Simple randomization with envelopes was used. The authors did not specify blocks or stratification by acute/subacute or the number of eligible patients (rejection log).

PVP was performed in the operating theatre and under local anaesthetics by orthopaedic surgeons specializing in spine surgery. Bone cement (PMMA) was injected under continuous fluoroscopy. Both groups were offered pain medication and physiotherapy if necessary until discharge. In addition, the patients in the conservative group were offered brace treatment.

A visual analogue scale (VAS) measuring self-reported pain was the primary outcome measure on which the sample size calculation was done. Secondary outcome measures for physical and mental outcome included the SF 36, Dallas Pain Questionnaire, EuroQol (EQ5D), and a modified mini-mental state examination (MMSE). Radiographic data was collected in an effort to characterize development of adjacent level fractures although this study was not powered to isolate this outcome.

The reduction in pain from the initial visit to the 3month follow-up was comparable in the 2 groups ($p = 0.33$) from approximately 8.0 to 2.0 visual analogue scale. The intragroup difference was significant ($P = 0.00$). While pain relief in the PVP group most often occurred within 12 to 24 hours after the procedure ($P = 0.00$), the other secondary parameters appeared similar between the two cohorts. They observed 2 adjacent fractures in the PVP group and none in the conservative group.

This study implies that the majority of patients with acute or subacute painful osteoporotic compression fractures in the spine will recover after a few months of conservative treatment and that further research is nec-

essary to characterize the risk of adjacent fractures. No major adverse events were observed in either group

CLINICAL INTERPRETATION

The authors should be congratulated for completing this randomized controlled trial in which they concluded there is no difference in pain in patients who received PVP or conservative treatment at a 3 month follow-up. This is a superiority trial so the power and sample size calculations were performed to find a difference. We cannot assume equivalence or non-inferiority because we would be committing a type 2 error; a large sample size is needed to establish equivalence or non-inferiority.

A few structural weaknesses in this study are concerning. The study had a long enrollment period so the authors need to explain why it took them 7 years to recruit 50 subjects. They also did not list the number of eligible subjects approached for recruitment. They should have maintained and reported a rejection log to track eligible patients that did not consent to see if they differed from the participating patients— this could have been part of Fig 1. There were only 36 of 50 patients who completed the baseline VAS which is the primary outcome, i.e. almost 30% of included subjects did not record baseline scores. Furthermore, there are too many p-values reported; as an option they may have only subjected their primary outcome measure to hypothesis testing or they could have used a corrective factor such as the Bonferroni correction. We recognize that this would not have changed their null result.

The main concern we have relates to their first key point: “PVP is comparable to conservative treatment for treating painful osteoporotic vertebral compression fractures” whereas it should read “immediate PVP is not superior to conservative treatment in the treatment of acute or subacute osteoporotic vertebral compression fractures at 3 month follow up”. The initial statement is not justified by their study. The most notable finding of this investigation is a statistically significant reduction in VAS pain score within 12–24 hrs of intervention in the population undergoing PVP, a potentially clinically relevant finding. Although this symptomatic benefit was preserved, the conservative treatment group also enjoyed gradual improvement so that by 3-months there was no statistically significant difference between treatment groups. Moreover, there was no significant difference in mental or physical outcome at 3 months. This paper has defined the natural history of osteoporotic fractures in that most patients feel better fairly quickly. At this point, many practitioners use conservative care as first-line treatment for osteoporotic compression fractures, reserving PVP for those patients who do not respond to conservative care. This study has not addressed the com-

parative effectiveness of PVP in those who fail this initial period which is arguably a more important question.

As patient involvement in clinical decision-making grows, it is imperative that physicians present a balanced perspective, drawing on experience and best available clinical evidence. The findings of Rousing et al. support the early use of PVP in the setting of acute or sub-acute osteoporotic VCFs if early pain control is a central priority or if conservative measures are not a viable option. Both groups enjoyed considerable improvement at 3 months without any significant difference in measured pain or functional outcome.

RhBMP-2 versus iliac crest bone graft for lumbar spine fusion in patients over 60 years of age. A cost-utility study.

Carreon LY, Glassman SD, Djurasovic M, et al. *Spine* 2009;34:238–43.

STUDY SUMMARY

Carreon et al. report the results of a cost-utility analysis (CUA) of rhBMP-2 compared to autologous iliac crest bone graft (ICBG) in patients over 60 years old based on a previously-published randomized clinical trial (RCT).^{xxii} Fifty patients were randomized to receive rhBMP-2 and 52 were treated with ICBG. All subjects underwent single or multilevel instrumented lumbar arthrodesis without interbody fusion. Patients were evaluated at regular intervals using multiple outcome measures including SF-6D utility scores and complications were recorded contemporaneously. Direct costs were calculated by considering the reimbursement amounts for all inpatient and outpatient events during the 2-year study period; however, indirect and non-health related costs were not considered. According to their study data, the authors constructed a decision-analysis tree with nodes for clinical improvement, complications, and reoperations.

The authors report that the mean total 2-year cost for care excluding complications and additional spinal treatments was \$34,235 for the ICBG group and \$36,530 for the rhBMP-2 cohort. Other mean costs for the entire study population were \$10,888 for a major complication, \$46,852 for a revision surgery to manage a pseudarthrosis, and \$5892 for additional spinal treatments. Rollback of the decision analysis tree yielded a total cost of rhBMP-2 of \$39,967 with a 0.11 mean improvement in SF-6D scores and a total cost of \$42,286 for ICBG with a mean improvement of 0.10. Based on this analysis, the authors calculated that the higher initial

RECOMMENDATION ON IMPACT TO CLINICAL PRACTICE

This is moderate quality or level 2 evidence that when put in the context of previous studies does not seem to have an impact on clinical practice. When taking this and previous evidence into consideration along with patient preference and clinical expertise there is a weak recommendation not to use PVP for acute VCFs as the results of conservative care at 3 months is favorable. This investigation suggests that PVP patients get better faster which could provide certain clinical advantages but study limitations prevent any recommendations regarding this outcome.

costs of rhBMP-2 were more than made up for by a decrease in the need for additional treatments by the individuals in this cohort and they concluded that the use of rhBMP-2 was both more effective and less expensive than ICBG.

While a number of studies have demonstrated that recombinant human bone morphogenetic proteins such as rhBMP-2 may represent an effective method for promoting spinal fusion, one major consideration that may ultimately determine whether these materials gain widespread acceptance as substitutes for autogenous bone in this era of limited health care resources is their significant expense. For this reason, several authors have performed economic analyses to determine the costs associated with these products. Ackerman et al. and Polly et al. compared the estimated expenditures of stand-alone anterior lumbar interbody arthrodesis procedures using either autograft procured from the iliac crest or rhBMP2.^{xxiii,xxiv} These investigations suggested that the higher charges arising from rhBMP-2 may be offset by savings incurred by fewer complications and more rapid functional recovery. These results were recently corroborated by a similar evaluation conducted in Europe.^{xxv} Unfortunately, these previous reports were based on largely theoretical models derived from peer-reviewed literature and expert opinion. In an effort to overcome these limitations, Carreon et al. employed actual cost data obtained from a prospective, randomized clinical trial in their cost-utility study quantifying the relative value of rhBMP-2 for instrumented posterolateral fusions.

METHODOLOGICAL REVIEW

Many of the key elements of a high-quality CUA are incorporated in this evaluation. The authors present their decision tree with the relative probabilities associated with each chance node. In addition, the costs and incremental utility for each branch are also listed. The probabilities, costs, and utilities were all determined from a single study which serves to minimize potential

sources of error. Health-state utilities were also determined using the validated SF-6D instrument. A 2-year follow-up period for the study is probably valid based on the assumption that symptomatic nonunions are identified within that time; however a longer follow-up would be desirable.

The model seems to be a reasonable representation of the intervention in terms of both its possible sequelae and outcomes. Nevertheless, one concern is that the baseline analysis indicates an overall improved outcome for the rhBMP-2 cohort (0.11) compared to the ICBG group (0.10) yet Figure 2 shows that the overall mean utility for the rhBMP-2 group was lower at each time point. A falloff of the mean rhBMP-2 utility at 2 years does not seem to support the claim that the incremental utility benefit for this group was greater compared to the ICBG cohort. This discrepancy raises concerns about the validity of the model itself that may compromise the overall study.

The investigators in this study used patient-level reimbursements in their model which represents the amounts paid to a health care provider by a government or private insurer. A more accurate reflection of the true economic cost, however, is the value of the resource utilization for providing that service. In general, Medicare reimbursement rates are employed to calculate the costs for physician services whereas hospital expenses are estimated by applying the ratio of costs-to-charges (RCC) method to the institutional reports prepared for the Centers for Medicare & Medicaid Services (CMS). Simply considering reimbursement data will tend to bias the results depending on the insurance status of the patients. In this study population, most patients were likely covered by Medicare but the reimbursement values do not exclude the hospital “markup” for these services.

Probably the most significant shortcoming of this study is that it does not include an assessment of the sensitivity of the model to changes in input parameters. As noted by the authors, the patients’ costs exhibited a substantial degree of variability; as an example, the costs of ICBG ranged from \$19,988 to \$139,583 and \$21,955 to \$109,027 for rhBMP-2. In these situations, mean values do not always give a comprehensive picture of the model. Due to sampling error, the “true” cost might actually be significantly higher or lower than the calculated mean value. Additional measures of dispersion such as the median values, interquartile ranges, and standard deviations would allow the reader to assess the validity of the mean values as summary point estimates. Even with randomization of a moderately large patient cohort, inequalities between the groups may occur by chance which is even more likely to occur with relatively rare events. For instance, the mean cost of readmission within 3 months of surgery for the ICBG group was

almost twice that of the rhBMP-2 group which was based on a total of 7 readmissions.^{xxvi} Because the estimated utility benefit of an intervention may vary significantly depending on the instrument used, extensive sensitivity testing comprising both one-way (i.e. assessing the impact of changing one variable at a time) and two-way (i.e. altering input values in pairs) analyses should be part of every CUA.^{xxvii}

A review of this model reveals significant sensitivity to the probability of improvement after additional treatments were administered to the subset of patients in the rhBMP-2 cohort who had no complications. All 7 patients in this group improved in the baseline model but if 2 of those individuals had not responded favorably the probability of improvement would have been similar to that observed for the corresponding ICBG subgroup such that the difference between the interventions in terms of both cost and outcome would have essentially disappeared. This finding suggests that the results may be very sensitive to small changes in the input parameters. Similar to subgroup analyses in clinical trials, sensitivity analyses should be selected a priori if possible in order to minimize the likelihood of bias.

Finally, it is important to point out that this investigation involves the use of rhBMP-2 for posterolateral lumbar fusions which is an indication that is not currently approved by the Food and Drug Administration. As with all industry-supported research it is also critical for the reader to recognize that one or more of the authors acknowledged a financial relationship that has the potential to create a conflict of interest.

CLINICAL INTERPRETATION

Cost-utility analysis is widely considered the gold standard design for comparative effectiveness studies because it allows for the comparison of the costs and effectiveness of different interventions for a single or multiple conditions. Because of their allure, high regard, and potential for significant impact from a clinical and reimbursement perspective it is critical that the results of these types of studies are not taken out of context and are appropriately reviewed. In this case, it would be inappropriate to conclude that spinal fusion is not cost-effective because its cost per quality adjusted life year (QALY) is roughly \$400,000 based on the findings of this investigation. The relevant metric is the incremental cost-effectiveness ratio (ICER) that compares the costs of two different treatments.^{xxviii}

RECOMMENDATION ON IMPACT TO CLINICAL PRACTICE

Although there are several significant methodological shortcomings in this study, this study should be viewed as an important first step in an iterative and evolving

process which will hopefully be refined in the future by the original authors or by other groups. Sensitivity analyses should be performed on the existing data to identify key areas requiring additional examination in an attempt to improve the estimates of cost, outcome, and probability which would be expected to yield more valid models and a better understanding of the relationship between BMP-2 as a re-

placement for iliac crest bone graft for active cost-effectiveness of various surgical options for fusion applications; thus, we do not currently recommend this patient population. The findings of this investigation do not justify the widespread incorporation of evidence provided by this study.

Does incorrect level needle localization during anterior cervical discectomy and fusion lead to accelerated disc degeneration?

Nassr A, Lee JY, Bashir RS, et al. *Spine* 2009;34:189–92.

STUDY SUMMARY

In a recent retrospective cohort study, Nassr et al. attempted to elucidate the risk of accelerated disk degeneration associated with the improper placement of a localizing needle during anterior cervical discectomy and fusion with plating (ACDFP) procedures by reviewing the preoperative, intraoperative, and postoperative cervical radiographs of 247 individuals who were managed with either 1- or 2-level operations. Sixteen patients were excluded because of non-degenerative pathology, a history of previous cervical spine interventions, inadequate radiographic studies, or insufficient follow-up; of the remaining 87 subjects, 15 (17%) were noted to have had the 22-gauge localizing needle inadvertently placed into the segment rostral to the intended fusion construct. Follow-up x-rays obtained at a mean of 21.8 to 24.6 months after surgery demonstrated that a greater proportion of patients with incorrect needle placement (9/15) exhibited radiographic evidence of progressive degeneration according to a 3-tier scale compared to those whose needles had been inserted into an appropriate disk space (23/72) which corresponded to an approximately 3-fold increased risk. There were no statistically significant differences between the incidences of progressive degeneration exhibited by subjects undergoing either 1- or 2-level procedures (28% and 43%, respectively).

The incidence and etiology of adjacent segment degeneration continue to be a matter of some debate, with many studies suggesting that it is a consequence of increased stresses arising from a contiguous arthrodesis whereas others have attributed this pathology to the natural history of cervical spondylosis. In a landmark paper, Hilibrand et al. retrospectively examined a large series of anterior cervical fusions and calculated the incidence of symptomatic adjacent segment disease to be 2.9% annu-

ally. Many surgeons elect to insert a needle within an intervertebral disk as a means of confirming the correct operative level during these procedures; however, with this method it is inevitable that the incorrect disk space will be marked incorrectly in some cases. Puncturing the annulus with a needle has been shown to trigger a number of biomechanical and biochemical alterations that collectively bring about progressive disk degeneration in animals. Elliott et al. examined the effect of needle diameter on this phenomena in an animal model and established that a needle:disk height ratio of less than 0.4 is unlikely to give rise to significant degeneration.

METHODOLOGICAL REVIEW

Although this is a retrospective review, the authors included a few design elements to minimize the potential biases to which these types of studies are susceptible. First of all, the raters of the pre- and postoperative radiographs were blinded to the patient's group (i.e. incorrect vs. correct needle placement). Furthermore, the authors calculated kappa values which represent a measure of interobserver agreement. Together these two factors serve to increase the validity of these findings.

The authors mention a few potentially confounding operative factors that may have also precipitated adjacent level degeneration such as the additional muscle and soft tissue dissection that occurred as part of the exposure of an adjacent disk space level. The authors also measured the distance between the upper edge of the anterior plate and the rostral disk space because any instrumentation in close proximity to uninvolved spinal segments has been shown to predispose these levels to progressive degeneration.

At first glance, the control and experimental groups in this study appear to be comparable; however, there were non-statistically significant differences between the mean ages, follow-up times and the proportion of each group that was male and female between these populations. It is also important to consider what other factors might have led to incorrect localization that also may have influenced the risk of developing degenerative disk disease. One such variable is the distribution of operative levels treated in each group. For example, there may have been a greater proportion of subjects with improper needle

placement who underwent surgery for C6–7 pathology in which case the puncture injury would have occurred at the C5–6 level which is known to be one of the most common segments to be affected by cervical spondylosis.² Similarly, body habitus is another potentially confounding characteristic that was not addressed by the authors.

The statistical analysis performed in this study also raises several questions. The sizes of the groups are relatively small which may affect the overall validity of the results. Although the odds ratio was reported to be 3.2, the 95 percent confidence interval is quite wide (1.02 to 10.05). It is impossible to determine from a single, small study whether the point estimate (i.e. 3.2) is the most accurate reflection of the true underlying value. Therefore, the authors' suggestion that "the results do reveal a significant increase in the radiographic evidence of disk degeneration with an odds ratio of 3.2 resulting from incorrect level needle localization" appears to overstate the strength of the evidence.

The finding that there were no differences in the likelihood of adjacent level disk degeneration between 1- and 2-level constructs is also limited by the study size. In this series 28% and 43% of 1- and 2-level fusion patients, respectively, demonstrated progression of degeneration of at least 1 grade which was associated with a P-value of 0.143. Unfortunately post-hoc power analysis is not possible because the number of individuals in each group is not provided but it is likely that a larger sample size would give rise to a statistically significant P value if these reported proportions were maintained. Given the insufficient power of this investigation, the authors' conclusions must be considered accordingly.

RECOMMENDATION ON IMPACT TO CLINICAL PRACTICE

The results of this investigation are quite striking in that incorrect needle localization had occurred in 17% of these cases. More importantly, these individuals were found to be approximately 3 times more likely to develop radiographic evidence of adjacent segment degeneration than those whose needles were introduced into an appropriate disk space. The authors attributed this finding to either a puncture injury to the annular tissue or unnecessary surgical dissection.

The small cohorts and retrospective design of the study significantly limit its ability to establish a causal link between improper needle placement and the radiographic degeneration. There is significant potential for confounding and bias in this study that may account for the observed difference in degeneration between the groups. The retrospective design limitations could have been minimized with an appropriate sample size as it is unlikely this research question will be addressed through a prospective design.

While it may still be prudent to develop alternative techniques for marking the cervical spine intraoperatively such as fixing a retractor pin in the vertebral bodies which may serve to minimize the development of secondary spondylotic changes at the levels that are not intended to be included into the fusion construct, the insufficient power of this investigation provides only low quality evidence to make any definitive conclusions about the effects of inadvertent disk injury. Based upon the results of this study along with the animal data and the insignificant clinical implications of marking the vertebral body, we offer a weak recommendation not to place a needle in the disk space to confirm levels during anterior cervical spine surgery.

REFERENCES

1. Boden SD, Davis DO, Dina TS, et al. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg Am* 1990;72:403–8.
2. Kerry S, Hiltons S, Dundas D, et al. Radiography for low back pain: a randomised controlled trial and observational study in primary care. *Br J Gen Pract* 2002;52:469–74.
3. Chou R, Fu R, Carrino JA, et al. Imaging strategies for low-back pain: systematic review and meta-analysis. *Lancet* 2009;373:463–72.
4. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus non-surgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med* 2007;356:2257–70.
5. Atlas SJ, Keller RB, Robson D, et al. Surgical and nonsurgical management of lumbar spinal stenosis: four-year outcomes from the maine lumbar spine study. *Spine* 2000;25:556–62.
6. Fischgrund JS, Mackay M, Herkowitz HN, et al. 1997 Volvo Award winner in clinical studies. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal fusion. *Spine* 1997;22:2807–12.
7. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *J Bone Joint Surg Am* 2009;91:1295–304.
8. Sethi MK, Schoenfeld AJ, Bono CM, et al. The evolution of thoracolumbar injury classification systems. *Spine J* 2009;9:780–8.
9. Boehler L. Die Technik der Knochenbruchbehandlung im Grieden und im Kriege. Vienna, Austria: Verlag von Wilhelm Maudrich, 1930. 9–11.
10. Vaccaro AR, Lehman RA Jr, Hurlbert RJ, et al. A new classification of thoracolumbar injuries: the importance of injury morphology, the integrity of the posterior ligamentous complex, and neurologic status. *Spine* 2005;30:2325–33.
11. Thomas KC, Bailey CS, Dvorak MF, et al. Comparison of operative and nonoperative treatment for thoracolumbar burst fractures in patients without neurological deficit: a systematic review. *J Neurosurg Spine* 2006;4: 351–8

12. Dai LY, Jiang LS, Jiang SD. Posterior short-segment fixation with or without fusion for thoracolumbar burst fractures. A five to seven-year prospective randomized study. *J Bone Joint Surg Am* 2009; 91:1033–41.
13. McCormack T, Karaikovic E, Gaines RW. The load sharing classification of spine fractures. *Spine* 1994;19:1741–4.
14. Denis F. The three column spine and its significance in the classification of acute thoracolumbar spinal injuries. *Spine* 1983;8: 817–31.
15. Sanderson PL, Fraser RD, Hall DJ, et al. Short segment fixation of thoracolumbar burst fractures without fusion. *Eur Spine J* 1999; 8:495–500.
16. Wang ST, Ma HL, Lu CL, et al. Is fusion necessary for surgically treated burst fractures of the thoracolumbar and lumbar spine? A prospective, randomized study. *Spine* 2006;31:2646–52.
17. Voormolen MH, Mali WP, Lohle PN, et al. 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;28:555–60.
18. Diamond TH, Champion B, Clark WA. Management of acute osteoporotic vertebral fractures: a non-randomised trial comparing percutaneous vertebroplasty with conservative therapy. *Am J Med* 2003;114:257–65.
19. Diamond TH, Bryant C, Browne L, et al. Clinical outcomes after acute osteoporotic vertebral fractures: a 2-year non-randomised trial comparing percutaneous vertebroplasty with conservative therapy. *Med J Aust* 2006;184:113–7.
20. Alvarez L, Alcaraz M, Perez-Higueras A, et al. Percutaneous vertebroplasty: functional improvements in patients with osteoporotic compression fractures. *Spine* 2006;31: 1113–8.
21. Rousing R, Andersen MO, Jespersen SM, et al. Percutaneous vertebroplasty compared to conservative treatment in patients with painful acute or subacute osteoporotic vertebral fractures. *Spine* 2009;34:1349–54.
22. Carreon LY, Glassman SD, Djurasovic M, et al. RhBMP-2 versus iliac crest bone graft for lumbar spine fusion in patients over 60 years of age. A cost-utility study. *Spine* 2009;34:238–43.
23. Ackerman SJ, Mafilios MS, Polly DW Jr. Economic evaluation of bone morphogenetic protein versus autogenous iliac crest bone graft in single-level anterior lumbar fusion. *Spine* 2002;27: S94–9.
24. Polly DW Jr, Ackerman SJ, Shaffrey CI, et al. A cost analysis of bone morphogenetic protein versus autogenous iliac crest bone graft in single-level anterior lumbar fusion. *Orthopedics* 2003; 26:1027–37.
25. Alt V, Chhabra A, Franke J, et al. An economic analysis of using rhBMP-2 for lumbar fusion in Germany, France and UK from a societal perspective. *Eur Spine J* 2009;18:800–6.
26. Glassman SD, Carreon LY, Campbell MJ, et al. The perioperative cost of Infuse bone graft in posterolateral lumbar spine fusion. *Spine J* 2008;8:443–8.
27. Marra CA, Marion SA, Guh DP, et al. Not all “quality-adjusted life years” are equal. *J Clin Epidemiol* 2007;60:616–24.
28. Torrance GW, Siegel JE, Luce BR: Framing and designing the cost effectiveness analysis, in Gold MR, Siegel JE, Russell LB, Weinstein MC (eds): *Cost-effectiveness in health and medicine*. New York, Oxford University Press, 1996, pp 54–81.
29. Nassr A, Lee JY, Bashir RS, et al. Does incorrect level needle localization during anterior cervical discectomy and fusion lead to accelerated disc degeneration? *Spine* 2009;34:189–92.
30. Hilibrand AS, Carlson GD, Palumbo MA, et al. Radiculopathy and myelopathy at segments adjacent to the site of a previous anterior cervical arthrodesis. *J Bone Joint Surg Am* 1999;81:519–28.
31. Masuda K, Aota Y, Meuhleman C, et al. A novel rabbit model of mild, reproducible disc degeneration by an annulus need puncture: correlation between the degree of disc injury and radiological and histological appearances of disc degeneration. *Spine* 2004;30:5–14.
32. Sobajima S, Kompel JF, Kim JS, et al. A slowly progressive and reproducible animal model of intervertebral disc degeneration characterized by MRI, x-ray, and histology. *Spine* 2004;30:15–24.
33. Kim SK, Yoon ST, Li J, et al. Disc degeneration in the rabbit: a biochemical and radiological comparison between four disc injury models. *Spine* 2004;30:33–7.
34. Elliott DM, Yerramalli CS, Beckstein JC, et al. The effect of relative needle diameter in puncture and sham injection animal models of degeneration. *Spine* 2008;33:588–96.