A review of lumbar spinal instrumentation: evidence and controversy

Maya A Babu,1 Jean-Valery C Coumans,2 Bob S Carter,3 William R Taylor,3 Ekkehard M Kasper,4 Ben Z Roitberg,5 William E Krauss,1 Clark C Chen4

ABSTRACT
Disability secondary to disorders of the spine is a significant problem worldwide. In the USA, there has been a recent surge in the costs associated with caring for spinal pathology; from 1997 to 2005, there was a growth of 65% in healthcare expenditures on spinal disease, totalling $86 billion in 2005. Increasingly, there has been media and public scrutiny over the rapid rise in the volume of procedures with spinal instrumentation; some have suggested that this rise has been fuelled by non-medical drivers such as the financial incentives involved with the use of instrumentation; others suggest that innovation in spine technology and devices has led to improved options for the treatment of spine pathology. In this context, we conducted a review of the literature to assess the use of instrumentation in lumbar procedures and its relationship to successful fusion and patient outcome. Our review suggests that there is data supporting the thesis that lumbar instrumentation improves rates of fusion. However, there is no consistent correlation between increased rates of fusion and improved patient outcomes.

INTRODUCTION
The leading cause of disability in the USA involves diseases of the musculoskeletal system. Of these disorders, those affecting the spine most commonly require medical or surgical care.1 Interestingly, healthcare expenditures related to disorders of the spine have grown 65% between 1997 and 2005, totalling $86 billion in 2005.2 In the USA, this rapid rise in healthcare expenditures has prompted media and Congressional scrutiny of the drivers behind this escalation. Specifically, much attention recently has been paid to the increased use of instrumented spinal fusion. Consider the following excerpt from the 11 September 2010 New York Times editorial on the topic of instrumented spinal instrumentation:

[There is a rapid rise in] the use of this surgery, which fuses multiple discs in the spine, in patients who would have done better, and faced fewer risks, with simpler surgery that eases pressure on the nerves without fusion... The explanation for the boom was likely economic. Surgeons were paid 10 times as much for the complex surgery, hospitals were paid three and a half times as much, and manufacturers reaped a bonanza selling $50 000 worth of implants for the complex surgery, compared with little or no profit from the simpler surgery.

Spine fusion is a surgical technique used to induce bone formation between adjacent vertebrae. In doing so, the adjacent vertebrae become fused and function as a single mechanical unit. This fusion is achieved by decorticating opposing bone surfaces and packing the gap between these surfaces with bone-graft material. Most frequently, the fusion is accompanied by the insertion of rigid implants that serve to minimise motion between the opposing bone surfaces. This practice is known as spinal instrumentation. The basic principle behind spinal instrumentation is built upon the observation that movement of opposing fractured bone surfaces impairs bone fusion and increases the likelihood of non-union. By connecting the implants in ways that augment the stability of the spine, it is thought that instrumentation functions as an ‘internal’ brace to facilitate bony fusion.

While instrumented spine fusion was initially developed for the treatment of conditions associated with a grossly unstable spine; including traumatic/pathological fractures and deformity correction, these indications now account for a small fraction of the procedures performed.3 The most common indication for spinal instrumentation is now degenerative spinal disorders without gross mechanical instability. This expansion in surgical indication has led to a significant rise in the frequency of spinal fusions. Over the period spanning 2002 to 2007, there was an approximate 15-fold increase in lumbar instrumentation in Medicare recipients.4

It is possible that the increased utilisation of spinal fusion reflects improved quality in care for an ageing population. Technological advances in spine instrumentation and imaging have certainly contributed to this increase.5–7 However, some have linked potential financial incentives to increased procedure utilisation. On average, a hospital bills more than $34 000 per instrumented fusion, excluding professional fees.8 The market for spinal implants is estimated to be $3.7 billion in 2008, with a projected annual growth rate of 18–20%.9–11

Adding to this controversy is a lack of general consensus in the neurosurgical and orthopaedic communities as to the indications for spinal fusion, particularly in the treatment of degenerative lumbar spine disorders. While guidelines have been published by various professional societies, these guidelines remain vague and subject to a wide range of interpretations. Importantly, despite significant regional variation in the frequency of spinal fusion performed in the USA, there is no evidence of differences in quality between regions with varying fusion frequency.12

1Department of Neurologic Surgery, The Mayo Clinic, Rochester, Minnesota, USA
2Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA
3Division of Neurosurgery, University of California, San Diego, California, USA
4Department of Neurosurgery, Beth Israel Deaconess Hospital, Harvard Medical School, Boston, Massachusetts, USA
5Section of Neurosurgery, The University of Chicago, Chicago, Illinois, USA

Correspondence to
Dr Clark C Chen, Department of Neurosurgery, Beth Israel Deaconess Hospital/Harvard Medical School, Dana-Farber Cancer Institute, Jimmy Fund 620A, 44 Binney Street, Boston, MA 02115-6084, USA; cchen1@partners.org

Received 12 October 2010
Revised 30 March 2011
Accepted 7 April 2011
Published Online First
20 May 2011
SPINAL INSTRUMENTATION: BACKGROUND AND RATIONALE

Spinal instrumentation was initially developed in the 1950s as a treatment for paediatric populations afflicted with polio and subsequent spinal deformity. The deformity was corrected by surgically destabilising the spine in a way that would allow for correction. The instrumentation was implanted to maintain the corrected alignment while facilitating fusion across the destabilised segments. The most famous and successful construct developed in this period was the Harrington Rod, a simple distraction device where rods are secured to the spine at two ends using hooks (figure 1A).15,16

The spectrum of instrumentation design has expanded greatly since the development of the Harrington Rod. Despite the multitude of instrument designs, instrumentation can be conceptualised within two categories: rigid and non-rigid. Rigid devices include segmental systems and interbody devices, such as cages. Non-rigid devices include prostheses such as artificial discs. Since non-rigid devices are generally designed for motion preservation and not used in conjunction with fusion, the remainder of the discussion will focus on rigid devices.

Rigid spinal implants are generally divided into two types: segmental devices and interbody devices. Segmental systems utilise instruments that are fixed onto each individual vertebra. The instruments are then connected in a rigid manner. The pedicle screw (figure 1B) is an example of this type. In this design, screws are placed into the vertebral bodies and connected by a rigid construct (a plate or a rod) thereby immobilising the spanning segments. Interbody devices utilise constructs to span a discontinuous bony surface. The intervertebral cage is an example of this design (figure 1C). The cage is frequently packed with autologous bone and inserted into fracture or discectomy sites to maintain spinal alignment. Each type of instrumentation has been used alone or in combination, depending on the instrumented level and anatomical considerations.

Instrumented spinal fusion in the lumbar region is typically performed when back pain is thought to be related to: (1) abnormal motion between lumbar vertebrae, (2) aberrant spinal alignment such as spondylolisthesis (anterior subluxation of one lumbar vertebra on another) or (3) painful motion segment which can be discogenic, facet mediated, or both. It is thought that elimination of these ‘pain generators’ by spinal fusion would afford symptomatic relief. However, the determination of spinal instability, aberrant spinal alignment, or painful disc/facet disruption is highly dependent on the evaluating surgeon. For instance, some surgeons will insist on radiographic demonstration of gross movement on flexion and extension, radiographic evidence of nerve root compression with corroborating neurological findings, and back pain before considering surgery. Others may require only a clinical history of back pain that worsens with physical motion. There is little or no consensus among practitioners in terms of surgical indication for instrumented spinal fusion.

EFFICACY OF SPINAL FUSION: A CENTRAL QUESTION

Fusion in the lumbar spine can be performed using a number of different techniques. However, if one accepts the premise that the purpose of rigid instrumentation is to facilitate the rate of bony fusion, the specifics of how the fusion is achieved (ie, what techniques were used to facilitate fusion) are then less pertinent than the question of whether successful fusion is correlated with improved clinical outcome. We examined this central question using the available data for lumbar instrumented postero-lateral fusion (defined below).

Postero-lateral fusion refers to the surgical technique whereby bone grafts are placed in between decorticated bony surfaces of the transverse processes (figure 1B) and, if feasible, the facet joints. Importantly, postero-lateral fusion can be achieved without spinal instrumentation. However, in recent years, postero-lateral fusion has been accompanied by segmental instrumentation (with screws placed through the pedicle and into the vertebral body and connected by rigid rods). The instrumentation is thought to immobilise the connected vertebrae and facilitate bony fusion. Analysis of randomised trials comparing non-instrumented and instrumented postero-lateral fusion, thus, affords one the opportunity to study whether spinal instrumentation affected clinical outcome in patients who essentially underwent comparable surgeries.

METHODS

We performed a search of the PubMed database and CENTRAL (The Cochrane Library 2008, Issue 2) for literature published from 1966 to 2010 using key words and MeSH headings including ‘lumbar fusion’, ‘outcome’, ‘surgery’, ‘simple’, ‘instrumentation’, and ‘complex.’ The search was restricted to the English language and yielded 318 references. The title and abstracts of each of these references were reviewed, and papers not providing class I/II data comparing non-instrumented and instrumented lumbar fusion were discarded. In the end, we identified seven well-designed randomised controlled trials (class I/II data).

RESULTS

Of the seven studies, four concluded that segmental instrumentation did not affect the likelihood of successful fusion or clinical outcome17–20 (table 1); three studies provided data that

Figure 1 Illustrations of spinal instrumentation. (A) Harrington Rod. Left: schematic of a Harrington Rod; middle: photograph of a Harrington Rod; right: x-ray (lateral view) of an implanted Harrington Rod. (B) Pedicle screws. Left top: schematic of a pedicle screw; left bottom: photograph of pedicle screws placed at L3, L4 and L5, connected with a rigid rod. The white rectangular region denotes where bone grafts are placed for postero-lateral fusions. (C) Intervertebral cage construct. Left: photograph of an intervertebral cage construct; right: x-ray (lateral view) of an implanted cage construct.
segemental instrumentation enhanced the rate of successful fusion. We reviewed the latter studies to determine whether the increased fusion rate corresponds to improved clinical outcome. Zdeblick21 randomised 124 patients with lower-back pain attributable to aberrant lumbar motion. Patients were randomised into groups that underwent postero-lateral onlay fusion or instrumented fusion. Patients with osteoporosis were moved from the instrumentation group to the non-instrumentation group owing to concerns that osteoporosis impairs pedicle screw fixation. Clinical outcome was assessed by patient report using a scale of excellent, good, fair and poor. Follow-up was 9–28 months. Fusion occurred in 65% of non-instrumented patients and 95% of the instrumented patients (p=0.002). Forty-nine per cent of the non-instrumented patients reported an excellent outcome, while 70% of the instrumented patients reported an excellent outcome.

A second study by Fritzell et al22 randomised 294 patients to postero-lateral onlay fusion versus different types of instrumentation supplemented with fusion, including pedicle screw. Ninety-eight per cent of the patients were followed for 2 years. The clinical outcome was assessed by the Oswestry Disability Index, General Function Score and Visual Analogue Scale (for pain assessment). The fusion rates for the on-lay fusion and pedicle screw group were 72% and 87%, respectively (p=0.004). There was no significant difference in any of the outcome scores between the groups, whether or not they fused.

A third study by Fischgrund et al23 randomised 76 symptomatic spondylolisthesis patients to postero-lateral onlay fusion or instrumented fusion. The clinical outcome was assessed using patient satisfaction surveys (graded as excellent, good and poor). The 2-year follow-up was obtained in 88% of the patients. Successful fusion was achieved in 82% of the instrumented patients and 45% of the non-instrumented patients (p=0.0015). However, the occurrence of successful fusion did not correlate with clinical outcome. Outcome was graded as excellent or good in 76% of the instrumented patients and 85% of the non-instrumented patients.

These studies suggest that the successful fusion rate for lumbar non-instrumented fusion ranged from 45% to 72%. Between 15% and 38% of the patients benefitted from instrumentation in terms of achieving fusion. However, the enhancement in bony fusion did not necessarily translate into improved clinical outcomes. Two of the three studies reviewed showed that an improved fusion rate related to spinal instrumentation did not correlate with clinical outcome. While the Zdeblick study demonstrated that instrumentation improved both fusion rate and clinical outcome, there is a noticeable discrepancy in the magnitude of these respective improvements. In this context, the data suggest a poor correlation between successful fusion and improved clinical outcome.

### DISCUSSION AND CONCLUSION

Extrapolating results from the various studies reviewed and offering one coherent clinical thesis is difficult. From a study design perspective, there was significant variation across the publications reviewed in terms of indications, inclusion and exclusion criteria, duration from the onset of symptoms to the time of surgery, methods of outcome assessment, and the level of supportive care provided to patients. For instance, only one study22 of the studies reviewed here examined the effect of instrumentation on low-back pain without laminectomy, a procedure that would further destabilise the spine. All other studies included variable number of patients who underwent laminectomy in addition to either instrumented or non-instrumented fusion. As another example, two of the studies22 23 included only patients who did not undergo prior surgery whereas other studies contained variable numbers of patients with previously failed back surgery.

From a surgical perspective, differences in surgical expertise, technique and instrumentation design confound any direct comparisons. From a patient selection point of view, differential society attitudes towards pain, differing practice patterns among surgeons as well as inherent genetic variations in the studied population may limit the generalisability of any particular study.

As a whole, however, the studies reviewed here suggest that there appears to be a ‘disconnect’ between spinal instrumentation and clinical outcome. There were four studies yielding class I b data suggesting that spinal instrumentation in the lumbar

<table>
<thead>
<tr>
<th>Author(s) (year)</th>
<th>Design</th>
<th>No of patients</th>
<th>Outcome assessment</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zdeblick (1993)21</td>
<td>Prospective randomised</td>
<td>124</td>
<td>Scale of excellent, good, fair and poor</td>
<td>— 40% increase in fusion with instrumentation (p=0.002) — 21% increase in rating as ‘excellent outcome’ with rigid instrumentation (p=0.00)</td>
</tr>
<tr>
<td>Fritzell (2002)22</td>
<td>Prospective randomised</td>
<td>294</td>
<td>Oswestry Disability Score, general function score, visual assessment score</td>
<td>— 15% increase in fusion with instrumentation (p=0.004) — No difference in outcome assessment</td>
</tr>
<tr>
<td>Fischgrund (1997)23</td>
<td>Prospective randomised</td>
<td>76</td>
<td>Scale of excellent, good and poor</td>
<td>— 37% increase in fusion with instrumentation (p=0.0015) — 9% increase in rating as ‘excellent or good outcome’ in non-instrumented patients relative to instrumented patients</td>
</tr>
</tbody>
</table>
spine does not enhance the rate of successful bony fusion. To the extent that the ultimate goal of rigid instrumentation is to facilitate fusion, this discrepancy raises questions as to the rationale for instrumented fusion. Of three studies yielding data suggesting that instrumentation facilitates bony fusion, the improved fusion rate correlated poorly with clinical outcome. These results underscore the complexity of nociceptive transmission and higher-order processing of neural transmission. Ultimately, successful surgical treatment of degenerative spinal disorders will require a deeper understanding of the biology underlying these processes beyond the current paradigms of aberrant spinal motion, alignment or discs.

Among spine practitioners, the adaptation of new technologies seems to outpace rigorous investigations of the associated clinical efficacy. The increased use of instrumented spinal fusion occurred well before outcome data were available. Similarly, as the outcome data for instrumented fusion have just begun to mature to a point where rigorous scrutiny is possible, some practitioners have already abandoned the rigid implants in favour of newer implants, including non-rigid instruments and artificial discs. If the trend of technical adaptation without convincing efficacy data continues, there is no doubt that the cost related to spinal instrumentation will continue to escalate and exacerbate the strain on the US healthcare system. On the other hand, an excessive regulation or focus on cost will facilitate fusion, this discrepancy raises questions as to the extent that the ultimate goal of rigid instrumentation is to facilitate fusion, this discrepancy raises questions as to the rationale for instrumented fusion. Of three studies yielding data suggesting that instrumentation facilitates bony fusion, the improved fusion rate correlated poorly with clinical outcome. These results underscore the complexity of nociceptive transmission and higher-order processing of neural transmission. Ultimately, successful surgical treatment of degenerative spinal disorders will require a deeper understanding of the biology underlying these processes beyond the current paradigms of aberrant spinal motion, alignment or discs.

Among spine practitioners, the adaptation of new technologies seems to outpace rigorous investigations of the associated clinical efficacy. The increased use of instrumented spinal fusion occurred well before outcome data were available. Similarly, as the outcome data for instrumented fusion have just begun to mature to a point where rigorous scrutiny is possible, some practitioners have already abandoned the rigid implants in favour of newer implants, including non-rigid instruments and artificial discs. If the trend of technical adaptation without convincing efficacy data continues, there is no doubt that the cost related to spinal instrumentation will continue to escalate and exacerbate the strain on the US healthcare system. On the other hand, an excessive regulation or focus on cost will facilitate fusion, this discrepancy raises questions as to the extent that the ultimate goal of rigid instrumentation is to facilitate fusion, this discrepancy raises questions as to the rationale for instrumented fusion. Of three studies yielding data suggesting that instrumentation facilitates bony fusion, the improved fusion rate correlated poorly with clinical outcome. These results underscore the complexity of nociceptive transmission and higher-order processing of neural transmission. Ultimately, successful surgical treatment of degenerative spinal disorders will require a deeper understanding of the biology underlying these processes beyond the current paradigms of aberrant spinal motion, alignment or discs.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

A review of lumbar spinal instrumentation: evidence and controversy


J Neurol Neurosurg Psychiatry 2011 82: 948-951 originally published online May 20, 2011
doi: 10.1136/jnnp.2010.231860

Updated information and services can be found at:
http://jnnp.bmj.com/content/82/9/948

Supplementary material can be found at:
http://jnnp.bmj.com/content/suppl/2011/08/22/jnnp.2010.231860.DC1.html

These include:

References
This article cites 18 articles, 0 of which you can access for free at:
http://jnnp.bmj.com/content/82/9/948#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

Editor's choice (105)
Musculoskeletal syndromes (487)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/
Adjacent Segment Disease After Anterior Cervical Discectomy and Fusion in a Large Series

**BACKGROUND:** Adjacent segment disease (ASD) development is known to occur after anterior cervical discectomy and fusion (ACDF).

**OBJECTIVE:** To study the relationship between index ACDF levels and the location of ASD development (above/below), as well as the effect of fusion length on ASD development.

**METHODS:** We report 888 patients who underwent ACDF for cervical spondylosis over a twenty-year period at a single institution. Of these patients, 108 had re-do surgery due to symptomatic ASD. Patients were followed for an average of 92.4 ± 52.6 months after the index ACDF.

**RESULTS:** In agreement with previous ACDF case series, we found the highest rates of cervical spinal degenerative disease requiring surgery at C5/C6, followed by C6/C7. Interestingly, neither the inherent location of index ACDF nor the length of instrumented arthrodesis appeared to correlate with the propensity to develop ASD. However, patients were more likely to develop ASD above the index level of fusion. This was true even for patients undergoing a second revision surgery due to recurrent ASD. Importantly, our data are consistent with existing in vitro biomechanical data in cadaveric spines.

**CONCLUSION:** We describe in detail the location and length of arthrodesis for index ACDFs, as well as first and second revision fusion surgeries in one of the largest Western cohorts in the literature. Our findings support the theory that iatrogenically introduced stress and instability at adjacent spinal segments contribute to the pathogenesis of ASD.

**KEY WORDS:** Adjacent level disease, Adjacent segment disease, Anterior cervical discectomy and fusion, Cervical, Fusion

---

A djacent segment disease (ASD) after anterior cervical discectomy and fusion (ACDF) is a well known but potentially debilitating complication following arthrodesis. Within the past few decades, both in vitro and in vivo biomechanical evidence suggest that the pathophysiology of ASD is iatrogenic in nature—specifically, that the destruction of vertebral joints via fusion of spinal segments during ACDF augments forces experienced by adjacent spinal levels.1–6 This subsequently exacerbates the stress levied upon the adjacent spinal level, worsening its degeneration, ultimately culminating in symptomatic ASD.7–18

In Hilibrand et al’s seminal study, 374 patients underwent ACDF; of these, there was an annualized incidence of 2.9% per year for developing ASD. Hilibrand et al estimated that more than a quarter of all patients would develop radiographic degeneration of spinal elements adjacent to the index surgery, and two-thirds of those would go on to require additional surgery.19,20 Indeed, the advent of total disc arthroplasty is predicated upon the assumption that restoring normal spinal kinematics would alleviate further spinal degeneration and development of ASD.3,15,21–35 Nonetheless, despite a number of studies looking at the effectiveness of total disc arthroplasty, to date there exists no Level I evidence that restoration of spinal biomechanics using disc replacements prevents or even abrogates the incidence of ASD.
development, highlighting the need for a better understanding of this pathophysiologic process. A review of the literature reveals few studies that specifically analyze the location and spinal construct lengths in patients who develop ASD.

In this original study, we present a single-institution series of patients who underwent ACDF, and report an incidence of 108 patients who required repeat cervical fusion due to symptomatic ASD after an initial ACDF. We compare the location and fusion construct lengths of patients undergoing index ACDF surgery, and patients requiring 1 or 2 revision surgeries due to onset of symptomatic ASD.

PATIENTS AND METHODS

Patients

In conducting the study, we retrospectively reviewed clinical notes, operative narratives, and radiology reports for all neurosurgical patients undergoing ACDF at our institution over a 20-year period. Our inclusion criteria were patients who had undergone ACDF exclusively for symptomatic cervical spondylosis and who subsequently also developed symptomatic ASD, defined as radiographic evidence of immediate above or below ASD as well as associated clinical symptoms in patients with a history of ACDF at least 6 months prior. Patients undergoing ACDF for neoplastic, infectious, or traumatic etiologies were excluded. In addition, patients receiving combined circumferential surgeries were excluded. Patients who underwent revision surgery due to wound infection, dehiscence, instrumentation failure or pseudoarthrosis were also excluded from this study.

Utilizing the aforementioned criteria, we identified 888 patients who underwent ACDF exclusively for symptomatic cervical spondylosis. Notably, 124 patients (13.9%) had a previous ACDF done at an outside institution. We identified 108 patients from the total cohort who underwent revision surgery for ASD at our institution.

Given that this is a retrospective study, we invariably lost some patients who moved away from the region or sought the services of other surgeons. When possible, we attempted to mitigate this factor through telephone calls to the patients to inquire about their functional status and surgical history since our last follow-up. Patients who had follow-up times of less than 6 months were not included in this study.

Variables and Statistical Methods

Our main outcome variables were the location and fused levels of the index ACDF, as well as levels involved during revision fusion surgeries. These variables were ascertained via operative notes. In order to account for fewer patients in the ASD cohort (108 patients) compared to the non-ASD cohort (780 patients), data are presented both as raw data and percentages. Distributions of vertebral levels were analyzed during index ACDF as well as during ASD revision surgeries. ASD development was noted for each affected spinal level, as well as sub-categorized as above or below the index ACDF surgery. Similarly, each vertebral level was re-evaluated for ASD re-development after first revision fusion surgery, and compared to ASD vertebral involvement after initial ACDF. The distributions of affected spinal levels were compared via Kaplan-Meier analyses and log-rank tests (cumulative distributions). Statistical tests were analyzed using GraphPad Prism 6.0. Statistical significance was defined as \( P < .05 \). Descriptive statistical results are presented as mean ± standard deviations when applicable.

RESULTS

A total of 888 patients underwent ACDF for degenerative spinal disease over the past 20 years at our institution. Of these, 401 (45.1%) were men and 487 (54.9%) women. Mean age at surgery was 50.8 ± 10.8 years. Mean follow-up time was 92.4 ± 52.6 months (Figure 1). The vast majority of patients underwent ACDF at the C5/C6 spinal level (Figure 2A). This was followed by the C6/C7, C4/C5, and C3/C4 levels in descending order of frequency. A total of 822 patients (92.6%) were fused with constrained plates and screws. Bone allograft was utilized in 163 cases (18.4%) and bone allograft (V2G and MTF bone spacers, DePuy, Johnson & Johnson, Raynham, Massachusetts) was utilized in 725 (81.6%) cases. Specifically, of the 780 patients who only received index ACDF surgery, the most common fused levels were C5/C6 in 565 (40.13%) occasions and C6/C7 in 423 (30.04%) cases. For patients who eventually developed ASD, the most common fused levels were also C5/C6 in 75 (40.1%) cases and C6/C7 in 54 (28.8%) cases (Figure 2B).

Of the 780 patients who only received index ACDF surgery, the majority of cases (325; 41.61%) involved a 1-level fusion, followed by a 2-level fusion in 311 (39.82%) cases (Figure 3A). For the 108 patients who developed ASD, most (51; 47.22%) also involved single-level fusions (Figure 3B). We analyzed the development of ASD over time as a function of levels fused via Kaplan-Meier analysis, but found no statistical significance between the 2 via log-rank analysis, \( P = .91 \) (Figure 4). Of the total 888 patients, 108 (12.2%) developed symptomatic ASD requiring revision surgery during an average follow-up time of 92.4 ± 52.6 months after the index ACDF. As patients were lost to follow-up over time, this may underestimate the true rate of ASD development. We attempted to correct for this via Kaplan-Meier analysis. Thus, accounting for time and censored patients, the rate of ASD development was 31.0% at 10 years (Figure 1).

FIGURE 1. Kaplan-Meier analysis of all patients undergoing ACDF developing ASD over time.
For the 108 patients who developed ASD, we sought to determine the location of their ASD relative to their initial surgery (Figure 5A). The most common levels fused were C5/C6 in 75 cases and C6/C7 in 54 cases. The symptomatic ASD that developed afflicted more commonly the C4/C5 levels in 35 cases and C3/C4 in 29 cases. Patients who developed ASD at the C1/C2 levels were treated with posterior fusions. Notably, 4 patients developed ASD at the C1/C2 level when only 1 patient had an index fusion at C2/C3. This occurred because the other 3 patients in fact developed 2-level ASD, which in these cases involved C1/C2. We sought to determine whether the location of index ACDF correlated with ASD development over time via Kaplan Meier analysis. Log-rank tests showed similar rates of ASD development over all surgical cohorts, \( P = .09 \) (Figure 6).

We also analyzed whether ASD developed at levels above or below relative to the original index ACDF (Figure 5B). 7 and 6 ASD spinal levels occurred above and below the index C3/C4 level, respectively. At C4/C5, 27 and 19 spinal levels developed ASD above and below the index ACDF site, respectively. 50 spinal levels developed ASD above C5/C6, whereas 27 developed below. 45 levels acquired ASD above the index C6/C7 ACDF level, in contrast to 4 spinal levels below C6/C7. Thus, the majority of ASD developed above the index level of ACDF.

For the 108 patients who developed adjacent segment disease after index ACDF, 27 patients developed re-occurrence of adjacent segment disease and required a second revision fusion surgery. Of the 81 who did not require a second revision surgery, most (29 patients; 37.18%) had a 3-level fusion (Figure 7A). Amongst the 27 patients who required a second revision surgery for re-development of adjacent segment disease, 12 (44.44%) had been fused at 2 levels (Figure 7B). We analyzed potential differences in ASD development location between patients who developed reoccurrence of ASD after first revision fusion surgery and those who did not.

**FIGURE 2.** Number of cases each spinal level was involved in. (A) Location of index ACDF levels between patients who did not experience ASD (black) and patients who did experience ASD (gray). (B) Location of index ACDF levels, expressed as percentages, between patients who did not experience ASD (black) and patients who did experience ASD (gray). These distributions were not significant based on Mann-Whitney U \( (P = .63) \) or 2-sample Kolmogorov-Smirnov \( (P = .99) \) tests.

**FIGURE 3.** Number of cases that involved single-level, 2-level, 3-level and 3-plus level ACDFs. (A) The number of vertebral levels fused in the index ACDF surgery for patients without ASD (black), and patients with ASD (gray). (B) The number of vertebral levels fused in the index ACDF surgery for patients without ASD (black), and patients with ASD (gray), expressed as percentages. The distributions were not significant based on Mann-Whitney U test \( (P = .37) \) or Kolmogorov-Smirnov test \( (P = .92) \).
Amongst the 81 patients who did not develop recurrence of ASD, most cases involved single or multi-level spinal fusions at C4/C5 and C5/C6 (Figure 8A, B). For the 27 patients who did develop ASD following 1 revision fusion surgery after index ACDF, most (24 cases; 32%) were located at the C5/C6 level. Interestingly, ASD redevelopment was again more likely to occur cephalad to the first revision fusion surgery levels (Figure 8C).

DISCUSSION

Anterior cervical disectomy and fusion is a commonly applied procedure for cervical spondylosis. However, arthrodesis of the cervical spine segments during ACDF irreversibly destroys the intervertebral disks, resulting in alterations to the physiological biomechanics of the spine. A number of biomechanical studies have shown that loss of mobility at a given spinal level increases the range of motion and intradiscal pressure experienced at both cephalad and caudal adjacent spinal segments. These studies have been correlated with clinical series documenting the development of degenerative changes in adjacent spinal levels following ACDF.

Key Results

Given that different intervertebral discs of the cervical spine experience different physiological ranges of motion and intradiscal pressures, we wondered if the level of index ACDF surgery influenced the potential development of ASD. Our data suggest that the specific spinal level is a predictor of likelihood of degenerative spinal disease development, but is not directly related to ASD. Thus, patients were at highest risk of degenerative cervical spinal disease requiring surgery at C5/C6, followed by C6/C7, C4/C5, and the C3/C4 spinal levels. This was true for all patients undergoing ACDF—regardless of whether they eventually developed ASD (Figure 2). This suggests that while the development of degenerative spinal disease is location-dependent, the advent of ASD is not directly due to the level of the index fusion itself.

As longer fusion constructs exhibit higher forces and more torque, we sought to analyze whether the length of instrumented arthrodesis correlated with ASD development (Figure 3). The majority of our patients either had 1- or 2-level index ACDFs. There was no statistical difference between the rate of ASD development in patients undergoing 1 level index ACDF vs 2 or 3+ ACDF via log-rank test, P = .91. Thus, our data suggest that the number of vertebral levels fused is not an immediate factor related to ASD development in vivo. We also compared the rate of ASD development over time as a function of the location of

(Figure 8). Amongst the 81 patients who did not develop recurrence of ASD, most cases involved single or multi-level spinal fusions at C4/C5 and C5/C6 (Figure 8A, B). For the 27 patients who did develop ASD following 1 revision fusion surgery after index ACDF, most (24 cases; 32%) were located at the C5/C6 level. Interestingly, ASD redevelopment was again more likely to occur cephalad to the first revision fusion surgery levels (Figure 8C).
index ACDF surgery. There was no statistical association between index ACDF fusion location and ASD development as assessed via log-rank test, \( P = .09 \).

We sought to determine the location of ASD development relative to the index ACDF (Figure 5A). Interestingly, for all levels of the cervical spine, patients were more likely to develop ASD above the index fusion level. While this was most pronounced for C6/C7, this trend held true at all cervical spinal levels (Figure 5B). Upon subsequent analysis of patients who underwent 2 revision surgeries for recurrence of ASD after 1 revision fusion surgery, this trend still remained true.

**Interpretation and Generalizability**

Our data are commensurate with existing biomechanical data. In Eck et al’s biomechanical analysis of vertebral levels adjacent to C5/C6 simulated ACDF via instrumented fusion in cadaveric cervical spines revealed increased intradiscal pressures and ranges of motion at both superior and inferior adjacent levels during both flexion and extension.\(^3\) However, closer analysis of the data reveals that the C4/C5 spinal level had higher intradiscal pressures relative to the C6/C7 level during flexion (C4/C5 intradiscal pressure increased by 73.2% vs 45.3% at C6/C7 during flexion). During extension, the C4/C5 range of motion increased 32.5%, more than 145% times the range of motion experienced by the C6/C7 level. Thus, both in flexion and extension, vertebral levels superior to the index ACDF experience higher intra-discal pressures and increased range of motion. These results are further corroborated by Park et al’s findings in cadaveric spines experiencing 2-level fusions from C5-C7.\(^6\) When intradiscal pressure changes were compared between pre- and post-ACDF cervical spines, only flexion produced statistically significant increased intradiscal pressures and range of motion anterior to the fused levels at C4/C5. Our study provides critical corroboration of these in vitro biomechanical findings, and highlights in vivo the increased risks of developing ASD anterior to the index ACDF compared to below the level of arthrodesis. Importantly, this increased risk held true even after 1 revision fusion surgery for ASD. Thus, for patients who are considering a second revision fusion surgery for recurrent ASD, this heightened risk must be kept in mind.

**Limitations**

As with all retrospective clinical studies, certain biases may play a confounding role. Because this study takes place over a long period of time, certain patients may have been lost eventually to follow-up. However, we were able to identify specific ASD rates based on follow-up times. Moreover, we tried to follow-up each patient via phone calls as well as clinic notes, and so were able to contact a subset of patients who moved away from the immediate area. With this method, we were able to...
achieve an average 92.4-month follow-up time. Thus, this study represents one of the largest Western cohorts of ASD after ACDF and provides one of the longest follow-up periods for these patients. Notably, we found a total of 108 patients (12.2%) who required revision surgery for ASD, which is relatively similar to Hilibrand et al.’s prediction of 16% of patients requiring revision surgery for ASD after ACDF. Prospective, randomized studies would help to not only clarify the contribution of index ACDF location to the development of ASD, but also broaden our understanding of ACDF as part of the treatment of degenerative spinal disease.

CONCLUSION

ACDF is the preferred surgical procedure for symptomatic degenerative cervical spinal disease and cervical spondylolisthesis. However, ASD is a common effect of this procedure and remains a poorly understood and highly morbid condition. In this manuscript, we describe in detail the location and length of arthrodesis for index ACDFs, as well as first and second revision fusion surgeries in one of the largest Western cohorts in the literature. In agreement with previous ACDF case series, we found the highest rates of cervical spinal degenerative disease requiring surgery at C5/C6, followed by C6/C7. Interestingly, neither the inherent location of index ACDF nor the length of instrumented arthrodesis affected the propensity to develop ASD. However, patients were statistically more likely to develop ASD above—compared to below—the index level of fusion. This was true even for patients undergoing a second revision surgery due to recurrent ASD. Importantly, our data are consistent with existing in vitro biomechanical data in cadaveric spines. These findings support the theory that iatrogenically introduced stress and instability at adjacent spinal segments—as opposed to endogenous degeneration of existing vertebral levels—contribute predominantly to the pathogenesis of ASD.

Disclosures

Daniel Sciubba is the recipient of a research grant from Depuy Spine. He has consulting relationships with Medtronic, Nuvasive, Globus, and Depuy. Timothy Witham is the recipient of a research grant from Eli Lilly and Company. Ziya Gokaslan is the recipient of research grants from Depuy Spine, AO Spine North America, Medtronic, NREF, Integra Life Sciences, and K2M. He receives fellowship support from AO Spine North America. He holds stock in Spinal Kinetics and US Spine. Ali Bydon is the recipient of a research grant from Depuy Spine. He serves on the clinical advisory board of MedImmune, LLC. The other authors have no conflict of interest to declare relating to the work in this manuscript.

REFERENCES


COMMENT

The authors describe 888 patients who had an index operation and followed up over a period of 20 years (an average follow-up of 92.4 ± 52.6 months), and found that 108 patients required a second operation. 27 of those 108 patients required a third operation and second revision. Adjacent segment disease after an anterior cervical discectomy and fusion certainly exists, but is still unclear if this is secondary to the natural history of the degenerative cervical spine, or whether it is secondary to increased stress causing the iatrogenic instability or degenerative disease at the adjacent spinal segments. The authors have documented nicely that approximately 12 percent of patients that had an anterior cervical discectomy and fusion will require repeat operation over a period of 20 years. These findings are valuable to spine surgeons advising patients and their families of their chances of needing repeat surgery after the index operation.

Volker K. H. Sonntag
Phoenix, Arizona

CME QUESTIONS:

1. What is the approximate annualized incidence of adjacent segment disease (ASD) after anterior cervical discectomy and fusion (ACDF)?
A. 1%
B. 2%
C. 3%
D. 4%
E. 5%

2. What is the most significant factor that contributes to the development of adjacent segment disease after anterior cervical discectomy and fusion?
A. Location of the index fused level
B. Length of the instrumented arthrodesis
C. Alteration of biomechanics above the index fusion level
D. Alteration of biomechanics below the index fusion level

3. What is the effect of cervical arthroplasty on adjacent segment disease (ASD) relative to anterior cervical fixation and fusion?
A. Reduction in short term ASD incidence but no change in incidence on long-term follow-up.
B. Increase in short term ASD incidence but no change in incidence on long-term follow-up.
C. No change in short term ASD incidence but reduction in incidence on long-term follow-up.
D. No change in short term ASD incidence but increase in incidence on long-term follow-up.
E. No impact on the incidence of ASD
Methodological reporting quality of randomized controlled trials in three spine journals from 2010 to 2012

Xiao Chen · Xiao Zhai · Xue Wang · Jiacan Su · Ming Li

Abstract

Purpose To elucidate the methodological reporting quality of randomized controlled trials (RCTs) in three spine journals from 2010 to 2012.

Methods In this study, we summarized the methodological report of RCTs in three major spine journals, including the Spine Journal, Spine and the European Spine Journal from 2010 to 2012. The methodological reporting quality, including the allocation sequence generation, allocation concealment, blinding and sample size calculation, was revealed. Number of patients, funding source, type of intervention and country were also retrieved from each trial. The methodological reporting quality was descriptively reported.

Results Ninety trials were involved and 57.8 % (52/90) reported adequate allocation sequence generation, 46.7 % (42/90) reported adequate allocation concealment, 34.4 % (31/90) reported adequate blinding and 37.8 % (34/90) reported adequate sample size calculation.

Conclusions This study shows that the methodological reporting quality of RCTs in the spine field needs further improvement.

Keywords Methodology · Quality · Randomized controlled trial · Evaluation

Introduction

Since the introduction of evidence-based medicine (EBM) a decade ago, the number of clinical practice guidelines published has increased dramatically. EBM has been defined by Sackett et al. [1] as “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients”. It integrates the best research evidence combined with clinical expertise and patient values.

The evidence quality of clinical studies was stratified by many systems for better use and according to one by the US Preventive Services Task Force [2]; the Level I evidence was obtained from at least one properly designed randomized controlled trials (RCTs).

Low-quality RCTs can be very misleading. Flaws in the randomization of RCTs can overestimate intervention benefits by 30 % and trials without proper blinding may exaggerate them by 14 % [3–5]. When they are included in systematic reviews and guidelines without cautious evaluation, conclusions will be severely compromised [4, 5]. Therefore, the best research evidence should be clinically
relevant and valid and it is extremely important to improve the quality of the RCTs.

CONSORT, which stands for Consolidated Standards of Reporting Trials, encompasses various initiatives developed by the CONSORT group to alleviate the problems arising from inadequate reporting of RCTs. The CONSORT statement is an evidence-based minimum set of recommendations for reporting RCTs. It comprises a 25-item checklist and a flow diagram, along with some brief descriptive text. The checklist items focus on reporting how the trial was designed, analyzed, and interpreted; the flow diagram displays the progress of all participants through the trial. The CONSORT statement offers a standard way for authors to prepare reports of trial participants through the trial. The CONSORT statement is an evidence-based minimum set of recommendations for reporting RCTs. It comprises a 25-item checklist and a flow diagram, along with some brief descriptive text. The checklist items focus on reporting how the trial was designed, analyzed, and interpreted; the flow diagram displays the progress of all participants through the trial. The CONSORT statement offers a standard way for authors to prepare reports of trial findings, facilitating their complete and transparent reporting and aiding their critical appraisal and interpretation [6].

Assessment of the reported methodology of RCTs in various research fields including gastroenterology [7], surgery [8] and oncology [9] has been studied previously. However, assessment of the methodological reporting quality of RCTs in spine journals has never been reported before. Therefore, the current methodological reporting quality of RCTs in major spine journals remains to be established. To address this issue, we systematically assessed the reporting methodology quality of RCTs published in three major spine journals from 2010 to 2012.

Methods

The current study included all RCTs published as full-text articles in the *European Spine Journal*, the *Spine Journal* and *Spine* in recent 3 years including 2010, 2011 and 2012. We studied RCTs in these three journals because they are the leading spine journals focusing on spine diseases and their reported methodological quality has never been systematically studied. *The Journal of Bone and Joint Surgery* both *American* and *British* volume were also taken into consideration. After search, only three (JBJS Am) and one (JBJS Br) RCTs related to spine diseases were identified. Due to the reason that most of the RCTs in these two journals were not related to spine diseases and adding four RCTs would not cause significant changes, after careful discussion and calculation, we decided not to incorporate these two journals in this current study.

The methods we used were similar to those described previously [10, 11]. In short, trials were considered to be RCTs if the words “random”, “randomly”, “randomization”, “randomized” were used in the text to describe the allocation method. However, trials published as abstracts, quasi-randomized trials, trials with animals or subgroups analysis of RCTs, trials being part of some large RCTs, observational studies nested within RCTs, and trials without the outcomes of randomized patients were excluded from the study.

The relevant trials were identified by the two co-first authors, who hand-searched all the issues of the three journals published in 2010, 2011 and 2012. The Pubmed database was searched with the strategy by Robinson and Dickersin [12] to include all potentially eligible trials. Methodological reporting quality was critically appraised by the following elements according to the Cochrane Handbook for Systematic Reviews of Interventions (updated March, 2011) [13]:

1. Adequate generation of the allocation sequence? “Yes” (e.g., random number table, computer random number generator), “unclear” (insufficient information to permit judgment of “yes” or “no”), or “no” (the description involves some systematic, nonrandom approach).
2. Adequate allocation concealment? “Adequate” (e.g., central allocation or sequentially numbered, opaque, sealed envelopes), “unclear” (insufficient information to permit judgment of “yes” or “no”), or “no” (e.g., open table of random numbers, alternation or rotation, date of birth).
3. Adequate blinding? “Yes” (e.g., no blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding; blinding of participants and key study personnel ensured and unlikely that the blinding could have been broken; either participants or some key study personnel were not blinded, but the outcome assessment was blinded, and the non-blinding of others is unlikely to introduce bias), “unclear” (insufficient information to permit judgment of “yes” or “no”), “no” (e.g., no blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by the lack of blinding; blinding of key study participants and personnel attempted, but likelihood that the blinding could have been broken; either participants or some key study personnel were not blinded, and the non-blinding of others is likely to introduce bias).

We also evaluated the sample size calculation, assessed as “yes” (types 1 and 2 error and estimated sample size reported), or “no” (not reported; type 1 or 2 error, estimated sample size, or both not specified). The agreement of the two authors was rated by calculation of kappa value. Any disagreement was resolved by discussion between the two reviewers. If the disagreement could not be resolved by discussion, then the opinion of the senior reviewer was sought. The primary aim of this study was to describe the current quality of the methodology.
reported in three major spine journals. We performed a Chi-square test for strata comparisons. Descriptive statistics (mean, standard deviation, median) were used. All statistical analyses were performed with Statistical Program for Social Sciences (SPSS) version 11.0 for Windows (SPSS Inc., Chicago, IL, USA).

Results

A total of 310 studies were retrieved from the three journals. Of these 310 studies, 220 were excluded because they had a nonrandomized design, letters or reviews, subgroup analyses, a pooled analysis of RCTs or cost-effective studies alongside RCTs. Finally, 90 trials were suitable for the analysis including 9 from the Spine Journal, 25 from the European Spine Journal and 56 from Spine.

Major characteristics of the included trials are summarized in Table 1. Generally, the 90 trials included a median of 166 patients (range 20–1,758; 25th percentile 60, 75th percentile 105). In 75 trials with specified primary outcomes, there were more trials with positive results published than trials with negative results (64 vs. 11). More than half of all the trials were reported from Europe (57 %, 51/90), with Asia contributing 20 trials, and the United States reporting 17 % (15/90) of all the trials (Table 1).

Kappa values for the interobserver agreement between the two reviewers were 0.90 for the generation of the allocation sequence, 0.91 for the allocation concealment, 0.88 for the double blinding, and 0.91 for the sample size calculation. All these values indicated almost perfect or substantial agreement.

Among the 90 trials, the generation of the allocation sequence was adequate in 52 (57.8 %) trials and unclear in the remaining 38 trials. The allocation concealment was adequate in 42 (46.7 %) trials and unclear in 48 trials. Adequate blinding was reported in 31 (34.4 %) trials and no blinding in 59 trials. Adequate sample size calculation was reported for 34 (37.8 %) trials. A total of 16 (17.8 %) studies were adequately performed according to our statistics.

According to different strata, it was found generally that large-scale (n > 100), multi-center studies had better quality of reported methodology for RCTs than small-scale (n ≤ 100), single-center studies in allocation concealment and sample size calculation. The quality of the reported methodology is summarized in Table 2.

Discussion

In the current study, we described the current methodological reporting quality in three major spine journals. It was noted that 42.2 % of all RCTs did not report adequate generation of the allocation sequence, 53.3 % did not report adequate allocation concealment, 65.6 % did not have adequate blinding, and 62.2 % did not report adequate sample size calculation. These findings suggest that much effort still be needed to improve the quality of the reported methodology for RCTs in major spine journals.

Through the efforts of Chalmers [14] and the EBM movement in general, it is common for new interventions to be evaluated in an RCT context. For some thorny issues in the field of spine diseases, such as low back pain and cervical degenerative diseases, surgeons are beginning to evaluate the treatment options with the help of EBM. Therefore, recent years has witnessed an increasing number of RCTs in the field of spine diseases. However, it has been suggested that the design has a more limited role in assessing surgery than for drug interventions in RCTs [15]. Reviews of urological and orthopedic trauma trials have revealed poor compliance with the consort statement for reporting RCTs and reflected generally poorer trial methodology [16, 17]. A lack of understanding about RCTs in surgical communities and the need for better epidemiological and statistical training of surgeons have been noted [18]. Compared to the drug interventions, due to the nature of surgery and exercise in spinal diseases treatment, randomized trials with patients undergoing such treatment as surgery or rehabilitation are extremely difficult to design and perform [15, 19]. Many uncontrollable confounding factors possibly affecting the final outcomes will increase substantial risk of the introduction of bias. Furthermore,
because it is difficult to randomize patients into treatment and control groups for a comparison of their outcomes, studies of surgical treatment are often criticized for lack of an appropriate control arm. The number of RCTs in the field of spinal diseases, although increased in the past decade, still is relatively small. It is of vital importance to appraise the quality of methodology of these randomized studies critically. Because if the conclusions of these studies are accepted without close examination, the reliability of future guidelines, based on the results of these randomized trials, could be severely compromised, as demonstrated in the field of cardiology [20].

In this study, we evaluated a randomized trial’s practical usefulness with 2011 Cochrane Handbook for Systematic Reviews of Interventions [13]. Sequence generation, allocation sequence concealment, blinding and sample size calculation were taken into consideration.

Randomization allows for the sequence to be unpredictable. According to Kenneth et al. [21], randomization eliminates bias in treatment assignment, facilitates blinding (masking) of the identity of treatments from investigators, participants, and assessors, including the possible use of a placebo and permits the use of probability theory to express the likelihood that any difference in outcome between treatment groups merely indicates chance. The process of randomization begins with this sequence generation process. In this study, a research considered with adequate sequence generation should clearly report the exact sequence generation methods, such as using a random number table or a computer random number generator. Sequence generation based on systematic methods, such as alternation, the date of admission, or with only such simple statements as ‘we randomly allocated’ or ‘using a randomized design’, was regarded inadequate in this study. Nearly 42.2 % of RCTs did not report adequate sequence generation, indicating possibly inadequate randomization process. It may be very misleading and introduce significant biases, which should be avoided in the future study.

Randomized sequence generation is not a sufficient safeguard against bias in intervention allocation. Efforts made to generate unpredictable and unbiased sequences are likely to be ineffective when those sequences are open to those involved in the enrolment and assignment of participants. Allocation concealment has been considered to be more important than other components of allocation in reducing bias [22]. Thus, the allocation concealment reporting can be further improved. Although it is not always feasible to conduct a double-blinded trial (especially in the surgical field), according to the Cochrane Handbook 2011, it is always possible to adequately conceal the allocation sequence by remote telephone, fax or email service, a central independent unit, or sequentially numbered, opaque and sealed envelopes. Our study found that 46.7 % of trials reported adequate allocation concealment.

<table>
<thead>
<tr>
<th>Year</th>
<th>Adequate allocation sequence generation n (%)</th>
<th>Adequate allocation concealment n (%)</th>
<th>Adequate blinding n (%)</th>
<th>Adequate sample size calculation n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010 (n = 28)</td>
<td>11 (39)</td>
<td>10 (36)</td>
<td>9 (32)</td>
<td>11 (39)</td>
</tr>
<tr>
<td>2011 (n = 38)</td>
<td>23 (61)</td>
<td>20 (53)</td>
<td>16 (42)</td>
<td>12 (32)</td>
</tr>
<tr>
<td>2012 (n = 24)</td>
<td>18 (75)</td>
<td>12 (50)</td>
<td>6 (25)</td>
<td>11 (46)</td>
</tr>
<tr>
<td>P value</td>
<td>0.031</td>
<td>0.368</td>
<td>0.368</td>
<td>0.667</td>
</tr>
<tr>
<td>Journal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Spine Journal (n = 9)</td>
<td>6 (67)</td>
<td>5 (56)</td>
<td>4 (44)</td>
<td>3 (33)</td>
</tr>
<tr>
<td>Spine (n = 56)</td>
<td>35 (63)</td>
<td>29 (52)</td>
<td>21 (38)</td>
<td>25 (45)</td>
</tr>
<tr>
<td>European Spine Journal (n = 25)</td>
<td>11 (44)</td>
<td>8 (32)</td>
<td>6 (24)</td>
<td>6 (24)</td>
</tr>
<tr>
<td>P value</td>
<td>0.253</td>
<td>0.219</td>
<td>0.399</td>
<td>0.200</td>
</tr>
<tr>
<td>Setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-center (n = 62)</td>
<td>33 (53)</td>
<td>24 (39)</td>
<td>20 (32)</td>
<td>23 (37)</td>
</tr>
<tr>
<td>Multi-center (n = 28)</td>
<td>19 (68)</td>
<td>18 (64)</td>
<td>11 (39)</td>
<td>11 (39)</td>
</tr>
<tr>
<td>P value</td>
<td>0.193</td>
<td>0.024</td>
<td>0.516</td>
<td>0.843</td>
</tr>
<tr>
<td>No. of patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N &gt; 100 (n = 47)</td>
<td>30 (64)</td>
<td>26 (55)</td>
<td>18 (38)</td>
<td>24 (51)</td>
</tr>
<tr>
<td>N ≤ 100 (n = 43)</td>
<td>22 (51)</td>
<td>23 (53)</td>
<td>13 (30)</td>
<td>10 (23)</td>
</tr>
<tr>
<td>P value</td>
<td>0.224</td>
<td>0.862</td>
<td>0.421</td>
<td>0.007</td>
</tr>
</tbody>
</table>
However, the majority of the trials with adequate allocation concealment used the sealed envelope method. Only a few trials used central randomization. It is suggested that the envelope method is more vulnerable to manipulation than other approaches [13]. Some investigators have even considered this method inadequate unless it is performed by an independent third party [23]. With this stricter criterion, 23 of 42 trials with adequate allocation concealment would have been considered inadequate. Therefore, if future investigators use the envelope method, they should follow the process recommended by the Cochrane Handbook to guarantee adequate allocation concealment.

The double-blind method is an important way of avoiding bias. It aims to eliminate subjective bias on both the subjects and the investigators. However, due to the nature of surgery and rehabilitation in the treatment of spine diseases, it is difficult even impossible, to use the double-blind method to compare surgical or rehabilitation interventions. For it would be unrealistic to blind the surgeons performing surgical interventions or the subjects undergoing rehabilitation. In this study, only 31 trials out of 90 used adequate blinding. Nevertheless, 2011 Cochrane Handbook recommended solutions that the following situations should be considered as adequate blinding. First, the outcome and the outcome measurement are not likely to be influenced by lack of blinding. Second, either the participants or some key study personnel are not blinded, but the outcome assessment is blinded, although it is not easy to conduct a double-blind surgical trial. It is basically possible to find a blinded outcome assessor.

The median sample size of the 90 trials was relatively small. The practice of sample size calculation in the 34 trials was suboptimal. Less than half (37.8 %) of the trials clearly specified the process of sample size calculation. We should pay much more attention to sample size determination before the trial because adequate sample size calculation demonstrates how well the trial is designed. It also could determine whether the results of trials showing no difference between two interventions, namely, negative trials, are valid and clinically useful. We suggest that spinal investigators should always determine the power of their trials before they begin to enroll patients. When sample size is smaller than needed, it is easy to draw negative conclusions. When it is larger, it will cost extra use of time and money.

In the current study, we found that 21 trials (23.3 %) did not disclose whether they were sponsored by industry or not. However, industry sponsorship is an important source of bias. Conclusions in trials funded by for-profit organizations may be more positive due to biased interpretation of trial results. Readers should carefully evaluate whether conclusions in randomized trials are supported by data [24]. Future investigators should not only disclose fully all potential conflicts of interest but also should register all trials and make all results available in publicly accessible registries to minimize publication bias.

In other medical disciplines, methodological quality of RCTs was assessed with the similar method. Bai et al. [7] reported quality of RCTs in major gastroenterology and hepatology journals in 2006. They included one hundred five trials and found that 81 % (85/105) reported adequate generation of the allocation sequence, 61 % (64/105) reported adequate allocation concealment, 51 % (54/105) were double-blind, and 75 % (79/105) reported adequate sample size calculation. Gabriel et al. [25] reported methodological quality of RCTs of anticoagulation in cancer. The results were adequate sequence generation (85 %), adequate allocation concealment (61 %), participants’ blinding (39 %), data collectors’ blinding (44 %), providers’ blinding (41 %), outcome assessors’ blinding (75 %) and data analysts’ blinding (15 %). In medical oncology, Julien et al. [9] showed that from 2005 to 2009, adequate random allocation (29 %), proper blinding (41 %), adequate allocation concealment (51 %), and participant flow (59 %) were reported among 357 RCTs. Generally, the RCTs involving medicine were reported with better quality. Comparatively speaking, RCTs quality in spinal journals is not satisfactory. This is probably due to more difficult control of surgical techniques or rehabilitation methods in RCTs.

The current study had several limitations. First, only 90 trials were included. We therefore provided the proportions with 95 % confidence intervals. Second, the accuracy of the methodological quality assessment is affected by the quality of the reporting. A trial with severe methodological limitations may be considered as an excellent trial if the limitations were not reported, whereas a well-designed and conducted trial may be considered to carry substantial bias if it was inadequately reported [26]. For example, if the authors do not mention the use of sealed envelopes for sequence concealment, readers should not assume that this is not performed. On the other hand, the use of sealed envelopes does not necessarily mean that it is adequately done. Nevertheless, it has been demonstrated that the quality of the trial’s design and its performance were positively related to the reporting quality [27]. Thus, the reporting could, to some extent, reflect the quality of a trial’s design and conduct.

In summary, the current study showed that the quality of the methods of RCTs in three major spine journals needs further improvement. Importantly, we hope journal editors, authors, and readers can pay more attention to the reported methodology of randomized trials to keep RCT as one of the best methods for obtaining convincing evidence in the field of spinal diseases.
Acknowledgments  We are indebted to the authors of the primary studies.
Conflict of interest  No competing interest declared.

References

Disk degeneration of the upper lumbar disks is associated with hip pain

Evelien I. T. de Schepper · Jurgen Damen · Pieter K. Bos · Albert Hofman · Bart W. Koes · Sita M. Bierma-Zeinstra

Received: 28 June 2011 / Revised: 7 August 2012 / Accepted: 28 October 2012 / Published online: 8 November 2012
© Springer-Verlag Berlin Heidelberg 2012

Abstract

Purpose A possible cause of hip pain is the presence of radiating pain from the higher lumbar spine. Identification of factors associated with hip pain arising from the lumbar spine would aid the physician. The first step in identifying possible factors is to look at the association between hip pain and osteoarthritis of the lumbar spine.

Methods In an open population based study of people 55 years and older (Rotterdam study), 2,819 lumbar radiographs were scored for the presence and severity of individual radiographic features of disk degeneration. Hip osteoarthritis was scored on anteroposterior pelvic radiographs, and questionnaires including self-reported hip pain were taken. Logistic regression adjusted for possible confounders was used to determine the association between self-reported hip pain and the individual radiographic features of lumbar disk degeneration.

Results The presence of dis space narrowing grade \( \geq 1 \) at level L1/L2 was significantly associated with hip pain in the last month (men OR = 2.0; 95 % CI 1.1–3.8 and women OR = 1.7; 95 % CI 1.1–2.5). The presence of disk space narrowing grade \( \geq 1 \) at level L2/L3 was only significantly associated with hip pain in women. The strength of the associations increased for self-reported chronic hip pain, especially in men (L1/L2 OR = 2.5; 95 % CI 1.3–5.0). The presence of disk space narrowing at the lower levels (L3/L4/L5/S1) was not significantly associated with hip pain.

Conclusion Our data provide evidence for an association between hip pain and disk space narrowing at disk level L1/L2 and L2/L3. In case of uncertainty of the cause of hip pain, evaluation of lumbar radiographs may help to identify those hip pain patients who might have pain arising from the lumbar spine.

Keywords Hip pain · Back pain · Lumbar disk degeneration

Introduction

Hip pain is a common symptom among older adults, with a point prevalence of 14.3 % reported in the United States [1]. The differential diagnosis of hip pain is broad and includes intra-articular pathology, extra-articular pathology, and other causes like radiating pain from the lumbar spine. Differentiating back pain from hip pain in patients who present with classic signs and symptoms is mostly not difficult and generally does not require further testing to establish an accurate diagnosis. However, in some cases, patients present with nonspecific complaints of pain in the lumbar spine, buttock, lateral hip, or thigh [2]. The differentiation of signs and symptoms suggestive of hip disorders versus spine disorders is important in giving patients the most beneficial treatment, especially if the treatment includes a major reconstructive surgery, such as hip replacement.

Differentiating whether hip pain originates from the hip, the spine, or both may be challenging. Brown et al. [3]
attempted to determine which physical signs and symptoms best predict the primary source of pain in patients with hip-, spine-, or concomitant disorders. After final diagnosis with imaging studies, they found that although limited internal rotation, groin pain, and a limp are more commonly associated with a hip disorder, these symptoms are also seen in patients with spine alone or both hip and spine disorders.

In order to make a differentiation between hip and spine originated hip pain there have been a few studies about the usefulness of local anesthetic with(out) corticosteroid hip infiltrations, to differentiate intra-articular causes of hip pain from spinal causes [4–7]. To our knowledge, there have been no studies about the usefulness of local spine infiltrations to differentiate hip and spine originated hip pain. However, infiltration of every patient with atypical hip pain for possible coexistent lumbar spine osteoarthritis would be counterproductive and costly. Preoperative identification of factors associated with hip pain arising from the lumbar spine would aid the physician by identifying the subgroup of patients who might not experience full relief of pain with a hip arthroplasty.

One of the first steps to identify possible factors is to look at the association between hip pain and osteoarthritis of the lumbar spine. The purpose of this study was to explore the association of self-reported hip pain with the different individual radiographic features (IRF) of spinal osteoarthritis by vertebral level, including osteophytes and disk space narrowing.

Materials and methods

Study population

The data for this study originate from data of the Rotterdam Study, an open population prospective cohort of people aged 55 years and older. The objective of the Rotterdam Study is to investigate the incidence of, and risk factors for, chronic disabling diseases. The study design has been described previously [8]. All 10,275 inhabitants of Ommoord (a district in Rotterdam, the Netherlands) were invited to participate. The baseline measurements were conducted between 1990 and 1993. In total, 7,983 participants were examined.

For this study, 2,819 lumbar radiographs were scored. The selection was based on the availability of radiographs of the hip and spine at a follow-up measurement 6.6 years later [9, 12].

Radiographic scoring

Lumbar lateral radiographs were scored by a single observer trained by a radiologist for the presence of the individual radiographic features of disk degeneration. The observer was blinded to clinical characteristics of the participants. Each vertebral level from L1/2 to L5/S1 was reviewed for the presence and severity of osteophytes (anterior) and vertebral narrowing, using the Lane atlas [10, 11]. In this atlas grade 0 = none; grade 1 = mild; grade 2 = moderate; and grade 3 = severe. The lumbosacral disk space was defined as narrowed when its height was less than that of the disk space between the third and fourth lumbar vertebrae. This is due to a normal progression of increasing disk space height from the third and fourth to the fourth and fifth lumbar vertebrae, and then a relative narrowing of the height of the lumbosacral disk space. Sclerosis was not scored because of the earlier reported low ICC for this feature [11].

Inter-observer reproducibility was assessed by a second independent observer who evaluated a random selection of 140 (5 %) X-rays. The ICC was 0.83 for osteophytes and 0.77 for vertebral narrowing, indicating good reproducibility.

Weight bearing anteroposterior radiographs of the pelvis were obtained. One trained reader evaluated the radiographs obtained at baseline, unaware of the clinical status of the participants [9]. At baseline, radiological osteoarthritis of the hip was quantified by measurements following the Kellgren & Lawrence grading system (atlas-based) in five grades (from 0 to 4). A person was considered to have osteoarthritis of the hip, if the Kellgren & Lawrence score of one or both joints was equal to or larger than two [9].

Hip pain

Hip pain and low back pain were determined from interviewing the participants during the home visits. Participants were asked “Did you have complaints of the (left and/or right) hip during the last month?” Hip pain was defined to be present if the answer was positive. Participants were subsequently asked “What is the duration of the present hip complaints?” For low back pain similar questions were asked. We defined chronic hip pain to be present if the duration of the hip joint pain was more than 1 year.

Participants also visited the research center, where X-rays were obtained. Height and weight were measured with participants wearing indoor clothing and without shoes. Body mass index (BMI) was calculated as weight in kilograms divided by length in meters squared (kg/m²).

Statistical analyses

We defined disk space narrowing to be present if the grade was mild, moderate, or severe (grade ≥1). Because of the small proportion of subjects without osteophytes, we used a higher cutoff value for this feature. We defined osteophytes
to be present if the grade was moderate or severe (grade ≥2) [12]. Using these definitions we calculated the prevalence of the IRF by vertebral level (L1/2 to L5/S1) and gender.

In order to explore the association between the IRF by vertebral level and hip pain, hip pain was used as the dependent variable with adjustments made for age and gender. The assessments of the associations were also adjusted for radiological hip osteoarthritis, as this variable was shown to be associated with disk space narrowing and of course with hip pain. The same was true for low back pain [12]. In addition, the assessments of the association were also adjusted for BMI, as this variable has been reported to be associated with both hip pain and some of the individual radiographic features [13–15].

In a separate analysis we explored the association between the IRF by vertebral level and hip pain in subjects with no sign of radiological hip osteoarthritis. The results of the analyses are expressed as odds ratios (OR) with 95% confidence intervals (CI), stratified for gender. The same methods were used to explore the association between the IRF and chronic hip pain. Statistical analysis was performed using SPSS version 15 (SPSS Inc, Chicago, USA).

**Results**

**Subject characteristics**

Baseline characteristics are shown in Table 1. There were 1,204 men (mean age 65.3 years, standard deviation (SD) 6.4) and 1,615 women (mean age 65.9 years, SD 6.8). Hip pain during the last month was reported more often by women than men [244 (15.1 %) vs. 84 (7.0 %); p < 0.05] (Table 1). Chronic hip pain was reported in the majority (82 %) of the current hip pain cases and was also more often reported by women [208 (12.9 %) vs. 62 (5.1 %); p < 0.05].

Radiological hip osteoarthritis was observed in 209 (7.4 %) persons (Kellgren & Lawrence ≥2 in one or both hips).

**Influence of gender and vertebral level**

The prevalence of the IRF in men and women is shown in Table 1. Osteophytes were the most frequently observed radiographic feature and were slightly more common in men than women (95 vs. 91 %; p < 0.05). Disk space narrowing was more frequent in women than men (65 vs. 53 %; p < 0.05). In terms of their distribution by vertebral level, narrowing was more frequent at the lower lumbar disk levels.

Disk space narrowing grade ≥1 at level L1/L2 was more common in persons with hip pain (19 vs. 10 %; p < 0.05). Hip pain was more common in persons with disk space narrowing grade ≥1 at level L1/L2 (21 vs. 11 %; p < 0.05).

**Association with LDD**

Table 2 shows the association between hip pain and the IRF, adjusted for age, gender, BMI, hip arthritis, and low back pain. The presence of disk space narrowing grade ≥1 at level L1/L2 was significantly associated with hip pain in the last month, both in men and women (men OR = 2.0; 95 % CI 1.1–3.8 and women OR = 1.7; 95 % CI 1.1–2.5) (Table 2). The presence of disk space narrowing grade ≥1 at level L2/L3 was significantly associated with hip pain in the last month, only in women (OR = 1.6; 95 % CI 1.1–2.2). The strength of the associations increased for the participants with chronic hip pain, especially for men (L1/L2 OR = 2.5; 95 % CI 1.3–5.0). The strength of the associations also increased for the group of subjects with no radiological hip osteoarthritis (men chronic pain L1/L2 OR = 2.7; 95 % CI 1.3–5.5 and women chronic pain L1/L2 OR = 2.0; 95 % CI 1.3–3.2).

The presence of disk space narrowing at the lower levels (L3/L4/L5/S1) was not significantly associated with hip pain. The presence of disk space narrowing grade ≥2 was not explored, because of the low number of persons with disk space narrowing grade ≥2 at the upper levels.

The presence of osteophytes grade ≥2 was not significantly associated with hip pain at any level (data not shown).

**Discussion**

The differentiation of signs and symptoms suggestive of hip disorders versus spine disorders is important in giving patients the most beneficial treatment. The purpose of this study was to explore the association of self-reported hip pain with the different individual radiographic features (IRF) of spinal osteoarthritis. In this study, disk space narrowing at level L1/L2 appeared to be associated with pain in the hip region, especially in men. The strength of the associations increased for participants with chronic hip pain and in those without radiological signs of hip osteoarthritis. These results suggest that in case of uncertainty of the cause of hip pain, evaluation of lumbar radiographs may help to identify those hip pain patients who may benefit the most from further diagnostic evaluation.

Our data provide evidence for radiating pain from the higher lumbar spine as a possible cause of hip pain in a cross-sectional open population based study. One of the explanations that can be found for the association between hip pain and disk space narrowing at level L1/L2 and
L2/L3 is "referred pain." The term "referred pain" is used for pain localized not in the site of its origin but in areas that may be adjacent or at a distance from such a site. Several theories have been proposed to explain the "referred pain" phenomenon, with the convergence-projection theory the most widespread [16, 17]. Input from different tissue types converge on the same dorsal horn neurons [18]; After activation, increased nociceptive input is transmitted supraspinally and misinterpreted at the cortical level as pain from other tissues. It is possible that the reduction of space between the vertebrae as a consequence of the degenerative disk leads to increased pressure on spinal ligaments and other supporting tissues. This can be misinterpreted at the cortical level as pain from other tissues, like the hip region. Experimental studies have confirmed that noxious stimulation of interspinous ligament, facet joint, and paravertebral muscles causes referred pain that can radiate into the extremity [19–21].

Another explanation for the radiating pain from the higher lumbar spine can be found in the dermatomal innervations of the hip region. It is suggested that impingement of the higher lumbar spinal nerve roots (L1–L3) can cause pain in the dermatomal distribution surrounding the hip. The dermatomal distribution of the L1 spinal nerve is located in the groin and the upper part of the buttock. The distribution of the L2 spinal nerve is located in the outside thigh. It is possible that reduction of space between the vertebrae as a consequence of the degenerative disk is more likely to lead to impingement of the L1 and L2 nerve roots, and, therefore, causes pain in the dermatomal distribution. Spinal nerve roots pass through the intervertebral foramen as they travel from the spinal cord toward the periphery. It has been reported that narrowing of the disk space can reduce the vertical diameter of this intervertebral foramen [22].

The explanation for the stronger association between hip pain and disk space narrowing compared with the presence of osteophytes is unknown. This study evaluates the severity of anterior osteophytes, unfortunately we could not evaluate any bony aspects of the intervertebral foramen.

Table 1  Frequency of hip pain and individual radiographic features of the low back in men and women

<table>
<thead>
<tr>
<th></th>
<th>Men, N = 1,204</th>
<th>Women, N = 1,615</th>
<th>All, N = 2,819</th>
<th>Hip pain, N = 328</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years) Mean ± SD</td>
<td>65.3 ± 6.4</td>
<td>65.9 ± 6.8</td>
<td>65.7 ± 6.6</td>
<td>66.2 ± 6.8</td>
</tr>
<tr>
<td>Body mass index (BMI) Mean ± SD</td>
<td>25.9 ± 2.9</td>
<td>26.6 ± 3.8</td>
<td>26.3 ± 3.5</td>
<td>27.0 ± 3.9</td>
</tr>
<tr>
<td>Hip pain (%)a</td>
<td>84 (7.0)</td>
<td>244 (15.1)</td>
<td>328 (11.6)</td>
<td>328 (100)</td>
</tr>
<tr>
<td>Chronic hip pain (%)b</td>
<td>62 (5.1)</td>
<td>208 (12.9)</td>
<td>270 (9.6)</td>
<td>270 (82.3)</td>
</tr>
<tr>
<td>Hip osteoarthrosis (%)</td>
<td>94 (7.8)</td>
<td>115 (7.1)</td>
<td>209 (7.4)</td>
<td>51 (15.5)</td>
</tr>
<tr>
<td>Osteophytes low back (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade ≥1</td>
<td>1,148 (95.3)</td>
<td>1,467 (90.8)</td>
<td>2,615 (92.8)</td>
<td>306 (93.3)</td>
</tr>
<tr>
<td>Grade ≥2</td>
<td>832 (69.1)</td>
<td>929 (57.5)</td>
<td>1,761 (62.5)</td>
<td>217 (66.2)</td>
</tr>
<tr>
<td>Grade 3</td>
<td>536 (44.5)</td>
<td>505 (31.3)</td>
<td>1,041 (36.9)</td>
<td>134 (40.9)</td>
</tr>
<tr>
<td>Narrowing low back (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade ≥1</td>
<td>637 (52.9)</td>
<td>1,048 (64.9)</td>
<td>1,685 (59.8)</td>
<td>210 (64.0)</td>
</tr>
<tr>
<td>Grade ≥2</td>
<td>286 (23.8)</td>
<td>525 (32.5)</td>
<td>811 (28.8)</td>
<td>115 (35.1)</td>
</tr>
<tr>
<td>Grade 3</td>
<td>40 (3.3)</td>
<td>107 (6.6)</td>
<td>147 (5.2)</td>
<td>20 (6.1)</td>
</tr>
<tr>
<td>Osteophytes ≥2 (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1–2</td>
<td>282 (23.4)</td>
<td>297 (18.4)</td>
<td>579 (20.5)</td>
<td>84 (25.6)</td>
</tr>
<tr>
<td>L2–3</td>
<td>347 (28.8)</td>
<td>404 (25.0)</td>
<td>751 (26.6)</td>
<td>105 (32.0)</td>
</tr>
<tr>
<td>L3–4</td>
<td>428 (35.5)</td>
<td>364 (22.6)</td>
<td>792 (28.1)</td>
<td>100 (30.5)</td>
</tr>
<tr>
<td>L4–5</td>
<td>403 (33.5)</td>
<td>354 (21.9)</td>
<td>757 (26.9)</td>
<td>94 (28.7)</td>
</tr>
<tr>
<td>L5–S1</td>
<td>312 (25.9)</td>
<td>303 (18.8)</td>
<td>615 (21.8)</td>
<td>68 (20.7)</td>
</tr>
<tr>
<td>Narrowing ≥1 (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1–2</td>
<td>107 (8.9)</td>
<td>201 (12.5)</td>
<td>308 (10.9)</td>
<td>63 (19.2)</td>
</tr>
<tr>
<td>L2–3</td>
<td>135 (11.3)</td>
<td>307 (19.0)</td>
<td>442 (15.7)</td>
<td>81 (24.7)</td>
</tr>
<tr>
<td>L3–4</td>
<td>153 (12.7)</td>
<td>342 (21.1)</td>
<td>495 (17.6)</td>
<td>78 (23.8)</td>
</tr>
<tr>
<td>L4–5</td>
<td>268 (22.2)</td>
<td>526 (32.6)</td>
<td>794 (28.2)</td>
<td>111 (33.8)</td>
</tr>
<tr>
<td>L5–S1</td>
<td>408 (34.0)</td>
<td>662 (41.0)</td>
<td>1,070 (38.0)</td>
<td>127 (38.7)</td>
</tr>
</tbody>
</table>

[a] Hip pain: complaints of the hip joint during last month
[b] Chronic hip pain: duration present hip joint complaints >1 year
and disk space narrowing at L1/L2 in men compared with the association in women is also unknown. It is possible that even though women reported hip pain more often, only a small proportion of the complaints are due to disk space narrowing, whereas other factors determine the feeling of pain. Men and women could also report pain differently, therefore, affecting the association between hip pain, disk space narrowing, and gender. Cecchi et al. [23] showed that women presented with significantly more severe pain than men. Finally, the explanation for the absence of an association between hip pain and disk space narrowing at L2/L3 in men compared to women is also unknown. It is maybe due to an evidently lower prevalence of disk space narrowing at L2/L3 in men compared to women.

Our study had several advantages. It was population based with a relatively high number of subjects. We used a semi-quantitative score, using standard radiographs, to characterize the presence and severity of hip and spine osteoarthritis. Assessment of the radiographs was carried out without knowledge of the questionnaire data, and so any errors in classification are likely to have been non-directional. We defined chronic hip pain and chronic low back pain to be present if the duration of the hip joint pain was more than 1 year. In literature, others have chosen 3 months or even 6 months as the dividing line between acute and chronic pain [24]. However, with our definition, chronic pain included long lasting chronic complaints with long lasting impact on ones life.

However, there are several limitations in our explorative study that need to be considered when interpreting the results. Our data did not include the precise location of the hip pain. This limitation is partly undermined by the fact that the dermatomal distribution of L1 and L2 includes the upper part of the buttock, the groin, and the lateral thigh, which cover a wide area of the hip region. Further, our data did not include a clinical evaluation of the hip pain. In this way we could not account for the potential of soft-tissue pathology contributing to the reported hip pain.

Moreover, hip osteoarthritis was only considered when the Kellgren & Lawrence score of one or both joints was equal to or larger than two in agreement with conventional epidemiological definitions for hip osteoarthritis [25]. In this way there is still a possibility of the presence of hip osteoarthritis which is not clearly visible yet on radiographs at that time point. In order to exclude the possibility of this confounding, we reanalyzed the data with adjusting for the presence of radiographic hip osteoarthritis 6.6 years later. We defined a new variable that included all the participants with hip osteoarthritis at baseline and/or hip osteoarthritis 6.6 years later (n = 413). The strength of the associations was unchanged (for chronic pain the L1/L2 OR was 1.9; 95 % CI 1.3–2.7; again higher in men (OR = 2.7; 95 % CI 1.4–5.3) than in women (OR = 1.7; 95 % CI 1.1–2.6).

Furthermore, there could be some selection bias in favor of relatively healthy participants. The participants in the present study had to be mobile enough to visit the research center for X-ray examination, both for the baseline and follow-up appointments (mean 6.6 years) [9]. In other words, patients with the most severe symptoms were most

<table>
<thead>
<tr>
<th>Narrowing level</th>
<th>N (%)</th>
<th>Hip pain OR (95 % CI)</th>
<th>Chronic hip pain OR (95 % CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men, N = 1,204</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1–L2</td>
<td>107 (8.9)</td>
<td>2.0 (1.1–3.8)*</td>
<td>2.5 (1.3–5.0)**</td>
</tr>
<tr>
<td>L2–L3</td>
<td>135 (11.3)</td>
<td>0.9 (0.4–1.8)</td>
<td>1.1 (0.5–2.4)</td>
</tr>
<tr>
<td>L3–L4</td>
<td>153 (12.7)</td>
<td>1.1 (0.6–2.1)</td>
<td>1.1 (0.5–2.2)</td>
</tr>
<tr>
<td>L4–L5</td>
<td>268 (22.2)</td>
<td>1.2 (0.7–2.0)</td>
<td>1.4 (0.8–2.5)</td>
</tr>
<tr>
<td>L5–S1</td>
<td>408 (33.9)</td>
<td>0.7 (0.4–1.1)</td>
<td>0.6 (0.4–1.1)</td>
</tr>
<tr>
<td>Women, N = 1,615</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1–L2</td>
<td>201 (12.5)</td>
<td>1.7 (1.1–2.5)*</td>
<td>1.8 (1.1–2.7)**</td>
</tr>
<tr>
<td>L2–L3</td>
<td>307 (19.0)</td>
<td>1.6 (1.1–2.2)*</td>
<td>1.6 (1.1–2.3)*</td>
</tr>
<tr>
<td>L3–L4</td>
<td>342 (21.1)</td>
<td>1.0 (0.7–1.4)</td>
<td>1.1 (0.7–1.5)</td>
</tr>
<tr>
<td>L4–L5</td>
<td>526 (32.6)</td>
<td>0.9 (0.7–1.3)</td>
<td>1.0 (0.7–1.4)</td>
</tr>
<tr>
<td>L5–S1</td>
<td>662 (41.0)</td>
<td>1.0 (0.7–1.3)</td>
<td>0.9 (0.7–1.2)</td>
</tr>
<tr>
<td>All, N = 2,819</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1–L2</td>
<td>308 (10.9)</td>
<td>1.8 (1.3–2.5)**</td>
<td>2.0 (1.4–2.8)**</td>
</tr>
<tr>
<td>L2–L3</td>
<td>442 (15.7)</td>
<td>1.4 (1.0–1.9)*</td>
<td>1.5 (1.1–2.1)*</td>
</tr>
<tr>
<td>L3–L4</td>
<td>495 (17.6)</td>
<td>1.1 (0.8–1.4)</td>
<td>1.1 (0.8–1.5)</td>
</tr>
<tr>
<td>L4–L5</td>
<td>794 (28.2)</td>
<td>1.0 (0.8–1.3)</td>
<td>1.1 (0.8–1.5)</td>
</tr>
<tr>
<td>L5–S1</td>
<td>1,070 (38.0)</td>
<td>0.9 (0.7–1.1)</td>
<td>0.8 (0.6–1.1)</td>
</tr>
</tbody>
</table>

* p ≤ 0.05
** p ≤ 0.01

OR odds ratio
CI Confidence interval
likely not included, but this may be inevitable in long-term prospective cohort studies.

What are the implications of these findings for researchers and clinicians?

Accurate diagnosis of pain originating from the hip joint can be clinically challenging. There have been several studies about the usefulness of hip injections to differentiate intra-articular causes of hip pain from spinal causes [4–7]. To our knowledge, there have been no studies about the usefulness of local spine infiltrations to differentiate hip and spine origin hip pain. A possible explanation for this is the availability of a successful treatment for degenerated hip disease (hip arthroplasty), but less predictable treatment options for degenerative spine disorders.

The differentiation of signs and symptoms suggestive of a hip disorder is important in giving patients adequate information regarding their condition and for applying the most beneficial treatment. Our data provides evidence for an association between hip pain and disk space narrowing at disk level L1/L2 and L2/L3. In case of uncertainty of the cause of hip pain, evaluation of lumbar radiographs may help to identify those hip pain patients who might have pain arising from the lumbar spine. Perhaps hip infiltration in patients without higher lumbar disk degeneration is even unnecessary. However, well-designed studies are needed to verify this hypothesis.

**Conclusion**

In conclusion, this study explores the association of self-reported hip pain with lumbar spine osteoarthritis. Our data provide evidence for an association between hip pain and disk space narrowing at disk level L1/L2 and L2/L3. In case of uncertainty of the cause of hip pain, evaluation of lumbar radiographs may help to identify those hip pain patients who might have pain arising from the lumbar spine. Well-designed studies are needed to verify this hypothesis.

**Acknowledgments** No financial or material support for the research and the work was received.

**References**

Randomized, Double-blind, Placebo-Controlled, Trial of Transforaminal Epidural Etanercept for the Treatment of Symptomatic Lumbar Disc Herniation

Brian J. C. Freeman, FRCS (Tr & Orth), FRACS (Ortho), DM,*† Guy L. Ludbrook, FANZCA, PhD,*† Stephen Hall, FRACP,* Michael Cousins, FAznCA, FFPRANZCA,§ Bruce Mitchell, FACSp, FAcC, MpainMed,¶ Mark Jaros, PhD,** Michael Wyand, DVM, PhD,** and James R. Gorman, MD, PhD**

Study Design. Multicenter, randomized, double-blind, placebo-controlled trial.  
Objective. To examine the safety and efficacy of three different doses of the tumor necrosis factor alpha (TNF-α) inhibitor etanercept versus placebo for the treatment of symptomatic lumbar disc herniation (LDH).

Summary of Background Data. TNF-α is considered to be a major cause of radicular leg pain associated with symptomatic LDH. Systemic administration of TNF-α inhibitors for sciatica has indicated a trend toward efficacy.

Methods. Forty-nine subjects aged between 18 and 70 years, with persistent lumbosacral radicular pain secondary to LDH, and an average leg pain intensity of 5/10 or more were randomized to 1 of 4 groups: 0.5-mg, 2.5-mg, 12.5-mg etanercept, or placebo. Subjects received 2 transforaminal epidural injections, 2 weeks apart, and were assessed for efficacy up to 26 weeks after the second injection. The primary outcome measure was the change in mean daily worst leg pain (WLP). Secondary outcomes included average leg pain, worst back pain, average back pain, in-clinic pain, Oswestry Disability Index, patient global impression of change, and tolerability.

Results: Forty-three of the 49 randomized patients completed the study. Patients receiving 0.5-mg etanercept showed a clinically and statistically significant (P < 0.1) reduction in mean daily WLP compared with the placebo cohort from 2 to 26 weeks for both the per protocol population (−5.13 vs. −1.95; P = 0.066) and the intention-to-treat population (−4.40 vs. −1.84; P = 0.058). Fifty percent of these subjects reported a 100% reduction in WLP 4 weeks post-treatment compared with 0% of subjects in the placebo cohort. Improvements in all secondary outcomes were also observed in the 0.5-mg etanercept cohort. The overall incidence of adverse events was similar in placebo and all etanercept cohorts.

Conclusion. Two transforaminal injections of etanercept provided clinically significant reductions in mean daily WLP and worst back pain compared with placebo for subjects with symptomatic LDH. Epidural etanercept may offer patients with sciatica a safe and effective nonoperative treatment.

Key words: lumbar disc herniation, sciatica, epidural injection, tumor necrosis factor alpha (TNF-α) inhibitors, etanercept.

Level of Evidence: N/A


Tumor necrosis factor alpha (TNF-α) is considered to be a major cause of radiculopathy associated with symptomatic lumbar disc herniation.1,2 TNF-α is a critical mediator of nerve root inflammation, central sensitization, and neuropathic pain that underlie sciatica.3-5 Studies using TNF-α inhibitors for the treatment of sciatica have had varying results.6-11 These studies predominantly used either intravenous or subcutaneous routes of administration. Because the vertebral disc is poorly vascularized, systemic administration of TNF-α inhibitors would require high doses to achieve efficacy. Thus, localized administration of TNF-α inhibitors may be more beneficial and may lead to fewer adverse events (AEs).

Etanercept is a recombinant fusion protein of the p75 subunit of the TNF receptor with the Fc component of the human immunoglobulin G1. Studies have demonstrated that patients with sciatica treated with 3 subcutaneous injections of...
of etanercept demonstrated a significant decrease in leg and back pain and significant improvements in Oswestry Disability Index (ODI) and Roland Morris Disability Questionnaire compared with historical intravenous steroid controls. Etanercept has also been investigated via direct administration in patients with sciatica. Cohen et al. found intradiscal etanercept (0.1–1.5 mg) to be ineffective for the treatment of chronic radicular pain or discogenic low back pain. However, the same group later found significant improvements in leg and back pain for patients with subacute radiculopathy after epidural administration of etanercept (2–6 mg) compared with placebo.

This multicenter randomized, double-blind, placebo-controlled trial was designed to investigate the safety and efficacy of etanercept delivered via the transforaminal epidural route for the treatment of symptomatic lumbar disc herniation using an expanded dose range, in an expanded cohort, in a trial with highly structured inclusion and exclusion criteria, carried out in multiple centers.

MATERIALS AND METHODS

This study was conducted between March 2009 and December 2010 at 6 centers in Australia (Royal Adelaide Hospital, Adelaide, South Australia, Australia; Cabrini Medical Centre, Malvern, Victoria, Australia; Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia; Fremantle Hospital, Fremantle, Western Australia, Australia; Metro Spinal Clinic, Fremantle, Western Australia, Australia; Metro Spinal Clinic, Fremantle, Western Australia, Australia; Metro Spinal Clinic, Melbourne, Victoria, Australia; and Royal North Shore Hospital, Sydney, New South Wales, Australia). The human research ethics committees of all participating centers approved the study protocol. All subjects provided written informed consent.

Subjects

Subjects between the ages of 18 and 70 years, with the current diagnosis of lumbar disc radicular pain of 6 to 26 weeks duration secondary to lumbar disc herniation confirmed by radiological means with a mean pain score of 5/10 or more for average leg pain (ALP) were considered for inclusion. Inclusion and exclusion criteria are shown in Tables 1 and 2, respectively.

Potential candidates were screened for eligibility and underwent baseline assessments including history, physical examination, radiological studies, concomitant medications, vital signs, drug-screening test, pregnancy test (where applicable), tuberculin skin test, and laboratory tests (serum hematology, coagulation studies, chemistry, urinalysis) 3 to 14 days before administration of the first dose.

Randomization

One computer-generated randomization schedule was prepared before the start of the study and used across all sites. When notified that a subject had completed the screening period, the investigational pharmacist requested the next available randomization number in sequence and matched this number to the master randomization list and then prepared the assigned study treatment for that subject.

Subjects were randomized in a 1:1:1:1 block manner to 1 of 3 doses of etanercept (0.5, 2.5, 12.5 mg) or placebo, with 10 subjects in each cohort. Subjects who were withdrawn from the study or who were not evaluable per protocol (PP) because of a major protocol deviation were replaced with the replacement subject being allocated the same treatment dose level as the withdrawn or nonprotocol subject.

Blinding

All subjects, clinicians performing transforaminal injections, radiologists reading postinjection films, postinjection evaluators, statisticians, and study personnel remained blinded to treatment until after database lock, except those persons responsible for the packaging and distribution of the trial medications and the biostatistician preparing the random code.

After randomization, each subject received 2 consecutive injections (etanercept or placebo), 2 weeks apart, of the same “treatment” and dose. Doses of etanercept or placebo were administered at study week 2 and 4, with follow-up visits at study week 6, 8, 12, 16, and 28 (2, 4, 8, 12, and 26 weeks after the second injection, respectively). Subjects maintained a study diary as a daily record of pain levels, concomitant medication use, and AEs during the full study period.

TABLE 1. Inclusion Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Subject is 18 to 70 yr of age, is in good general medical and psychological health, and is capable of completing study assessments and procedures.</td>
</tr>
<tr>
<td>2. If female, the subject is either not of childbearing potential, or is practicing an acceptable form of birth control and has a negative pregnancy test at screening.</td>
</tr>
<tr>
<td>3. Subject has a mean score of at least 5/10 for “average leg pain during the past 24 hr” more than 3 to 7 d before the randomization visit.</td>
</tr>
<tr>
<td>4. Subject has a “current” diagnosis of lumbosacral radicular pain of between 6 and 26 wk of duration.</td>
</tr>
<tr>
<td>5. Pain must radiate into the leg in a dermatomal/myotomal distribution consistent with the suspected involved nerve root and the diagnosis of lumbosacral radicular pain.</td>
</tr>
<tr>
<td>6. Diagnosis must be confirmed by CT or MRI related to the symptoms present at screening with this investigation being performed within a maximum of 6 mo before the screening visit. This study should demonstrate disc herniation at a location consistent with the clinical symptoms of radicular pain and nerve root irritation.</td>
</tr>
<tr>
<td>7. At least one of the after: positive straight leg raise (L5, S1), positive femoral stretch test (L3, L4) or other positive test upon physical examination that is consistent with the presence of sciatic nerve root irritation at screening.</td>
</tr>
<tr>
<td>8. Disc herniation must affect the L3, L4, L5, or S1 nerve root.</td>
</tr>
<tr>
<td>9. Negative QuantiFERON-TB Gold test or negative tuberculin skin test.</td>
</tr>
<tr>
<td>10. Subject is willing to keep all analgesic medication and other therapy usage (e.g., physiotherapy, acupuncture, or transcutaneous electrical nerve stimulation) stable or decreased and use the protocol prescribed rescue pain medication as needed.</td>
</tr>
</tbody>
</table>

CT indicates computed tomography; MRI, magnetic resonance imaging.
TABLE 2. Exclusion Criteria

1. Subject has any history of sciatica (persistent radicular leg pain) in the same leg as the current episode, if the previous radicular leg pain either: (a) was accompanied by intervertebral disc herniation or (b) required prescription medication or visits to a health professional in the year before the onset of the current episode of sciatica.

2. Subject has a documented history of either an allergic reaction to, or a clinically significant intolerance to etanercept, oxycodone hydrochloride, or any iodinated contrast agent.

3. Subject has a BMI > 35 kg/m².

4. Subject has a HADS score >10 on either subscale or has an established history of major psychiatric disorder not controlled with medication.

5. Subject has clinically significant abnormalities in clinical chemistry, hematology or urinalysis, including serum glutamic-oxaloacetic transaminase/AST or serum glutamic-pyruvic transaminase/ALT ≥ 2 times the upper limit of the reference range or a creatinine clearance < 70 mL/min (Cockcroft Gault formula) at screening.

6. Subject has significant pain unrelated to the disc herniation (in addition to radicular buttock or leg pain related to the disc herniation).

7. Subject has clinical evidence of radicular pain at “more than” 1 spinal nerve.

8. Subject has received any investigational drug within 30 d before screening, or is scheduled to receive an investigational drug, other than blinded-study drug during the course of this study.

9. Subject has had lumbar or sacral back surgery related to the specific disc that is the cause of the sciatic pain upon presentation to the study, or currently plans to undergo spine surgical intervention while in the study.

10. Subject has received epidural corticosteroid injections in the back “within 6 mo” of screening.

11. Subject is involved in an ongoing worker’s compensation claim, disability, or litigation related to any pain problem.

12. Any active infection or malignancy.

13. History of chronic infection, tuberculosis, HBV, HCV, HIV or diabetes that is either: (a) type 1 (juvenile –onset, insulin dependent); or (b) type II (adult onset) and if type II, then also either currently on insulin or currently showing significant neuropathy with symptoms of pain or sensory loss.


17. Severe spinal stenosis, grade II or higher spondylolisthesis.

18. Coagulopathy.

19. Pregnant or breast-feeding mothers.

20. Systemic infection.

21. Plaque psoriasis, psoriatic arthritis, ankylosing spondylitis, or rheumatoid arthritis.

22. Patients receiving concurrent cyclophosphamide therapy.

23. Patients with Wegener granulomatosis.

BMI indicates body mass index; HADS, Hospital Anxiety and Depression Scale; AST, aspartate aminotransferase; ALT, alanine aminotransferase; HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; TNF, tumor necrosis factor.

study period. At each clinic visit, subjects were assessed for safety by review of AEs, and for efficacy by in-clinic collection of diary cards, assessment of pain, and completion of questionnaires regarding pain and quality of life.

Transforaminal Epidural Injection

After application of skin preparation, 2 mL of 1% plain lignocaine was injected into the skin surface at the point of planned injection. Using a 22-gauge 3.5-in needle, the relevant foramen was targeted via a posterolateral approach (for L3, L4, L5) (Figure 1) or a direct posterior approach (for S1) (Figure 2). Once the needle tip was in the ideal position, and after a negative cerebrospinal fluid aspiration, a maximum of 2.0 mL of nonionic contrast media (Iohexol-180, Amersham Health, Oslo, Norway) was injected. Once the operator confirmed satisfactory flow of contrast in the anteroposterior, lateral (and oblique if required) and after a further negative cerebrospinal fluid aspiration, the 2.0 mL of injectate allocated to the subject was injected slowly for 2 minutes. All patients had pulse and blood pressure checked every 15 minutes for 1 hour and hourly thereafter for a total of 4 hours. A neurological examination was performed before discharge. All
radiographical images (anteroposterior, lateral, and oblique) of the final needle-tip position and subsequent contrast flow were reviewed in a blinded fashion by an expert independent reader within 24 hours to confirm satisfactory needle placement and epidural contrast flow.

Outcomes Assessments
The prespecified primary efficacy endpoint was the difference between each treated cohort and the placebo cohort for average change in mean daily worst leg pain (WLP) comparing scores at baseline to 4 weeks after the second injection. The secondary efficacy endpoints included average change from baseline in mean daily WLP, ALP, worst back pain (WBP), average back pain (ABP), in-clinic pain assessment, ODI, and patient global impression of change at each postdose visit. Safety monitoring included vital signs, physical examination, and treatment-emergent adverse events (TEAEs).

Statistical Methods
Sample size was calculated using data from Cohen et al.\textsuperscript{13} These authors reported a mean numerical rating leg pain score of 6.3 at baseline with a standard deviation of 1.6 in a cohort of 24 subjects with subacute lumbosacral radiculopathy.\textsuperscript{13} The minimal clinically important difference of the visual analogue scale score for back/leg pain has previously been reported as 2.0.\textsuperscript{14} For this study, the risk of a type I error was set at 10% ($P < 0.10$) and the risk of a type II error was set at 20% (equating to a power = 80%). The estimated sample size was calculated with $n = 10$ for each group and a total sample size of $N = 40$. The prespecified level for statistical significance was set $P < 0.1$. This level was selected on the basis of the stage of the study (phase IIa), the size of the study and the expected treatment effect size and designed to detect a potential signal of efficacy in a small group of subjects.

Subjects were enrolled until 40 subjects were confirmed to have reached visit 5. The primary efficacy variable was summarized in terms of descriptive statistics by treatment. Treatments were compared using a linear model with treatment as a fixed effect and baseline pain score as a covariate. Pairwise contrasts of interest were primarily the etanercept doses (0.5 mg, 2.5 mg, and 12.5 mg) compared with placebo. The primary prespecified comparisons identified in the earlier text adjusted the $P$ values for multiple comparisons using the Benjamini-Hochberg method. Similar methods were used in analyzing the secondary efficacy variables.

The intention-to-treat (ITT) population included all subjects who received any planned study medication after randomization. The PP population included all subjects who received both doses of study medication with independent confirmation of epidural needle placement and contrast flow for the first and second injections and who took at least one pain assessment after the second injection (week 4) visit. The PP population excluded all subjects with protocol deviations and forms the

Copyright © 2013 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
basis for all primary analyses. The safety population included all subjects who received any amount of planned study medication (analyzed as treated). Safety and tolerability data were summarized by treatment group using appropriate descriptive statistics.

Continuous variables were summarized with mean, median, standard deviation, and minimum and maximum values, whereas discrete variables were summarized using number and percentage for each category. Confidence intervals when appropriate were provided at the 90% level for key efficacy variables.

RESULTS
A total of 156 subjects were screened for inclusion in the study. Fifty-one subjects were recruited and randomized, with 49 subjects receiving both transforaminal injections (Figure 3). The number of subjects enrolled by site is shown in Table 3. Six subjects did not complete the study. Two subjects

<table>
<thead>
<tr>
<th>TABLE 3. Number of Subjects Enrolled Per Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5-mg Etanercept</td>
</tr>
<tr>
<td>Site 1</td>
</tr>
<tr>
<td>Site 2</td>
</tr>
<tr>
<td>Site 3</td>
</tr>
<tr>
<td>Site 4</td>
</tr>
<tr>
<td>Site 5</td>
</tr>
<tr>
<td>Site 8</td>
</tr>
<tr>
<td>Total ITT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PP population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
</tr>
<tr>
<td>Site 2</td>
</tr>
<tr>
<td>Site 5</td>
</tr>
<tr>
<td>Site 8</td>
</tr>
<tr>
<td>Total PP</td>
</tr>
</tbody>
</table>

All sites were located in Australia: 1 = Royal Adelaide Hospital, South Australia; 2 = Cabrini Medical Centre, Victoria; 3 = Sir Charles Gairdner Hospital, Western Australia; 4 = Fremantle Hospital, Western Australia; 5 = Metro Spinal Clinic, Victoria; 8 = Royal North Shore Hospital, New South Wales.

A further 3 sites received human research ethics committee approval, but did not screen any potential subjects nor randomize or treat any subjects. The intention-to-treat population is also the safety population.

ITT indicates intention-to-treat; PP, per protocol.
TABLE 4. Baseline Characteristics of the PP Population

<table>
<thead>
<tr>
<th></th>
<th>0.5-mg Etanercept (N = 8)</th>
<th>2.5-mg Etanercept (N = 10)</th>
<th>12.5-mg Etanercept (N = 9)</th>
<th>Placebo (N = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yr (SD)</td>
<td>44.4 (14.2)</td>
<td>50.6 (13.2)</td>
<td>47.0 (9.7)</td>
<td>46.3 (12.1)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>5 (62.5)</td>
<td>6 (60.0)</td>
<td>6 (66.7)</td>
<td>7 (70.0)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>8 (100)</td>
<td>9 (90.0)</td>
<td>9 (100)</td>
<td>9 (90.0)</td>
</tr>
<tr>
<td>Asian</td>
<td>0 (0)</td>
<td>1 (10.0)</td>
<td>0 (0)</td>
<td>1 (10.0)</td>
</tr>
<tr>
<td>BMI, kg/m² (SD)</td>
<td>29.0 (5.1)</td>
<td>27.4 (4.0)</td>
<td>29.0 (3.6)</td>
<td>25.4 (3.7)</td>
</tr>
<tr>
<td>Sciatic pain history</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of condition, mo (SD)*</td>
<td>1.8 (1.8)</td>
<td>0.7 (1.3)</td>
<td>1.6 (1.8)</td>
<td>1.7 (1.8)</td>
</tr>
<tr>
<td>Duration of symptoms, mo (SD)+</td>
<td>3.5 (1.1)</td>
<td>3.9 (1.4)</td>
<td>3.4 (1.3)</td>
<td>3.7 (1.3)</td>
</tr>
<tr>
<td>Average pain during the last day, score (SD)</td>
<td>6.3 (1.3)</td>
<td>5.6 (0.8)</td>
<td>6.0 (1.7)</td>
<td>5.8 (0.8)</td>
</tr>
<tr>
<td>Worst pain during the last day, score (SD)</td>
<td>8.1 (1.5)</td>
<td>7.8 (1.1)</td>
<td>7.7 (2.1)</td>
<td>8.2 (0.8)</td>
</tr>
<tr>
<td>Efficacy variables (PP), mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worst leg pain</td>
<td>7.8 (1.2)</td>
<td>6.8 (1.0)</td>
<td>7.4 (1.8)</td>
<td>6.9 (1.3)</td>
</tr>
<tr>
<td>Average leg pain</td>
<td>6.3 (1.0)</td>
<td>5.6 (0.5)</td>
<td>6.2 (1.0)</td>
<td>5.5 (0.7)</td>
</tr>
<tr>
<td>Worst back pain</td>
<td>6.3 (1.5)</td>
<td>6.5 (1.3)</td>
<td>6.1 (0.7)</td>
<td>5.0 (2.7)</td>
</tr>
<tr>
<td>Average back pain</td>
<td>4.7 (0.6)</td>
<td>5.4 (1.1)</td>
<td>4.7 (1.3)</td>
<td>3.7 (2.0)</td>
</tr>
<tr>
<td>Oswestry Disability Index</td>
<td>36.6 (17.4)</td>
<td>29.6 (12.0)</td>
<td>36.3 (16.5)</td>
<td>31.5 (9.0)</td>
</tr>
<tr>
<td>In-clinic pain, previous 24 hr</td>
<td>6.9 (1.1)</td>
<td>5.9 (0.9)</td>
<td>6.9 (2.2)</td>
<td>5.9 (1.2)</td>
</tr>
<tr>
<td>In-clinic pain, previous week</td>
<td>6.5 (0.9)</td>
<td>5.8 (0.8)</td>
<td>6.6 (1.1)</td>
<td>6.1 (1.2)</td>
</tr>
<tr>
<td>Efficacy variables (ITT), mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worst leg pain</td>
<td>N = 12</td>
<td>N = 13</td>
<td>N = 8</td>
<td>N = 10</td>
</tr>
<tr>
<td>Average leg pain</td>
<td>7.8 (1.2)</td>
<td>7.0 (1.0)</td>
<td>7.3 (1.6)</td>
<td>7.2 (1.3)</td>
</tr>
<tr>
<td>Worst back pain</td>
<td>6.3 (1.2)</td>
<td>5.7 (0.5)</td>
<td>6.2 (1.0)</td>
<td>5.4 (0.7)</td>
</tr>
<tr>
<td>Average back pain</td>
<td>5.7 (2.0)</td>
<td>6.1 (2.0)</td>
<td>6.2 (0.8)</td>
<td>4.7 (2.9)</td>
</tr>
<tr>
<td>Oswestry Disability Index</td>
<td>39.1 (16.6)</td>
<td>31.8 (12.0)</td>
<td>39.6 (17.5)</td>
<td>31.1 (9.4)</td>
</tr>
<tr>
<td>In-clinic pain, previous 24 hr</td>
<td>6.9 (1.3)</td>
<td>6.3 (1.1)</td>
<td>7.2 (1.9)</td>
<td>5.7 (1.1)</td>
</tr>
<tr>
<td>In-clinic pain, previous week</td>
<td>6.7 (1.1)</td>
<td>6.1 (0.9)</td>
<td>6.5 (1.0)</td>
<td>5.9 (1.2)</td>
</tr>
</tbody>
</table>

*Computed from date of diagnosis to date of consent.
†Computed from date of onset of symptoms to date of consent.
SD indicates standard deviation; BMI, body mass index; PP, per protocol population; ITT, intention-to-treat population.

withdrew to undergo treatments for sciatica not permitted under the protocol (1 nonpermitted medication, 1 nonpermitted epidural steroid injection). Four subjects were withdrawn because of AEs (n = 2 for worsening of sciatica; n = 2 for recurrence of radicular pain). Both subjects withdrawing for radicular pain were included in the PP analysis set because withdrawal occurred after week 6.

Two subjects received low volume in 1 of the 2 injections (inadequate dose). For 8 subjects, independent reader review of the fluoroscopy images of the epidural injections failed to confirm appropriate epidural contrast flow at the affected spinal nerve root. In total, 12 subjects were excluded from the efficacy analysis because of protocol violation. The treatment groups were well matched with respect to baseline characteristics for the PP population including demographics, duration, and severity of radicular symptoms (Table 4).

**Efficacy Evaluation**

The safety population and the ITT population were the same population of 49 subjects. The PP population included 37 subjects and provided the key data set for analyses of primary and secondary efficacy endpoints.

The decrease from baseline in mean daily WLP score for the etanercept 0.5-mg group was 4.4 (ITT) and 5.1 (PP) versus a decrease of 1.8 (ITT) and 1.9 (PP) in the placebo group (Table 5). This difference in outcome was clinically and statistically significant (P < 0.1) for the 0.5-mg treatment group compared with the placebo group in both the ITT
Etanercept for Lumbar Radiculopathy • Freeman et al

Randomized Trial

The etanercept 2.5-mg and 12.5-mg treatment groups also demonstrated reduction in mean daily WLP scores compared with placebo (Figure 4). Although the reduction in these 2 groups was not statistically significant when compared with placebo, the pain scores in each drug-treatment group were reduced to a greater extent than seen in the placebo group. Responder rates evaluate the percentage of subjects in each group who show a specified pain response. Fifty percent of subjects in the etanercept 0.5-mg group reported a 100% reduction in the mean daily WLP scores to mean daily WLP scores of zero at 4 weeks post-treatment, whereas no subjects in the placebo group reported a 100% reduction. This difference was clinically and statistically significant.

TABLE 5. Primary Efficacy Endpoint—Change From Baseline in Mean Daily WLP at Visit 5

<table>
<thead>
<tr>
<th></th>
<th>0.5-mg Etanercept</th>
<th>2.5-mg Etanercept</th>
<th>12.5-mg Etanercept</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITT population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline, mean (SD)</td>
<td>7.77 (1.25)</td>
<td>7.04 (0.99)</td>
<td>7.34 (1.65)</td>
<td>7.16 (1.27)</td>
</tr>
<tr>
<td>Visit 5, mean (SD)</td>
<td>3.37 (3.40)</td>
<td>4.79 (2.60)</td>
<td>4.13 (2.47)</td>
<td>5.32 (2.71)</td>
</tr>
<tr>
<td>Change, mean</td>
<td>−4.40</td>
<td>−2.25</td>
<td>−3.21</td>
<td>−1.84</td>
</tr>
<tr>
<td>90% CI</td>
<td>−5.65 to −3.15</td>
<td>−3.45 to −1.06</td>
<td>−4.50 to −1.93</td>
<td>−3.07 to −0.60</td>
</tr>
<tr>
<td>P value vs. placebo</td>
<td>0.0577*</td>
<td>0.6855</td>
<td>0.3035</td>
<td></td>
</tr>
<tr>
<td>PP population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline, mean (SD)</td>
<td>7.85 (1.19)</td>
<td>6.83 (1.03)</td>
<td>7.36 (1.82)</td>
<td>6.95 (1.29)</td>
</tr>
<tr>
<td>Visit 5, mean (SD)</td>
<td>2.71 (3.19)</td>
<td>4.06 (2.51)</td>
<td>4.67 (2.29)</td>
<td>5.00 (2.65)</td>
</tr>
<tr>
<td>Change, mean</td>
<td>−5.13</td>
<td>−2.77</td>
<td>−2.69</td>
<td>−1.95</td>
</tr>
<tr>
<td>90% CI</td>
<td>−6.5 to −3.42</td>
<td>−4.23 to −1.52</td>
<td>−4.07 to −1.23</td>
<td>−3.37 to −0.67</td>
</tr>
<tr>
<td>P value vs. placebo</td>
<td>0.0664*</td>
<td>0.5896</td>
<td>0.5896</td>
<td></td>
</tr>
</tbody>
</table>

Mean values are arithmetic means.

*Significance at the 90% confidence level using the Benjamini-Hochberg adjustment for multiple comparisons.

ITT indicates intention-to-treat; WLP, worst leg pain; PP, per protocol; SD, standard deviation; CI, confidence interval.

Clinically and statistically significant changes in mean daily ALP (Figure 5), WBP (Figure 6), and ABP (Figure 7) were observed for the 0.5-mg etanercept group compared with placebo. Mean daily WBP and ABP were also reduced to a clinically and statistically significant degree in the 2.5-mg etanercept cohort and 12.5-mg etanercept cohort when compared with placebo.

The 0.5-mg etanercept group showed a statistically significant reduction in mean ODI from baseline and also percent ODI change from baseline to week 4. Three months after treatment, the 0.5-mg group consistently maintained at least a 10-point greater change from baseline than the placebo group, and also a percentage reduction at least 30 percentage points greater than that observed in the placebo group. The percentage...
of subjects rating their well-being with patient global impression of change as significantly improved was 63% for the etanercept 0.5-mg group, 30% for etanercept 2.5-mg group, 11% for etanercept 12.5-mg group, and 10% for placebo group.

Safety Evaluation

The overall incidence of AEs in the etanercept treated subjects (59%) was similar to that in the placebo group (67%). The rates in the 3 drug-treatment groups (38% to 88%) clustered around the incidence rate seen in the placebo group (67%), with no consistent dose relationship observed. Commonly occurring AEs included sciatica, headache, nausea, diarrhea, and constipation. Although there was no clear dose-related trend in the incidence of most of these commonly occurring AEs, the possibility of elevated incidence and severity of sciatica in the 12.5-mg etanercept group cannot be ruled out. Most of the total of 100 AEs were classified as mild (N = 66) or moderate (N = 24), with 10 classified as severe (6 worsening sciatic pain, 2 recurrent lumbar radiculopathy, 1 headache, and 1 irregular heart beat).

Most study discontinuations occurred in subjects who had grade 3 AEs of increased sciatic pain. In total, 5 from 49 subjects (10.2%) underwent discectomy surgery within the 6-month follow-up period. The percentage of patients undergoing this surgery was similar among the groups (1 subject from the placebo group, 1 from 0.5-mg etanercept group, 1 from the 2.5-mg group, and 2 from the 12.5-mg group).

DISCUSSION

Persistent moderate to severe leg pain is the most bothersome aspect for patients with radicular pain, contributing to pain burden, disability, decreased quality of life, and progression to surgery. This study achieved its prespecified primary efficacy endpoint by demonstrating a clinically and statistically significant reduction of mean daily WLP 4 weeks after treatment with 2 transforaminal injections of the TNF-α inhibitor etanercept. In addition, a clinically and statistically significant difference between the 0.5-mg etanercept and placebo cohorts was consistently seen across mean daily ALP, and mean daily WBP and ABP when compared with the placebo in the PP population. Results for the ITT population were similarly supportive of a therapeutic effect, with more variation in dose effect. These findings suggest that 2 transforaminal injections of 0.5 mg of etanercept provided substantial rapid reduction in WLP for the majority of treated subjects, persisting for 6 months after treatment.

Although WLP is the most bothersome aspect of radicular pain, and also the most sensitive to measurement of intervention, ALP is more reflective of the overall pain burden for the subject. The 0.5-mg etanercept cohort showed a reduction in ALP similar in magnitude and significance to the reduction in WLP. The etanercept treated groups also showed significant reductions in mean daily WBP and mean daily ABP 4 weeks after the second etanercept injection and were maintained for at least 3 months. These results suggest that epidural etanercept can reduce both radicular pain and back pain. Patients undergoing discectomy for lumbar disc herniation often report a prompt reduction in leg pain, but many complain of persistent or in some cases worse low back pain after surgery.14 In this study, etanercept at the 0.5-mg dose produced impressive reductions both in WBP and ABP out to 6 months.

The observed responder rates and improvement in mean ODI in the 0.5-mg etanercept group suggest that etanercept treatment can reduce disability by a clinically meaningful extent within 2 weeks of treatment and persisting for at least 3 months after the treatment. The observation that 2 weeks after the first injection, more than 63% of the 0.5-mg group, but only 10% of the placebo group, rated themselves as very much improved or much improved, suggests that the treatment not only affected pain scores, but also the subjects’ overall impression of recovery and well-being.

In this study, the lowest dose of etanercept (0.5 mg) was the most effective in reducing WLP and WBP. This finding is consistent with the results of a previously published study of epidural etanercept, where the lowest dose of etanercept (2 mg) was more effective than either the 4 mg or the 6 mg dose, accepting that the numbers in the previous study were small.13

Reverse-dose effects are not unusual in biological systems; one such effect has been reported with off-label use
of recombinant human bone morphogenetic protein-2 in transforaminal lumbar interbody fusion resulting in osteolysis rather than bony fusion. In this case, it has been suggested that osteolysis may be the result of supraphysiological doses of the osteoinductive agent. One possible explanation for the higher doses of etanercept being less efficacious in our study could be that the higher doses, through TNF inhibition, in some way inhibit macrophage-mediated disc resorption. This would need to be explored in larger trials with greater power and a wider dose range, requiring more preclinical safety data before dosing.

Cohen et al recently reported a randomized trial in 84 patients with lumbosacral radiculopathy of less than 6 months duration. Subjects received 2 epidural injections of corticosteroids, etanercept (4 mg) or saline, mixed with bupivacaine. A greater reduction in leg pain 1 month after the second injection was observed with corticosteroids than etanercept or saline. There are important differences between the Cohen study and the study presented here. The study by Cohen et al was less rigorous, for example, including participants with herniated discs or annular tears, mean improvement in visual analogue scale scores for leg pain postintervention were modest (−1.26 to −0.52) for all groups and not statistically significant. Differences in back pain and function associated with steroids versus saline also were also not statistically significant.

CONCLUSION
This randomized, double-blind, placebo-controlled trial has provided support for the hypothesis that transforaminal epidural etanercept is safe and effective (0.5-mg dose), with the potential to provide significant therapeutic relief for patients with persistent sciatica via a brief widely available nonoperative outpatient procedure.

This study was a phase IIa clinical trial. Such studies typically focus on type of patient, dose response, and the frequency of dosing as well as safety and efficacy. A signal of efficacy has been identified suggesting potential for transforaminal epidural etanercept in this patient cohort. This study provides valuable information regarding safety, patient population, dosing, effect size, and variability to assist with the design of larger phase III confirmatory trials.

Key Points

- In this RCT, two transforaminal injections of the TNFα inhibitor etanercept provided statistically and clinically significant reductions in mean daily worst leg pain compared to the placebo for patients with persistent sciatica.
- Transforaminal epidural etanercept (0.5 mg) provided 50-100% relief of pain for at least 3-6 months for more than one-half of subjects with persistent moderate to severe sciatica secondary to lumbar disc herniation.
- Clinically and statistically significant improvements were also noted in worst leg pain worst back pain, average back pain, and Oswestry disability index in etanercept-treated patients when compared to the placebo cohort.
- There was no difference in the incidence of adverse events in the etanercept cohorts when compared to the placebo cohort.
- Epidural etanercept could offer sciatica patients a safer and effective non-operative treatment.

References
HIZ’s relation to axial load and low back pain: investigated with axial loaded MRI and pressure controlled discography

Hebelka Hanna · Hansson Tommy

Abstract

Purpose The aims were to investigate if the detection of high-intensity zones (HIZ) is affected by axial load, and to study the correlation between HIZ and discogenic pain provoked with pressure controlled discography (PCD).

Methods 41 consecutive patients with chronic low back pain, referred for discography, were included. Each patient underwent PCD, CT, MRI, and axial loaded MRI (almRI) within 24 h. 35 patients completed all MRI sequences (140 discs). The detection of HIZ was compared between conventional MRI and almRI. PCD was performed in 119 of the discs examined at MRI. Provoked pain at PCD was classified into four categories (none/unfamiliar/similar/exact), with the patients’ daily pain as reference, and correlated with presence of HIZ.

Results AlMRI did not affect the detection of HIZ compared with conventional MRI. No significant correlation between HIZ and the 4-graded pain response at discography was found ($p = 0.34$), neither when combining similarly/exactly reproduced pain ($p = 0.08$). The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of HIZ in detecting discs with exactly reproduced pain were 49, 69, 39 and 76 %. When combining similarly/exactly reproduced pain, PPV was higher but still only 70 %.

Conclusions The detection of HIZ was not influenced by axial load. With strict PCD, discogenic pain can neither be confirmed when having HIZ (PPV 39 %) nor ruled out in discs without HIZ (NPV 76 %). Larger PCD studies including quantification of HIZ at conventional and almMRI are needed, before any dynamic component affected by axial load can be ruled out completely.

Keywords Discography · HIZ · Intradiscal pressure · Discogenic low back pain · Axial loaded MRI · Lumbar intravertebral disc

Introduction

High-intensity zones (HIZ), identified at T2-weighted MRI sequences within the posterior annulus [1], have been reported as reliable markers of discogenic pain [2, 3]. However, the diagnostic role of HIZ is limited due to low sensitivity [4–8]. The importance of HIZ in discogenic low back pain (LBP) remains therefore unclear. Pressure controlled discography (PCD), now criterion standard, has not been used in the majority of the numerous studies correlating HIZ and discogenic pain [1–7, 9–16] making previous results dubious, since discography without pressure registration can induce false positive responses [17, 18].

Histology of HIZ lesions, in discs believed to be painful, showed that they are composed of vascularised granulation tissue [2]. A dynamic component in the appearance of HIZ has been suggested by Alyas et al. [19], who reported that HIZ was unmasked by upright extension MRI in one patient with LBP and that HIZ appeared differently in another patient at the upright MRI compared with conventional...
sequences. The authors hypothesized that the load induced extrusion of fluid from the nucleus into the annular tear.

In order to investigate if HIZ constitutes dynamic tissue, affected by axial load, our primary aim was to compare if axial loaded MRI (alMRI) effects the detection of HIZ compared with conventional MRI. A secondary aim was to evaluate the relation between HIZ and discogenic pain with strict PCD.

Materials and methods

41 consecutive patients referred for preoperative lumbar discography were included between April 2007 and March 2010. They all had at least 6 months duration of LBP that had failed conservative therapy. Patients unable to undergo MRI or with allergies to contrast media were not eligible. With approval from the ethics committee, each patient underwent MRI, alMRI, discography and CT within 24 h. First, all were interviewed regarding the character and localisation of their symptoms. Then, MRI was performed with a 1.5 T equipment (Siemens Magnetom Symphony Maestro Class, Erlangen, Germany). T1 (TR 541 ms/TE 1 ms) and T2 (TR 4,000 ms/TE 124 ms) weighted sagittal images (4 mm/FoV 300 mm) were obtained and 4 mm T2 (TR 5,000–6,970 ms/TE 114–116 ms) axial sections were generated. AlMRI was performed with a non-magnetic compression device, and a harness according to Fig. 1 (DynaWell, Dynawell diagnostics AB, Las Vegas, Nevada USA) [20]. In addition to the sequences at the conventional MRI, 4 mm T1-weighted axial images (TR 500 ms/TE 15 ms) were added.

HIZ was evaluated by an experienced radiologist, blinded to the discography result, according to the criteria by Aprill and Bogduk [1]. PCD was performed, by one of two experienced radiologists, under standardized conditions. In all patients, discograms were performed of at least L4–L5 and L5–S1 and in addition 1–2 disc levels above these to include one negative control disc. Intravenous cefuroxime 1.5 g × 1 and midalozam 10 mg × 1 rectally was administered before the patients were placed for a prone posterolateral approach. 5–10 ml carbocain 10 mg/ml was administered locally before a 22-Gauge needle was introduced into each disc under fluoroscopy guidance. At the L5–S1 level, an 18-Gauge introduction needle was used. Contrast (Omnipaque, GE, Healthcare) was injected with a twist-manometer (Stryker Discmonitor®, Kalamazoo Michigan, USA), until one of the following endpoints was reached: pain intensity ≥5/10, pressure 100 pounds per square inch (psi), contrast volume 3.5 ml or a steady state in the pressure/volume curve (i.e., further pressurization impossible due to high resistance or epidural leak). 0.1–0.2 ml was injected at each twist, resulting in an injection speed of approximately 0.03 ml/s. Absolute intradiscal pressure was registered by the pressure limits including the intrinsic hydrostatic disc pressure. The patients were asked to relate any provoked pain in relation to their daily LBP and classify it into one of the four grades: none or pressure sensation, unfamiliar, similar and exact pain. During injection, the patient’s pain response was registered directly by the second radiologist. Discs were only graded as similarly or exactly reproduced, if the pain intensity was >5/10 on a numerical rating scale (0 = no pain and 10 = worst experienced pain). The patients were awake, alert and could correspond adequately when asked about their pain experience. The subsequent CT was performed within an hour after the contrast injection, and the annular disruptions was graded according to a modified Dallas discogram description (DDD) [21].

Statistics

Wilcoxon rank sign test was used to compare HIZ before and after axial load. Chi-square test, Fisher’s exact test and logistic regression analysis were used to analyze associations between HIZ and pain/annular disruptions. To correlate the grade of annular disruption with pain, contrast volumes and disc pressures, non-parametric Spearman test was used. The intra- and inter-observer agreement of HIZ was tested with kappa-coefficient. Kappa values >0.75 represent excellent agreement beyond chance, 0.60–0.74 good agreement, 0.40–0.59 moderate agreement and <0.40 poor agreement [22]. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated considering discography as a reference standard. A p value of <0.05 was considered statistically significant. The statistical analyses were performed with SPSS software.

Results

Of 41 patients, initially included 35 (19 female and 16 male, age 26–64 years, mean 46.9 years) completed all

Fig. 1 Compression device used at the alMRI. The alMRI was performed with a non-magnetic compression device in an out-stretched leg position, with an axial load of 50 % of the total body weight to simulate an upright position. The compression device is composed of a patient harness attached to a footplate with adjustable side straps. The external load was adjusted by loosening or tightening the straps.
required MRI sequences. The reasons for incomplete examinations were four cases handling errors with incomplete sequences, and in the remaining two patients due to motion artifacts. In the remaining 35 patients, 140 discs were examined with both conventional and alMRI. No significant difference in the detection of HIZ before and during axial load was found. 48 discs (34%) displayed HIZ at MRI and 49 (35%) at alMRI. However, in three discs, HIZ appeared at axial loaded sequences (two at L4–L5, one at L5–S1), whereas at four levels HIZ was detected before but not after axial load (two at L3–L4, two at L5–S1). Kappa value for intra- and inter-observer agreement of the 164 discs at the conventional MRI was 0.87 and 0.84, respectively. Intra- and inter-observer agreement of HIZ at the 140 discs at the alMRI was 0.82 and 0.80, respectively.

Of the 164 discs examined with conventional MRI, PCD was performed at 124 discs. Due to anatomic unfavorable positioning of 5 L5–S1 discs, 119 discs were successfully injected. The distribution of disc levels injected and their pain response is shown in Table 1. Maximal pressure, total contrast volume and both pressure and contrast volume at the initial evoked pain are displayed in Table 2. Since the absolute pressure was used in this study, the opening pressure (i.e., the intrinsic hydrostatic pressure) was estimated to 14 psi, the mean opening pressure found by Derby et al. [23] when using manometry. Mean maximal pressure in the current study was 43 psi, subtracting the estimated opening pressure results in a mean maximum pressure of 29 psi above opening pressure in the current study. Low pressure positive disco grams, i.e., positive at ≤15 psi above opening pressure [23] were 14% for exactly reproduced pain, and 23% for similar/exact pain in combination. Seven patients had exact pain at two or more discs.

Statistically significant associations were found neither between HIZ and discogenic pain (4 grades) (p = 0.34), nor between HIZ and similar/exact pain in combination

<table>
<thead>
<tr>
<th>Disc</th>
<th>None</th>
<th>Unfamiliar</th>
<th>Similar</th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2–L3</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>L3–L4</td>
<td>10</td>
<td>15</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>L4–L5</td>
<td>3</td>
<td>5</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>L5–S1</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>28</td>
<td>36</td>
<td>35</td>
</tr>
</tbody>
</table>

The positive rate of discograms calculated per disc was 29% (35/119) for exact provoked pain and 60% (71/119) when combining similar/exact pain. The pain provoked at discography is correlated with the patients’ daily pain.

Table 2 Maximal pressure, total contrast volume and both pressure and contrast volume at initial evoked pain at discography

<table>
<thead>
<tr>
<th>Total contrast volume</th>
<th>Disc N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume at initial evoked pain</td>
<td>107</td>
<td>0.6</td>
<td>3.5</td>
<td>1.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Maximal pressure</td>
<td>96</td>
<td>0.1</td>
<td>2.5</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>Pressure at initial evoked pain</td>
<td>119</td>
<td>10</td>
<td>127</td>
<td>43</td>
<td>25</td>
</tr>
<tr>
<td>Pressure at initial evoked pain</td>
<td>94</td>
<td>3</td>
<td>72</td>
<td>24</td>
<td>16</td>
</tr>
</tbody>
</table>

Due to technical and human factors (invaluable figures, when doubt existed if pain was evoked etc.) all parameters were not registered in every disc. Volume is given in ml and pressure in pounds per square inch (psi).

SD Standard deviation

Discussion

The principal findings of this study were that the detection of HIZ did not change between conventional and alMRI, and that HIZ is a poor predictor of discogenic pain provoked by strict PCD. We hypothesized that alMRI, possibly by forcing nuclear material into the annular tear or just by the pressure elevation, may change the flow in the vascularised HIZ tissue, and consequently change its detection. The present study did not reveal any significant change in the appearance of HIZ due to axial load. We found, however a small discrepancy in HIZ between unloaded and loaded position in some of the more than 40 discs presenting with HIZ (Fig. 2).

With a 1.5 T the signal-to-noise ratio ought to be high enough for detecting HIZ, but as Peng et al. [2] pointed out HIZ might be undetected even in 3–4 mm sagital images if the granulation tissue is too small or less vascularised.

According to Aprill and Bogduk's [1] definition HIZ must be as intense as the adjacent CSF, suggesting that less
bright zones, “low intensity zones”, represents less inflamed tears. Carragee et al. [5] used quantitative digital signal analysis in the center of HIZ and classified it into “high-intensity” (HIZ) and “medium-intensity” (MIZ) zones, where MIZ might be more inactive tears. Further O’Neill et al. [8] classified HIZ into mild, moderate and marked hyper intense zones, also showing that there exist a spectra of intensity variations within HIZ. HIZ was not quantified in the current study which is a limitation. However in consistency with the case report by Alyas et al. [19], the impression was that HIZ often existed at unloaded and loaded MRI examination but varied in intensity/size/shape.

AlMRI increases the lordosis [20], and alters the position of nucleus pulposus [25]. These alterations might be another explanation to the discrepancy of HIZ appearance between the examinations in one and the same patient. AlMRI is developed to simulate the loading conditions of the lumbar spine in standing position particularly to the L3–L4-level, where usually the peak of the lumbar lordosis is found. Depending on at what lumbar level the tears are located, their granulation tissue might be more or less prone to react to the changes in the intradiscal pressure, if the peak stress is at the anterior or posterior annulus. At some discs, the axial load might result in a disc pressure that exceeds the intravascular pressure of the vascularised granulation tissue in HIZ, obstructing the vascular supply of the latter. Theoretically this could lead to a reduced signal in the posterior annulus.

Despite above discussed topics, the most probable explanations to the discrepancy in HIZ between the MRI examinations in the current study are partial volume effect and the observers’ validation of HIZ, deciding if zones are as bright as CSF or less. A more detailed investigation is needed to clarify if axial load affects HIZ in terms of intensity/size/shape.

Significant association between HIZ and discogenic pain was not found in contrary to many previous studies [1, 6, 9, 10, 15, 26]. Granulation tissue is likely to change over time rendering this study an edge, since MRI and discography were performed within 24 h. The time interval between the two procedures varied between 4 weeks [5] and 5 months [13] in the previous studies correlating HIZ and discography findings, and could be a possible explanation of the discrepancy with this study. The PPV of HIZ in terms of exactly reproduced pain was only 39 % compared to, between 83 and 95 % [1, 6, 9–11] in studies favoring HIZ as a reliable marker of discogenic pain. Among numerous authors correlating HIZ and discogenic pain [1, 2, 4–13, 15, 26], only Carragee et al. [5] and O’Neill et al. [8] used PCD. The lack of controlled pressure injection can be an important reason to the discrepancy in PPV between the current and previous reports. Uncontrolled high intradiscal pressures and injection speeds have been shown to generate false positive responses [17, 18, 27]. It is likely that the lack of PCD was a strong confounding factor to many of those previous results. The use of 4-graded pain scale instead of three or less [2, 6, 7, 9–11, 15, 21] may also have

Table 3  Pain provoked at discography in relation to HIZ

<table>
<thead>
<tr>
<th>Provoked pain at discography</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Unfamiliar</td>
</tr>
<tr>
<td>HIZ Absent</td>
<td>18</td>
</tr>
<tr>
<td>Present</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
</tr>
</tbody>
</table>

Discography was considered reference standard and a discogram positive when pain was either exactly provoked or similarly/exactly provoked. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of HIZ in detecting painful discograms were 49 % (17/35), 69 % (57/84), 39 % (17/44) and 76 % (57/75) for exactly reproduced pain. Corresponding figures for similar/exact pain were 44 % (31/71), 73 % (35/48), 70 % (31/44) and 47 % (35/75).

Fig. 2  T2-weighted sagital images of one patient at conventional MRI and at axial loaded MRI  
(a) Conventional MRI, (b) Axially loaded MRI. There is a discrepancy in HIZ between the conventional and axially loaded MRI examination. At the conventional MRI HIZ was observed at L3–L4 and L5–S1 whereas when applying axial load HIZ was not shown at L5–S1. Note the increased lordosis at the alMRI
influenced our results and that is why we also combined similar/exact pain in our calculations, by doing so PPV became higher though still only 70%.

As previously shown [2, 21], significant correlation was found between annular disruption and both HIZ and symptomatic discograms. In this study, 97% of the discs with exactly reproduced pain exhibited grade 3 or higher annular disruptions, indicating that annular tears are important in generating discogenic pain.

Limitations
The general lower pressurization in the current study might have resulted in some false negative discograms. However, since intradiscal pressure during discography potentially is transferred to adjacent discs [28] and consequently might result in false positive diagnosis when pressurizing discs too much, a lower pressurization was used. O’Neill et al. [17] also favors a lower pressurization when reporting 50% false positive discograms at 25 psi above opening pressure. Further, Derby et al. [18] stated that mean pressure initially evoking pain in positive discograms was 20 psi, also supporting a lower pressure range. The rate of low pressure positive discograms in the current study was in accordance with Derby et al. [23], who showed a rate of 15% when using manometry. The limitations of not quantifying HIZ have been discussed above. Further limitations are the relatively small number of patients and that an asymptomatic control cohort was not used to evaluate any change in HIZ between the MRI examinations.

Conclusions
There was no significant change in appearance of HIZ between conventional and axial loaded MRI, indicating that the appearance of HIZ is not in any distinct way related to the loading of the spine. HIZ is a poor predictor of discogenic pain using PCD. Discogenic pain can neither be confirmed when having HIZ, nor ruled out in discs lacking the sign. Larger PCD studies, including quantification (size, shape, intensity) of HIZ without and with axial load as well as a likewise study on asymptomatic individuals are needed before any dynamic component affected by axial load completely can be ruled out.

Acknowledgments The study was supported by funds from University of Gothenburg (ALF) and Gothenburg Medical Society.

Conflict of interest None.

Ethical standard Ethical approval was granted by the Ethics Committee of Gothenburg, Sweden.
Dose response and structural injury in the disability of spinal injury

Mohammed Shakil Patel · Philip Sell

Abstract

Introduction In traumatic injury there is a clear relationship between the dose of energy involved, structural tissue damage and resultant disability after recovery. This relationship is often absent in cases of non-specific chronic low back pain that is perceived by patients as attributed to a workplace injury. There are many studies assessing risk factors for non-specific low back pain. However, studies addressing causality of back pain are deficient.

Purpose To establish whether there exists a causal relationship between structural injury, low back pain and spinal disability.

Methods Retrospective analysis of prospectively gathered validated spinal outcome measures [Oswestry disability index (ODI), low back outcome score (LBO), modified somatic perception (MSP), modified Zung depression index (MZD)] between patients with healed high energy thoracolumbar spinal fractures and patients with self-perceived work-related low back pain. Causality was established according to two of Bradford Hill’s criteria of medical causality, temporal and dose–response relationships.

Results Twenty-three patients with spinal fractures (group 1) of average age 44 years were compared to 19 patients with self-reported back pain in the workplace pursuing claims for compensation (group 2) of average age 48 years. Both groups were comparable in terms of age and sex. The average ODI in group 1 was 28 % (SD 19) compared to 42 % (SD 19) in group 2 ($P < 0.05$). Similarly, LBOS was 39.7 versus 24.3 ($P < 0.05$), MSP 4.3 versus 9.3 ($P < 0.05$) and MZD 20.2 versus 34.8 ($P < 0.05$) in groups 1 and 2, respectively.

Conclusion Despite high-energy trauma and significant structural damage to the spine, patients with the high energy injuries had better spinal outcome scores in all measures. There is no ‘dose–response’ relationship between structural injury, low back pain and spinal disability. This is the reverse of what would be anticipated if structural injury was the cause of disability in workplace reported onset of low back pain.

Keywords Non-specific low back pain · Vertebral fractures · Outcomes · Work-related injury · Dose–response relationship

Introduction

The lifetime incidence of back pain in western countries is 70 % [1, 2]. This is often associated with occupational injuries [3] with 52–60 % being reported as work-related [2] and 47–77 % of such patients seeking compensation [4]. Each year in the UK 120 million working days are lost because of back pain [5]. The socio-economic burden of LBP is therefore significant and increasing with the indirect costs of this loss of productivity thus far exceeding the substantial medical costs involved [6].

The majority of cases of non-specific low back pain (NSLBP) are self-limiting. The prognosis for most patients on sick leave to resume work is good [7] with 70 % of patients on sick leave returning to work within 1 week, and 90 % within 2 months [1]. The complexity of the spine
often renders it difficult to identify the specific causes or trigger points in patients presenting with low back pain. 85% of patients with back pain have unexplainable symptoms [8]. The aetiology of NSLBP is therefore considered as multifactorial [9], with clear morphological alterations only being found in 10–20% of cases [8].

The percentages of patients with acute LBP that go on to a chronic state varies between studies from 2 to 34% [10]. Why some individuals with NSLBP progress to develop disabling chronic pain in the absence of any significant structural injury still remains unknown with psychosocial rather than medical factors being held responsible [11, 12].

Medicalisation [13], sick leave, work absence [14] and a claim for financial compensation [15] have all been shown to be strongly linked with poor prognosis in patients struggling with NSLBP.

Establishing a causal relationship in patients with back pain is therefore challenging. Subgroups of patients with low velocity injury and self-reported LBP in the workplace have poor functional outcomes in long-term studies. The opposite is true in patients with significant high-energy spinal fractures. These groups of patients have been shown to have satisfactory long-term outcomes in terms of both pain and functional improvement with between 60 and 80% of patients returning to work [16–19].

The majority of existing studies relating to non-specific low back pain identify an association between potential risk factors and outcomes. However, such associations do not infer a causal relationship [20]. Establishing causality is important particularly with regards to work-related injury and disability from low back pain. Patients with work-related injuries are more likely to commence litigation [21]. However, such injuries may not be a causative factor for back pain and related disability, but may be associated with psychosocial factors. Why should industry and economy suffer from non-causal back pain when in fact such perceived injuries may not be responsible for patient disability?

The purpose of this study was to establish whether there exists a relationship between structural injury, low back pain and spinal disability according to Sir Bradford Hill’s criteria of medical causality [22].

**Methodology**

To address this causality, a retrospective analysis of prospectively gathered spinal outcome measures consisting of the Oswestry disability index (ODI), low back outcome score (LBO), modified somatic perception (MSP) and modified Zung depression index (MZD) was compared in two contrasting cohorts of patients at the extremes of an energy spectrum. Contrasting cohorts were selected to highlight and amplify any definite structural injury effect on disability.

Group 1 consisted of 23 consecutive patients admitted to a university hospital between 1996 and 2009 under the care of the senior author for high energy isolated unstable thoracolumbar and lumbar spinal fractures which were treated with surgical stabilisation. They were then followed-up until satisfactory clinical and radiographic fusion and completed standard spinal assessment questionnaires relating to spinal disability at final review. This was taken as the time of removal of metalwork as was standard practice at that time. Polytrauma and conservatively managed patients as well as those who failed to return for follow-up were all excluded.

Group 2 consisted of 19 consecutive patients under the care of the senior author who were recruited from a medico-legal clinic. All of these patients were presenting for medico-legal compensation claims relating to self-reported low back pain perceived to be due to a workplace injury over the same time frame. Patients with neck pain, workplace accident, road traffic accidents or other physical injuries were excluded. The energy involved in such cases was negligible or difficult to quantify. A thorough history and examination was performed along with patient completion of validated standard spinal assessment questionnaires.

Statistical analysis was performed using the Mann–Whitney U test using SPSS version 16. Causality was assessed using Sir Bradford Hill’s criteria for medical causality [22]. He proposed a minimal set of nine conditions required to establish a causal relationship. For the purpose of this study two of these criteria were used. A temporal relationship, that is, did the cause of back pain precede the effect of back pain and spinal disability; secondly as the dose of energy involved in the traumatic event increased, did the resulting disability also increase, that is a dose–response relationship.

**Results**

Patient demographics are presented in Table 1.

Patients in group 1 were assessed following satisfactory clinical and radiographic union of isolated surgically stabilised thoracolumbar spinal fractures. The average time from surgical stabilisation and removal of metalwork was 41 months. The majority of fractures were at the level of the first lumbar vertebrae (Fig. 1).

All 19 patients in group 2 related their symptoms to be work-induced injuries as a result of manual handling whilst performing their regular duties. All of these patients had manual handling training and all complained of predominant low back pain. 47% of patients within this group had
back pain prior to the alleged incident. All but one patient within this group were not undertaking any form of physical exercise and all remained on a cocktail of analgesia. 47% of patients within this group remained in some form of employment with the remainder being unemployed as a result of their symptoms. All patients within this group had obstacles to recovery or yellow flags in addition to that of their compensation status and all had completely normal neurological examinations. On national guidelines, none of these patients warranted any form of specialised imaging but only 3 (16%) patients were not investigated with specialised imaging techniques. The remainder had either X-rays of their lumbar spine ± MR ($n = 16$) imaging. All of these had been organised by physicians prior to review. The majority of these showed normal age related changes and all failed to identify focal pathology that may have correlated with patient symptoms.

The average ODI for patients in group 1 was 28% (SD 18.5), (moderate disability) compared to 42% (SD 18.6) in group 2, (severe disability) ($P = 0.031$). The average LBO was 39.7 (SD 16.1) in patients in group 1 representing fair back function versus 24.3 (SD 9.9) in patients in group 2, indicating poor function, ($P = 0.008$). The modified somatic perception (MSP) was 4.3 (SD 4.3) versus 9.3 (SD 6.3), ($P = 0.027$); and MZD was 20.2 (10.9) versus 34.8 (8.4) ($P = 0.001$) in groups 1 and 2, respectively. Patients with self-reported back pain secondary to a self-perceived injury in the workplace were significantly more disabled; had poor levels of back function; were more somatized and had higher levels of depression when compared to those with thoracolumbar spinal fractures. The average visual analogue score for leg and back pain in group 2 was 3.7 and 5.7, respectively, ($P = 0.016$) indicating that this group of patients were back pain dominant. Visual analogues scores for the spine fracture cohort were not recorded and therefore no comparisons could be made relating to these (Table 2; Fig. 2).

A subgroup analysis was performed comparing the outcomes between patients with and without previous back pain within group 2. Patients with previous back pain had lower levels of back function and greater scores on the MZD index (Table 3). Any further constructive analysis of this was not plausible due to very small numbers. Statistical analysis was therefore not performed.

**Discussion**

Disability associated with work-related LBP is an increasingly serious societal problem. Although most injured workers return quickly to work, a substantial number do not [23]. Significant proportions of such claimants do not have any verifiable evidence of significant injury and are therefore self-reported episodes of back pain occurring in the workplace.

Identifying the cause of low back pain secondary to workplace accidents is important and often overlooked in the literature. Structural causes which can be identified make up to only 20% of back pain cases in total. In the remainder no cause is often found. Many publications assess risk factors for chronicity and continuing work absence. These are primarily psychosocial. Although these are important in enabling and directing intervention thereby preventing or reducing the chronicity of patient symptoms and ultimately societal costs, they fail to address the issue of causality. Is non-specific work-related injury responsible for chronic low back pain?

The answer to this question is important and one which many medical and legal experts try to address on a regular basis within our compensation and litigation-orientated culture. If back pain is caused by such work-related injury then a claim for compensation may be valid and institutions should be made to address such causative factors. However, if not, then awareness of this needs to be raised to prevent opportunistic claims.

Sir Bradford Hill, a medical statistician, proposed a minimal set of nine conditions which are required to establish a causal relationship. The purpose of this study was to assess one of these criteria, the dose–response relationship. As the dose increases in magnitude the cause or effect should also increase in similar magnitude. The

**Table 1** Patient demographics [mean/SD (range)]

<table>
<thead>
<tr>
<th>Groups</th>
<th>No. of patients</th>
<th>Age (range)</th>
<th>Sex</th>
<th>Final follow-up (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thoracolumbar spinal fractures</td>
<td>23</td>
<td>44 (26–66)</td>
<td>8F; 15M</td>
<td>41 (14–89)</td>
</tr>
<tr>
<td>Workplace injury</td>
<td>19</td>
<td>48 (30–64)</td>
<td>4F; 15M</td>
<td>38 (18–53)</td>
</tr>
</tbody>
</table>

*a Average time between onset of symptoms and review in clinic (workplace injury); fracture fixation and removal of metalwork (thoracolumbar fractures)*
The presence of this relationship is strong evidence for a causal relationship. This is often of a threshold relationship manifesting itself as a sigmoid-shaped curve (Fig. 3).

Patients in this study with the high energy spinal fractures had less disability and better back function when compared to the patients with work-related injuries. There is therefore no dose–response relationship. The findings of this study also correlate with that by Giannoudis et al. [24] in which they conclude the absence of a dose–response relationship between the magnitude of trauma severity and the incidence of whiplash injury.

47 % of patients in group 2 had back pain even prior to the onset of their perceived work-related insult. This is quite interesting as this defies causality. The only essential criterion for causality according to Bradford Hill is the presence of a temporal relationship, the cause must precede the effect. However, almost half of the patients in this group already had symptoms prior to the perceived injury. There is therefore also an absence of a temporal relationship and again further refutes the presence of a causal relationship between non-specific back pain and self-perceived work-related injury.

There are many studies in literature assessing the long-term outcomes in patients with work-related injury and spinal fractures but none of these have compared the long-term outcomes between these 2 groups of patients. This comparative contrasting cohort study is therefore unique. Although both groups of patients were comparable in terms of age and sex, the employment and compensation status as well as the visual analogue score for back pain in patients with thoracolumbar fractures was not known. This is due to the limitations of a retrospective study. However, previous studies on patients with thoracolumbar spinal fractures have shown between 50 and 80 % of these patients returning to full-time employment with up to 64 % returning to their previous level of employment and similar numbers reporting minimal or no pain [16–18]. Furthermore, if there were some patients within the fracture group that had poor outcomes due to psychosocial factors such as compensation or failure of return to employment this would further emphasise the clinically significant difference between structural and non-structural injury.

The study also comprised a very selective group of patients to reduce heterogeneity within groups. The purpose was to assess for the presence of a dose–response relationship and consequently, causality. The group of patients used thereby allowed for the comparison at the two extremes of an energy spectrum. Patients with self-perceived work-related injury were recruited from a medicolegal clinic and all were pursuing compensation claims. All patients with actual accidents at work were excluded. The dose of energy involved in this group was minimal and unquantifiable. This may for obvious reasons confound the results as such claims have been shown to result in greater levels of disability. Nevertheless, the highly selective group of patients were only available through a medicolegal clinic setting.

Furthermore, only patients with surgical treatment of their isolated thoracolumbar spinal fractures were included. Conservatively managed patients were excluded. Again the

<table>
<thead>
<tr>
<th>Table 2 Summary of spinal outcome measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Group 1 (Fracture)</td>
</tr>
<tr>
<td>Group 2 (NSLBP)</td>
</tr>
<tr>
<td>P value</td>
</tr>
</tbody>
</table>

| Table 3 Subgroup analysis of patients with and without previous back pain (mean/SD) |
|------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                  | Age         | VAS L       | VAS B       | ODI         | LBO         | MSP         | MZD         |
| Previous back pain | 47/11      | 4.1/1.6     | 5.9/2.0     | 45.6/13.1   | 20/12.8     | 8.7/5.3     | 39.5/8.9    |
| No previous back pain | 50/8.9     | 4.3/2.3     | 6.1/2.1     | 40.6/17.4   | 27/8.6      | 10/7.6      | 30/4.8      |
The purpose of this selection bias was to include only those patients with the higher energy injuries at the other extreme of the injury spectrum.

**Conclusion**

Despite high energy trauma and significant structural damage to the spine, those with high energy injury and thoracolumbar spinal fractures had better spinal outcome scores in all measures (ODI, LBOS, MSP, MZD). We have failed to identify a temporal or a dose–response relationship between structural injury and spinal disability. Uniquely, the disability is greater in the lower energy injury which is unique in trauma care. The reasons for such differences are primarily psychosocial. Self-perceived work-related injury is not a cause of low back pain. There is no causal relationship between structural injury, low back pain and spinal disability (Table 4; Fig. 3).

**Conflict of interest** None.

**References**


Incidence and risk factors for early adjacent vertebral fractures after balloon kyphoplasty for osteoporotic fractures: analysis of the SWISSspine registry

Christian Spross • Emin Aghayev • Rouven Kocher • Christoph Röder • Thomas Forster • Fabrice A. Kuelling

Abstract
Purpose The SWISSspine registry (SSR) was launched in 2005 to assess the safety and effectiveness of balloon kyphoplasty (BKP). In the meantime, repeated reports on high rates of adjacent vertebral fractures (ASF) after BKP of vertebral insufficiency fractures were published. The causes for ASF and their risk factors are still under debate. The purpose of this study was to report the incidence and potential risk factors of ASF within the SSR dataset.

Methods The SSR data points are collected perioperatively and during follow-ups, with surgeon- and patient-based information. All patients documented with a monosegmental osteoporotic vertebral insufficiency fracture between March 2005 and May 2012 were included in the study. The incidence of ASF, significant associations with co-variates (patient age, gender, fracture location, cement volume, preoperative segmental kyphosis, extent of kyphosis correction, and individual co-morbidities) and influence on quality of life (EQ-5D) and back pain (VAS) were analyzed.

Results A total of 375 patients with a mean follow-up of 3.6 months was included. ASF were found in 9.9 % (n = 37) and occurred on average 2.8 months postoperatively. Preoperative segmental kyphosis >30° (p = 0.026), and rheumatoid arthritis (p = 0.038) and cardiovascular disease (p = 0.047) were significantly associated with ASF. Furthermore, patients with ASF had significantly higher back pain at the final follow-up (p = 0.001). No further significant associations between the studied co-variates and ASF were seen in the adjusted analysis.

Conclusions The findings suggest that patients with a preoperative segmental kyphosis >30° or patients with co-morbidities like rheumatoid arthritis and a cardiovascular disease are at high risk of ASF within 6 months after the index surgery. In case of an ASF event, back pain levels are significantly increased.

Level of evidence IV.

Keywords Balloon kyphoplasty • Adjacent segment fracture • SWISSspine registry • Risk factors • Rheumatoid arthritis

Introduction

Vertebral compression fractures (VCF) are an increasingly prevalent disease in the aging population and may cause pain, limited physical function, decreased mobility, reduced quality of life (QoL) and increased mortality as well as a resulting heavy economic burden on society [1–4]. Technical advances in minimal invasive spine surgery have offered new treatment possibilities.

Balloon kyphoplasty (BKP) of VCF was first described by Garfin et al. [4]. With the transpedicular insertion of two inflatable balloons into the fractured vertebral body, a void can be created which is subsequently filled with polymethyl methacrylate bone cement (PMMA) to realign and stabilize the fractured and often kyphotic vertebra (kyphoplasty).

The Swiss Federal Office of Health mandated a nationwide health technology assessment registry for BKP to assess its benefits and to determine the reimbursement of this procedure. Therefore, the SWISSspine registry (SSR)
was launched in March 2005. In 2010, first results of the registry were published [5] indicating that BKP is a safe and effective procedure in the treatment of vertebral fractures with a significant and clinically relevant reduction of back pain as well as improvement of QoL. This was in accordance with previous studies [6–8]. SSR data also showed that within the first postoperative year 20 % of all patients sustained new fractures, 72 % adjacent to the cemented levels [5]. The adjacent segment fracture (ASF) seems to be a non-negligible complication of BKP in osteoporotic vertebral fractures. This requires further investigation, especially since the incidence is still controversial [9–14], and potential risk factors are not well studied.

The aim of our study was hence to analyze the early ASF after kyphoplasty for osteoporotic fractures within the SSR.

Materials and methods

The registry

SSR data are collected perioperatively and during follow-ups. Main follow-up intervals are at 3 and 6 months, 1 year and annually thereafter. The documentation forms and outcome instruments are: (1) primary intervention form (surgeon-administered), (2) follow-up form (surgeon-administered), (3) comorbidity questionnaire (patient assessment), (4) EQ-5D questionnaire (patient assessment), (5) North American Spine Society outcome instrument (NASS) (patient assessment) between 2005 and 2009. In 2009, the evaluation obligation was released and the NASS patient assessment was replaced by the Core Outcome Measures Index (COMI Back). The main reason to change to the COMI instrument was a better comparability of SSR BKP results with those of the international Spine Tango registry [15, 16]. Both outcome instruments, NASS and COMI, have a visual analog scale (VAS 0–10) for pain assessment, which was used as one of the outcome measures in the study.

The EQ-5D is one of the most frequently used generic QoL instruments in health economic and outcome studies today. The instrument includes a global visual analog scale (VAS), anchored in worst and best imaginable health status, and the EQ-5D descriptive system which is subdivided into five dimensions; (1) mobility, (2) self-care, (3) usual activities, (4) pain/discomfort and (5) anxiety/depression [17]. The EQ-5D score ranges from −0.6 to 1, where 1 is the best imaginable health and 0 represents a state of death.

All patients sign an informed consent form agreeing with data collection in the registry and regular follow-ups. Patient questionnaires are completed before surgery and at all follow-ups, except for the preoperative comorbidity form. Surgeons complete the primary form at the time of surgery and the follow-up form at each follow-up.

Sample characteristics

From March 2005 to May 2012, 817 surgeries and 1,797 follow-ups were documented. The Swiss Society for Spinal Surgery recommends the following criteria for BKP: (a) the VCF is responsible for the symptoms, (b) the pain level on VAS is persistently above five points, (c) the segmental kyphosis for thoracic vertebrae needs to be at least 15°, (d) the segmental kyphosis for lumbar vertebrae needs to be at least 10°, (e) height reduction of the fractured vertebra is larger than 1/3 of the height of the adjacent vertebrae.

Patients with an osteoporotic vertebral fracture and at least one follow-up within the first 6 postoperative months were included in the study. To simplify the interpretation of our study results, patients with more than one vertebral body fracture were excluded. The in- and exclusion criteria resulted in a sample of 375 patients (91 males and 284 females) with an average age of 72.7 years for females and 73.1 years for males. A total of 44 single surgeons were involved in the treatment of these 375 patients (8–9 patients per surgeon on average; range 1–44 cases). Detailed demographic characteristics of the sample are shown in Table 1 and co-morbidities in Table 2.

Statistical analysis

Patients with and without ASF were compared regarding their demographic and clinical characteristics. Wilcoxon signed-rank test was used for comparisons of continuous variables such as pain on VAS between patients with and without ASF. When comparing proportions, the Chi-square test was used.

A multivariate logistic regression model was built to search for significant associations for ASF after BKP treatment. The binary outcome was the occurrence of ASF (yes/no). Co-variates considered in the analysis were: patient age, gender, fracture location (Th4–Th11, Th12, L1, L2–L5), volume of injected cement (<4.5 ml, ≥4.5 ml), preoperative segmental kyphosis (none, 10–20°, 21°–30°, >30°), and extent of achieved correction of segmental kyphosis immediately after surgery (worsening, no change, 1 category, 2 categories, ≥2 categories improvement—see below).

The registry contains information on segmental kyphosis. The preoperative categories are: none, 10°–15°, 16°–20°, 21°–25°, 26°–30° and >30°. Postoperatively, the categories are the same, except for an extra category of 5°–10°. For the adjustment in the multivariate regression models the four possible changes of the postoperative segmental kyphosis category relatively to the perioperative were used: worsening of kyphosis category, no change, 1
category improvement, 2 categories improvement, ≥2 categories improvement. The extra category (5°–10°) together with the category “none” on the follow-up form was thereby equalized to the category “none” on the surgery form.

\(\alpha\) was set to 0.05 throughout the study. All statistical analyses were conducted using SAS 9.3 (SAS Institute, Inc., Cary, NC, USA).

**Ethics**

Ethical approval was not needed as anonymized data were collected and evaluated within a mandated registry based on informed written patient consent.
Results

A total of 37 patients (9.9 %) showed an ASF, in 72 % the cranial and in 28 % the caudal adjacent vertebra was involved. The average follow-up was 3.6 months (range 0.8–6.6 months). The adjacent new fracture was detected on average 2.8 months (range 1.2–6.6 months) postoperatively.

Univariate analysis

The non-adjusted comparisons between the patients with and without ASF as shown in Table 1 revealed significant difference regarding preoperative segmental kyphosis \((p < 0.001)\). Furthermore, comparison of back pain characteristics demonstrated that patients with and without ASF have significantly different postoperative back pain \((p = 0.001)\). The latter was higher in patients with ASF. The average back pain relief, however, was not significantly different between the patient groups, differing by only 6 VAS points on the 100 points scale. Regarding EQ-5D score characteristics, no significant differences were observed.

The proportions of individual co-morbidities for patients with and without ASF were compared in Table 2. Cardiovascular disease and rheumatoid arthritis were significantly more frequently observed in patients with ASF.

Multivariate analysis

The multivariate logistic regression revealed preoperative segmental kyphosis to be significantly associated with an early postoperative ASF \((p = 0.026)\). According to the model, patients with preoperative segmental kyphosis higher than \(30^\circ\) had an 8.36-times (95 % CI 1.61–43.5) higher likelihood for an ASF than those without segmental kyphosis (“no segmental kyphosis preoperative”). In addition, rheumatoid arthritis (OR 2.96, 95 % CI 1.07–8.21; \(p = 0.038\)) and cardiovascular disease (OR 2.66, 95 % CI 1.01–7.0; \(p = 0.047\)) as comorbidity were associated with higher likelihood for an ASF. Patient age \((p = 0.24)\), gender \((p = 0.78)\), fracture location \((p = 0.70)\), volume of injected cement \((p = 0.55)\) and extent of achieved correction of segmental kyphosis \((p = 0.53)\) were non-significant co-variates.

Discussion

This study presents the analysis of the so far largest cohort of patients treated with BKP, focusing specifically on early incidence of adjacent segmental fractures and their association with surgical and patient characteristics. In this analysis, a collective of patients with osteoporotic fractures and short-term follow-up up to 6 months was considered. Prior studies reported ASF after BKP to mainly occur within the first 3 months postoperatively and most often in patients with osteoporosis as cause of the primary vertebral fracture [12, 18–20]. In addition to osteoporosis as fracture etiology, we restricted the analysis to patients with a single level fracture for creating a uniform cohort and simplifying interpretation of results. In our sample nearly each tenth patient developed an early ASF. A preoperative segmental kyphosis \(>30^\circ\) was significantly associated with the occurrence of ASF both in the univariate and multivariate analyses. We could demonstrate that patients with ASF had a significantly higher risk of persistent back pain in the univariate model.

In the literature, ASF rates after BKP vary between 6.5 and 25 % [12, 19–22]. Our observed rate of nearly 10 % is well within this reported range. The higher rates of adjacent fractures in other studies may be explained by inclusion of patients treated on multiple segments and having long-term follow-ups [12, 18, 22]. The cumulative
incidence of new vertebral fractures at any spine level within 1 year after an incident vertebral fracture has been found to be 6.6% in postmenopausal women [13]. This percentage seems to be well comparable to the 10% seen in the BKP treated cohort.

The natural cause of ASF after insufficiency fractures is still unknown. Evidence exists that rather the primary disease (osteoarthritis) and secondary malalignment of the vertebral column may be the reason for further fractures [11, 13, 23, 24] rather than the cement augmentation itself. Primary and secondary osteoporosis [18, 19, 25] are already identified risk factors for ASF after BKP [19, 25]. Our results, in terms of significant association of the segmental kyphosis prior to surgery for the occurrence of ASF, also support this thesis. However, the extent of achieved correction of segmental kyphosis did not result in a significant association with the incidence of ASF.

Both univariate and multivariate analyses showed a significant association of cardiovascular disease and rheumatoid arthritis with ASF. Those associations were not previously reported and it is difficult to explain them. One may speculate that advanced cardiovascular disease, rheumatoid arthritis or their combination (with or without other co-morbidities) may be associated with less mobility and, therefore, a potentially higher degree of osteoporosis. Furthermore, in patients with rheumatoid arthritis, an involvement and weakening of the adjacent discovevertebral joints and steroid treatments may contribute to the higher rate of ASF. However, those are hypotheses that require further studies.

The extent of achieved correction of segmental kyphosis was not significant in the statistical model. In our opinion, the amount of preoperative kyphosis may reflect the severity of the underlying disease, i.e., osteoporosis. Another possible explanation could be the residual kyphotic deformity due to surgical difficulties in fully realigning such a high preoperative kyphosis angle. The sagittal malalignment due to an increased segmental kyphosis may lead to potentially higher load through the anterior column and consequent higher incidence of ASF [26, 27]. Movrin et al. [19] reported a decreased risk of ASF in patients with a kyphotic angle lower than 9° postoperatively in comparison with kyphosis higher than 9°.

Weaknesses and strengths

Some limitations of the present study require a further explanation. A potential under-reporting of surgeon-based outcomes like ASF cannot be completely excluded in this observational unmonitored study. All surgeons participating in the documentation are accredited by the Swiss Spine Society for BKP procedures. To get accredited each surgeon has to participate in an advanced training including a cadaver lab. Furthermore, they need an adequate infrastructure to conduct the procedure in their hospital. An audit or any other control mechanism in a national registry would need strong financial and organizational resources and was considered as not feasible by the stakeholders of the project.

No direct treatment comparator is included in the SSR documentation. A complete documentation of a comparator procedure like non-surgical treatment in a national registry would need additional and substantial administrative and financial efforts, which were considered even less feasible. Therefore, it is impossible to distinguish between ASF due to BKP and adjacent “natural” vertebral insufficiency fractures in this study. Controlled studies of non-surgical treatment versus BKP are needed.

The mean follow-up of 3.6 months in our study is relatively short and the incidence of ASF might even be higher with longer follow-up. However, most ASF usually occur within the first 3 months after BKP [12, 18, 19, 28], the vast majority of the BKP-associated ASFs should, therefore, be included in our study sample. One of the main reasons for considering early follow-ups was the goal to minimize the basic influences of patient’s primary disease such as osteoporosis on the incidence of ASFs.

Also, the analysis did not include data on bone mineral density (BMD) and data on sagittal alignment, as it was not part of the SSR dataset. Impaired sagittal balance may potentially be associated with a mechanical risk of further fractures.

Finally, the registry uses predefined categories for pre- or postoperative segmental kyphosis, which were grouped into four categories for the current analysis. The relation between the segmental kyphosis at follow-ups and occurrence of ASF could be more accurately studied in linear multivariate regression models considering kyphosis angle as a continuous variable.

On the other hand, this is the first study of ASF after BKP based on a national registry, and thus less dependent on a single treating institution or individual surgeon and related experience. The real-life data enable a more objective analysis with higher external validity, which is the major strength of the current study. Furthermore, the analysis contains the so far largest collective of patients treated with BKP for osteoporotic VCF.

Conclusions

A preoperative segmental kyphotic angle >30°, rheumatoid arthritis and cardiovascular disease are significantly associated with an early ASF and increased back pain after BKP in patients with mono-segmental osteoporotic insufficiency fractures. Controlled trials are required to compare non-surgical treatment and BKP regarding their incidence of ASFs.
Acknowledgments  We are thankful to the Swiss Spine Society and the SWISSspine registry group who made this research possible by populating the database with their valuable and much appreciated entries. The Kyphoplasty group contributing data to the current study is comprised of the following surgeons (in order of case contribution): Michael Schuler (n = 44), Thomas Forster (n = 38), Gianluca Maestrelli (n = 26), Urs Mueller (n = 23), Stefan Schäfer (n = 23), Max Aebi (n = 20), Markus Kröber (n = 19), Bruno Beele (n = 16), Ivan Broger (n = 16), Friedrich Sigier (n = 16), Martin Blay (n = 13), Ulrich Berlemann (n = 11), Thomas Lutz (n = 11), Christian Eiter (n = 10), Christian Bärlocher (n = 8), Christoph Binkert (n = 8), Patrick Moulin (n = 7), Fabrice Külling (n = 6), Stefan Kunz (n = 5), Thomas Markwalder (n = 5), Uwe Schwarz (n = 5), Bijan Cheikh-Sarraf (n = 4), Bernhard Jeanneret (n = 4), Marc Morard (n = 4), Enrico Tessitore (n = 4), Oliver Hausmann (n = 3), Karen Heimberger (n = 3), Philip Otten (n = 3), Michael Payer (n = 3), Andreas Schirm (n = 3), Christoph Hamburger (n = 2), Thomas M. Stoll (n = 2), Kristof Van Dommelen (n = 2), Martin Baur (n = 1), Michael Heinzellmann (n = 1), Knutti Oliver (n = 1), Rosa Martinez (n = 1), Aymen Ramadan (n = 1), Othmar Schwarzenbach (n = 1), Oliver Vernet (n = 1), Guido Wanner (n = 1).

Conflict of interest  The authors, or any member of their family, did not have any conflict of interest or financial support related to the subject of this article.

References


The Role of Drains in Lumbar Spine Fusion

Mohammad Sami Walid1, Moataz Abbara2, Abdullah Tolaymat2, James R. Davis3, Kevin D. Waits3, Joe Sam Robinson III2, Joe Sam Robinson Jr.2

OBJECTIVE: To study the role of drains in lumbar spine fusions.

METHODS: The charts of 402 patients who underwent lumbar decompression and fusion (LDF) were retrospectively reviewed. Patients were classified per International Classification of Diseases, 9th Edition (ICD-9) procedure code as 81.07 (lateral fusion, 74.9%) and 81.08 (posterior fusion, 25.1%). The investigators studied the prevalence of drain use in lumbar fusion procedures and the impact of drain use on postoperative fever, wound infection, posthemorrhagic anemia, blood transfusion, and hospital cost.

RESULTS: No significant differences in wound infection rates were noted between patients with and without drains (3.5% vs 2.6%, P = 0.627). The difference in postoperative fever rates between patients with and without drains (63.2% vs 52.6%, P = 0.05) was of borderline significance. Posthemorrhagic anemia was statistically more common in patients with drains (23.5% vs 7.7%, P = 0.000). Allogeneic blood transfusion was also statistically more common in the drained group (23.9% vs 6.8%, P = 0.000). Postoperative hemoglobin levels were lower in patients with drains who underwent one-level (9.5 g/dL vs 11.3 g/dL) or two-level (9.3 g/dL vs 10.2 g/dL) spine fusions. In this series in which drains were liberally used, no patient had to return to the operating room because of postoperative hematoma. An increased rate of allogeneic blood transfusion was noticed with posthemorrhagic anemia and drain use. The rate of allogeneic blood transfusion increased from 5.6% in patients without drains or posthemorrhagic anemia to 38.8% in patients with drains and posthemorrhagic anemia as a secondary diagnosis. The use of drains was associated with statistically insignificant increases in length of stay and cost in posterior procedures. Drain use was associated with shorter length of stay and hospital charges in lateral fusions of three or more levels.

CONCLUSIONS: Drain use did not increase the risk of wound infection in patients undergoing LDF, but it had some impact on the prevalence of postoperative fever. Drain use was significantly associated with posthemorrhagic anemia and allogeneic blood transfusion. Drain use did not have a significant economic impact on hospital length of stay and charges except in lateral procedures involving three or more levels.
performed by a neurosurgeon and an orthopedic surgeon, and a Jackson-Pratt closed-suction drain is habitually installed at the end of the procedure to decrease the risk of hematoma formation and neurologic compromise. In this study, the charts of 402 patients undergoing LDF operated between October 2007 and September 2009 were retrospectively reviewed. Patients were classified per International Classification of Diseases, 9th Edition (ICD-9) procedure code as 81.07 (lateral fusion, 74.9%) and 81.08 (posterior fusion, 25.1%):

- **81.07**
  1. Lateral transverse process technique
  1.781.08
  1. Arthrodesis of lumbar or lumbosacral region
    - a. Posterior (interbody) technique
    - b. Posterolateral technique
  2. Posterior lumbar interbody fusion
  3. Transforaminal lumbar interbody fusion

Postoperative temperature threshold was 100°F (37.7°C). Wound infection was determined by hospital course and follow-up. Posthemorrhagic anemia was extracted from the ICD-9 secondary diagnosis codes. Hospital charges were used as reflective of hospital cost.

Limitations
This study is based on hospital administrative data, which are coded using the ICD-9 system. These codes may not differentiate minimally invasive from open lumbar fusion cases. Estimated blood loss, recorded in anesthesia or surgery records, was not collected in this study; instead, postoperative hemoglobin and secondary ICD-9 diagnosis of posthemorrhagic anemia were used. This choice may have caused a selection bias.

### RESULTS

#### Demographic Characteristics and Comorbidities
The patient cohort comprised 57.6% women and 43% men. The mean age was 57.3 years.

<table>
<thead>
<tr>
<th>Principal Diagnosis</th>
<th>1 Level</th>
<th>2 Levels</th>
<th>≥3 Levels</th>
<th>1 Level</th>
<th>2 Levels</th>
<th>≥3 Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>721.3 Lumbosacral spondylosis</td>
<td>6</td>
<td>17</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>722.10 Lumbar disc displacement</td>
<td>37</td>
<td>48</td>
<td>15</td>
<td>23</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>722.52 Lumbar/lumbosacral disc degeneration</td>
<td>18</td>
<td>19</td>
<td>21</td>
<td>12</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>722.73 Lumbar disc disease with myelopathy</td>
<td>9</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>724.02 Spinal stenosis—lumbar</td>
<td>15</td>
<td>31</td>
<td>16</td>
<td>7</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>738.4 Acquired spondylolisthesis</td>
<td>15</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>756.12 Spondylolisthesis</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Diagnosis Codes, Procedure Codes, and Number of Levels

One-level (n = 37) and two-level (n = 48) lumbar/lumbosacral lateral fusions performed for lumbar disk displacement were the most common cases (Table 2).

Prevalence of Outcome Variables

The prevalence of outcome variables was as follows:

- 70.9% of patient had drains
- 60.1% had postoperative fever
- 3.2% had wound infection
- 18.9% had posthemorrhagic anemia
- 18.9% had allogeneic blood transfusion

- 0.0% returned to operating room because of hematoma

No significant differences in wound infection rates were noted between patients with and without drains (3.5% vs 2.6%, P = 0.627, Figure 1). The difference in postoperative fever rates between patients with and without drains (63.2% vs 52.6%, P = 0.05) was of borderline significance. In diabetic patients, drains were associated with a statistically insignificant increase in infection rate (7.1% vs 3.0%, P > 0.05).

Posthemorrhagic anemia was statistically more common in the group with drains (23.5% vs 7.7%, P = 0.000). Allogeneic blood transfusion was also statistically more common in the drained group (23.9% vs 6.8%, P = 0.000), especially with fewer operated levels (Figure 2). Postoperative hemoglobin levels were lower in patients with drains who underwent one-level (9.5 g/dL vs 11.3 g/dL) or two-level (9.3 g/dL vs 10.2 g/dL) spine fusions.

In a model including obesity, type of procedure, age, gender, number of levels, drain use, and blood transfusion, drain use and number of levels were significant predictors (P < 0.01) of posthemorrhagic anemia and allogeneic blood transfusion. Drain use was not a significant predictor of postoperative fever or wound infection. An increased rate of allogeneic blood transfusion was noticed with posthemorrhagic anemia and drain use.
use. The rate of allogeneic blood transfusion increased from 5.6% in patients without drains or posthemorrhagic anemia to 38.8% in patients with drains and posthemorrhagic anemia as a secondary diagnosis (Figure 3).

Economic Impact of Drain Use

The use of drains was associated with a statistically insignificant increase in length of stay and cost in posterior procedures. In contrast, drain use was associated with shorter length of stay and hospital charges in lateral fusions involving three or more levels (Figure 4).

DISCUSSION

Our study has several limitations inherent to any retrospective study. Not all variables could be normalized in regard to intraoperative factors and comorbidities other than obesity and diabetes (3, 6). Nevertheless, this investigation generated several useful findings.

On the positive side of drain use, there was no return to the operating room because of neurologic compromise related to postoperative bleeding, a finding that supports the use of drains in lumbar fusions. Also, the liberal use of drains was not associated with an increased risk of wound infection. This finding seems to be consistent with the findings of other researchers such as Payne et al. (4), who conducted a study on closed suction drainage after single-level lumbar laminectomy and found “no significant difference in the rate of infection or wound healing” between patients with and without drainage, and no patient developed postoperative neurologic deficit. Scuderi et al. (7) studied drain use after lumbar fusions at a single level for degenerative disease and concluded that drains do “not appear to increase the risks of wound related complications.” More recently, Kanayama et al. (2) showed that “the risk of wound infection and hematomas” in single-level lumbar decompression surgery “was not influenced by use of a drain.” From an economic point of view, drain use did not significantly increase hospital charges or length of stay.

On the negative side of drain use, we found an increased prevalence of postoperative fever associated with drain use, which could be a reaction to the invasiveness of surgery and the nature of drains as a foreign body. Drain use was associated with a significant increase in blood transfusion requirements. There was no way to know if drains have a role in impairing the coagulation cascade and preventing wound tamponade or whether drains are more willingly used in cases with excessive intraoperative bleeding. Because our study was retrospective, the rationale behind the decision to use a drain in each case was unknown.

CONCLUSIONS

To resolve these contradictory results, we recommend larger prospective controlled
studies in which drains are employed in a random fashion with the caveat that a surgeon may opt to remove a patient from the study and insert a drain if the surgeon believes the hemostatic status of the patient warrants this maneuver. Until such data become available, the use of drains will be based on individual surgical judgment.

REFERENCES

Conflict of interest statement: The authors declare that the article content was composed in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

received 08 December 2010; accepted 27 May 2011
Citation: World Neurosurg. (2012) 77, 3/4:564-568.
DOI: 10.1016/j.wneu.2011.05.058
Journal homepage: www.WORLDNEUROSURGERY.org
Available online: www.sciencedirect.com
1878-8750/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
Anterior approach versus posterior approach for the treatment of multilevel cervical spondylotic myelopathy: a systemic review and meta-analysis

Bin Zhu · Yilan Xu · Xiaoguang Liu · Zhongjun Liu · Gengting Dang

Abstract

Purpose To compare the clinical outcomes, complications, and surgical trauma between anterior and posterior approaches for the treatment of multilevel cervical spondylotic myelopathy.

Study design Systematic review and meta-analysis.

Methods We searched MEDLINE, EMBASE, and the Cochrane Central Register of Controlled Trials databases for randomized controlled trials or non-randomized controlled trials that compared anterior and posterior surgical approaches for the treatment of multilevel cervical spondylotic myelopathy. Exclusion criteria were non-controlled studies, combined anterior and posterior surgery, follow-up <1 year, cervical kyphosis >15°, and cervical myelopathy caused by ossification of the posterior longitudinal ligament. The main end points included: recovery rate; Japanese Orthopedic Association (JOA) score; reoperation rate; complication rate; blood loss; and operation time. Subgroup analysis was conducted according to the mean number of surgical segments.

Result A total of eight studies were included in the meta-analysis; none of which were randomized controlled trials. All of the selected studies were of high quality as indicated by the Newcastle–Ottawa scale. In five studies involving 351 patients, the preoperative JOA score was similar between the anterior and posterior groups ($P > 0.05$, WMD: $-0.00 (-0.56, 0.56)$). In four studies involving 268 patients, the postoperative JOA score was higher in the anterior group compared with the posterior group ($P < 0.05$, WMD: 0.79 (0.16, 1.42)). For recovery rate, there was significant heterogeneity among the four studies involving 304 patients, hence, only descriptive analysis was performed. In seven studies involving 447 patients, the postoperative complication rate was significant higher in the anterior group compared with the posterior group ($P < 0.05$, odds ratio: 2.60 (1.63, 4.15)). Of the 245 patients in the 8 studies who received anterior surgery, 21 (8.57 %) received reoperation. Of the 285 patients who received posterior surgery, only 1 (0.3 %) received reoperation. The reoperation rate was significantly higher in the anterior group compared with the posterior group ($P < 0.001$). In the 3 studies involving 236 patients compared subtotal corpectomy and laminoplasty/laminectomy, blood loss and operation time were significantly higher in the anterior subtotal corpectomy group compared with the posterior laminoplasty/laminectomy group ($P < 0.05$, WMD: 150.10 (63.53, 236.66) and $P < 0.05$, WMD: 59.17 (45.69, 72.66)).

Conclusion The anterior approach was associated with better postoperative neural function than the posterior approach in the treatment of multilevel cervical spondylotic myelopathy. There was no apparent difference in the neural function recovery rate. The complication and reoperation rates were significantly higher in the anterior group compared with the posterior group. The surgical trauma associated with corpectomy was significantly higher than that associated with laminoplasty/laminectomy.

Keywords Anterior approach · Posterior approach · Cervical spondylotic myelopathy · Systemic review · Meta-analysis
Introduction

Cervical spondylotic myelopathy (CSM) is the most common cause of myelopathy. The degeneration of the intervertebral disc and secondary degeneration of stable structures such as the uncovertebral joint, facet joint, posterior longitudinal ligament, and ligamentum flavum cause spinal cord compression and cervical myelopathy [1]. The presence of degenerative changes is very common in the general population and becomes more common with increasing age [2–4]. Individuals with congenital cervical canal stenosis are more vulnerable to cervical myelopathy caused by cervical degeneration and spinal cord compression [1–4].

The choice of surgical treatment for CSM remains controversial; however, most physicians agree that patients who have experienced symptoms for an extended period of time or who experience disease progression require surgical intervention [5, 6]. Generally, surgical approaches can be divided into anterior cervical canal decompression approaches and posterior cervical canal decompression approaches. Anterior cervical canal decompression approaches typically comprise corpectomy and anterior cervical discectomy and fusion (ACDF), whereas posterior cervical canal decompression approaches typically comprise laminoplasty and laminectomy.

For single-level CSM, ACDF is the gold standard. Indeed, several clinical case series’ have demonstrated >90 % fusion rates with satisfactory clinical results [7, 8]. However, for patients with CSM involving multiple segments associated with congenital cervical canal stenosis, the optimal surgical approach remains unclear.

Physicians who support the use of posterior decompression surgery believe that the surgical trauma and incidence of postoperative pseudarthrosis significantly increase with the number of involved segments when anterior surgical approaches are used [9, 10]. Findings from a recently published meta-analysis demonstrated that the fusion rate associated with the anterior procedure and plating was 97.1 % at the one-disc level, 94.6 % at the two-disc level, and 82.5 % at the three-disc level [11]. Anterior surgery also appears to be associated with a high rate of adjacent degeneration [12–14]. However, some authors have argued that although the adjacent degeneration rate is high with anterior cervical canal decompression approaches, it rarely leads to significant clinical symptoms or the need for reoperation [15, 16].

Physicians who support the use of anterior decompression surgery have expressed concern about the surgical trauma and high incidence of complications, such as axial pain and C5 root palsy, associated with posterior approaches [17, 18]. Further, some authors have suggested that patients who undergo posterior laminoplasty or laminectomy are more prone to develop cervical kyphosis or instability than those who receive treatment via anterior approaches [19].

Reports describing the advantages and disadvantages of anterior and posterior approaches for the treatment of multilevel CSM vary considerably; most are retrospective studies or single approach observational studies. Therefore, we performed a meta-analysis to compare the posterior approach with the anterior approach for the treatment of multilevel CSM.

End points of our analysis included: clinical outcome-related end points [recovery rate, Japanese Orthopedic Association (JOA) score, and Nurick grade]; complication-related end points (reoperation rate and complication rate); and operation-related end points (blood loss and operation time).

Methods

Inclusion criteria

Studies were included if they met the following criteria: (1) randomized or non-randomized controlled study; (2) included patients with CSM caused by multi-segmental spinal stenosis (≥2 segments); (3) included patients who underwent surgical treatment; (4) posterior cervical canal decompression and anterior cervical canal decompression were compared (regardless of the specific surgical approach); and (5) included patients >18 years of age.

Exclusion criteria

Studies were excluded if they: (1) were non-controlled; (2) combined anterior and posterior surgery; (3) had an average follow-up time of <1 year; (4) included patients with cervical kyphosis >15°; (5) included patients with cervical myelopathy caused by ossification of the posterior longitudinal ligament (OPLL). Studies involving patients with cervical kyphosis >15° were excluded because this is considered to be a contraindication for posterior surgery by many researchers [20]. Studies involving patients with cervical myelopathy caused by OPLL were excluded because this condition is different from CSM in terms of etiology, pathogenesis, and natural history; hence, this may have affected the surgeon’s decision making regarding the surgical approach used.

Search methods and selection of studies

We searched Ovid MEDLINE (1950–present), EMBASE (1980–present), and the Cochrane Central Register of Controlled Trials and bibliographies between May and July 2012. The search was not restricted to any specific
language or by year of publication. The following search terms and strategies were used: (1) cervical myelopathy OR CSM OR myelopathy OR cervical spondylosis OR cervical vertebrae OR cervical stenosis; (2) Corpectomy OR ACDF OR anterior cervical discectomy and fusion OR anterior decompression and fusion OR anterior decompression OR ventral decompression OR ventral approach OR ventral; (3) laminoplasty OR laminectomy OR posterior decompression OR posterior decompression and fusion OR dorsal decompression OR dorsal approach OR dorsal; (1) and (2) or (3).

One reviewer conducted the initial search of all databases. Reviewers were not blinded to the authors, journal, or source of financial support. In stage one, two reviewers independently reviewed the titles and abstracts identified in the initial search according to the inclusion and exclusion criteria. In stage two, the full text of articles identified in stage one was reviewed. If additional data or clarification was necessary, we contacted the study authors. Any disagreement between reviewers was resolved by discussion with three other reviewers.

Data extraction and management

The following information was collected from each study using a standardized form: (1) study ID; (2) study design; (3) study location; (4) main inclusion/exclusion criteria; (5) patient demographics; (6) length of follow-up; (7) number of surgical segments; (8) surgical approach for each group; (9) JOA scores before and after surgery; (10) recovery rate; (11) reoperation rate; (12) number of complications, type of complications, and rate of complications; and (13) operation time and blood loss.

Subgroup analysis

Subgroup analysis was conducted according to the mean number of surgical segments; subgroup A included studies in which the mean number of surgical segments was between 2 and 3, whereas subgroup B included studies in which the mean number of surgical segments was more than 3.

Statistical analysis

Heterogeneity was tested using a Chi-square test, for which $P < 0.1$ was considered to be statistically significant. Inconsistency was quantified by calculating the $I^2$ statistic. Continuous variables are presented as mean differences and 95% confidence intervals, whereas dichotomous variables are presented as odds ratios and 95% confidence intervals. For the pooled effects, a weighted mean difference (WMD) was calculated. Random-effects or fixed-effects models were used depending on the heterogeneity of the studies included. All statistical tests were performed using SAS software, version 9.1 (SAS Institute, Cary, NC, USA) and Review Manager, version 5.1 (The Cochrane Collaboration).

Results

Search results

The initial search identified 1396 studies in Ovid MEDLINE, 652 studies in Embase, and 196 studies in the Cochrane Central Register of Controlled Trials. Of these studies, 2,171 were excluded after review of the abstract and title. The reasons for exclusion were as follows: unrelated studies; not human studies; not comparative studies; case reports; review articles; and involved patients with OPLL related cervical myelopathy. The remaining 73 studies underwent full text review. A further 65 studies were subsequently excluded for the following reasons: single level CSM ($n = 14$); mean cervical kyphosis $> 15^\circ$ ($n = 1$); OPLL related cervical myelopathy ($n = 13$); combination of anterior and posterior surgical approaches used ($n = 5$); different surgical indications between groups ($n = 7$); and non-controlled studies ($n = 25$). Hence, a total of 8 studies were included in the meta-analysis [21–28]. The detail selection process is shown in Fig. 1. Tables 1 and 2 summarize the study characteristics and outcomes.

Quality assessment of the studies

No randomized controlled trials were identified. All eight studies included were non-randomized controlled studies: five were retrospective studies and three were prospective

---

**Fig. 1 Flow of studies through review**
Table 1 Common characteristics of included studies

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Study design</th>
<th>Study location</th>
<th>Number of patients</th>
<th>Patient age statistics</th>
<th>Follow-up time</th>
<th>Surgical segments</th>
<th>Surgical approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yonenobu et al. [21]</td>
<td>Non-randomized</td>
<td>Japan</td>
<td>Total: 83</td>
<td>A: 56.0 ± 11.5</td>
<td>A: 53 months</td>
<td>A: 2.51</td>
<td>A: corpectomy with fusion</td>
</tr>
<tr>
<td></td>
<td>Retrospective</td>
<td></td>
<td>A: 41, P: 42</td>
<td>P: 54.3 ± 8.9</td>
<td>P: 43 months</td>
<td>P: 2.64</td>
<td>P: laminoplasty</td>
</tr>
<tr>
<td></td>
<td>Comparative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wada et al. [22]</td>
<td>Non-randomized</td>
<td>Japan</td>
<td>Total: 47</td>
<td>A: 52.7 ± 7.6</td>
<td>A: 15 ± 2.7 years</td>
<td>A: 2.3 ± 0.7</td>
<td>A: corpectomy with fusion</td>
</tr>
<tr>
<td></td>
<td>Retrospective</td>
<td></td>
<td>A: 23, P: 24</td>
<td>P: 56.5 ± 11.2</td>
<td>P: 11.7 ± 0.9 years</td>
<td>P: 2.5 ± 0.8</td>
<td>P: laminoplasty</td>
</tr>
<tr>
<td></td>
<td>Comparative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edwards et al. [23]</td>
<td>Non-randomized</td>
<td>USA</td>
<td>Total: 26</td>
<td>A: 53</td>
<td>A: 49 months</td>
<td>A: 3.15</td>
<td>A: corpectomy with fixation and fusion</td>
</tr>
<tr>
<td></td>
<td>Retrospective</td>
<td></td>
<td>A: 13, P: 13</td>
<td>P: 54</td>
<td>P: 40 months</td>
<td>P: C3–C7 12</td>
<td>P: laminoplasty</td>
</tr>
<tr>
<td></td>
<td>Matched-cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C4–C7 1</td>
<td></td>
</tr>
<tr>
<td>Kristof et al. [24]</td>
<td>Non-randomized</td>
<td>Germany</td>
<td>Total: 103</td>
<td>A: 62.5 ± 10.61</td>
<td>A: 196.56 ± 212.0 months</td>
<td>A: 2</td>
<td>A: corpectomy and plate-screw-instrumented fusion</td>
</tr>
<tr>
<td></td>
<td>Retrospective</td>
<td></td>
<td>A: 42, P: 61</td>
<td>P: 66.0 ± 12.4</td>
<td>P: 66.53 ± 34.21 months</td>
<td>P: 3</td>
<td>P: laminoplasty and rod-screw-instrumented fusion</td>
</tr>
<tr>
<td></td>
<td>Comparative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu et al. [25]</td>
<td>Non-randomized</td>
<td>China</td>
<td>Total: 52</td>
<td>A: 54.64 ± 11.49</td>
<td>A: 25.40 ± 13.76 (months)</td>
<td>A: 3.36 ± 0.57</td>
<td>A: ACDF with PCB system</td>
</tr>
<tr>
<td></td>
<td>Prospective</td>
<td></td>
<td>A: 25, P: 27</td>
<td>P: 57.33 ± 10.09</td>
<td>P: 27.47 ± 11.06 (months)</td>
<td>P: 3.67 ± 0.55</td>
<td>P: laminoplasty</td>
</tr>
<tr>
<td></td>
<td>Controlled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghogawala et al. [26]</td>
<td>Non-randomized</td>
<td>USA</td>
<td>Total: 50</td>
<td>A: 60</td>
<td>1 year</td>
<td>A: 2.1</td>
<td>A: multi-level of discectomy with fusion and fixation</td>
</tr>
<tr>
<td></td>
<td>Prospective</td>
<td></td>
<td>A: 28, P: 22</td>
<td>P: 64</td>
<td></td>
<td>P: 3.1</td>
<td>P: laminoplasty with lateral mass fixation and fusion</td>
</tr>
<tr>
<td></td>
<td>Pilot</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hirai et al. [27]</td>
<td>Non-randomized</td>
<td>Japan</td>
<td>Total: 86</td>
<td>A: 59.2 ± 10.7</td>
<td>5 years</td>
<td>A: 2.18</td>
<td>A: corpectomy with fusion and fixation</td>
</tr>
<tr>
<td></td>
<td>comparative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C3–6 20</td>
<td></td>
</tr>
<tr>
<td>Shibuya et al. [28]</td>
<td>Non-randomized</td>
<td>Japan</td>
<td>Total: 83</td>
<td>A: 60.4 ± 8.4</td>
<td>A: 11 years 11 months</td>
<td>A: 3.18</td>
<td>A: corpectomy with autologous bone grafting and halo-vest external fixation</td>
</tr>
<tr>
<td></td>
<td>Retrospective</td>
<td></td>
<td>A: 34, P: 49</td>
<td>P: 64.8 ± 11.7</td>
<td>P: 8 years 3 months</td>
<td>P: 3.10</td>
<td>P: laminoplasty</td>
</tr>
<tr>
<td></td>
<td>Analysis of case series</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A anterior surgery group, P posterior surgery group
<table>
<thead>
<tr>
<th>Study ID</th>
<th>JOA score</th>
<th>Recovery rate</th>
<th>Nurick grade</th>
<th>Complication</th>
<th>Reoperation rate and reason for reoperation</th>
<th>Blood loss, operation time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yonenobu et al. [21]</td>
<td>BO: A: 8.2 ± 2.2, P: 9.3 ± 3.0</td>
<td>A: 44.9 ± 26.2</td>
<td>–</td>
<td>A: 12/41 29.3 % graft complications (10) esophageal fistula (1) retrolisthesis (1) P: 3/42 7.1 % C5 palsy(3)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>PO: A: 13.3 ± 2.6, P: 12.8 ± 2.7 (final follow-up data)</td>
<td>P: 55.3 ± 30.2</td>
<td>(final follow-up data)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wada et al. [22]</td>
<td>BO: A: 7.9 ± 1.8, P: 7.4 ± 2.2</td>
<td>–</td>
<td>–</td>
<td>A: 8/23 34.8 % graft complication (7) esophageal fistula (1) P: 4/24 17 % C5 palsy (4)</td>
<td>A: 7/23(6 for non-union of the graft and 1 for adjacent deterioration), P: 0</td>
<td>A: 264 ± 65 min; 986 ± 751 g P: 182 ± 43 min; 608 ± 212 g</td>
</tr>
<tr>
<td></td>
<td>1 year PO: A: 13.3 ± 1.6, P: 13.1 ± 2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 years PO: A: 13.9 ± 2.0, P: 12.9 ± 2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Final PO: A: 13.4 ± 2.8, P: 12.2 ± 3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edwards et al. [23]</td>
<td>–</td>
<td>–</td>
<td></td>
<td>A: 9/13 69.2 % progression of myelopathy (1) pseudoarthrosis (1) subjacent ankylosis (1) persistent dysphagia (4) persistent dysphonia (2) P: 1/13 7.7 % radiculopathy (1)</td>
<td>A: 0</td>
<td>A: 224 min; 572 ml P: 216 min; 360 ml</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kristof et al. [24]</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu et al. [25]</td>
<td>BO: A: 8.16 ± 3.41P: 8.59 ± 2.98</td>
<td>A: 59.79 ± 23.43 %</td>
<td>–</td>
<td>A: 9/25 36.00 % late deterioration (2) screw loosening (1) pseudoarthrosis (1) subjacent ankylosis (1) temporary odynophagia (2) temporary dysphonia(2) P: 3/27 11.11 % C5 root palsy (2) axial neck pain (1)</td>
<td>A: 3/25(2 late deterioration and 1 screw loosening) P: 0/27</td>
<td>A: 115.92 ± 24.14 min; 118.48 ± 27.62 ml P: 187.78 ± 25.01 min; 361.11 ± 57.8 ml</td>
</tr>
<tr>
<td></td>
<td>PO: A: 13.20 ± 2.72P: 13.67 ± 2.70</td>
<td>P: 59.54 ± 29.37 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study ID</td>
<td>JOA score</td>
<td>Recovery rate</td>
<td>Nurick grade</td>
<td>Complication</td>
<td>Reoperation rate and reason for reoperation</td>
<td>Blood loss, operation time</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
<td>---------------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------------------------------------</td>
<td>-----------------------------</td>
</tr>
</tbody>
</table>
| Ghogawala et al. [26] | BO: A: 13.40 ± 0.44, P: 11.60 ± 0.5  
              PO: A: 15.44 ± 0.39, P: 13.54 ± 0.45 | – | – | A: 5/28 (17.9 %)  
P: 3/22 (13.6 %) | | |
| Hirai et al. [27]  | BO: A: 9.9 ± 3.1, P: 9.7 ± 2.9  
              1 year PO: A: 14.0 ± 2.6, P: 13.3 ± 2.5  
              2 years PO: A: 14.8 ± 2.0, P: 13.5 ± 2.5  
              3 years PO: A: 15.0 ± 2.3, P: 13.5 ± 2.6  
              5 years PO: A: 14.9 ± 2.3, P: 13.1 ± 2.9 | 1 year PO: A: 59.9 ± 27.4 %, P: 49.5 ± 25.8 %  
              2 years PO: A: 63.52 ± 28.6 %, P: 50.4 ± 27.3 %  
              3 years PO: A: 74.1 ± 25.4 %, P: 52.5 ± 27.3 %  
              5 years PO: A: 72.9 ± 28.3 %, P: 50.2 ± 26.6 % | – | A: 7/39 17.95 % airway problems (3)  
meralgia (2)  
C5 palsy (1)  
pseudoarthrosis (1)  
P: 3/47 6.38 % C5 palsy (3) | A: 1/39(pseudoarthrosis)  
P: 0 | A: 211 ± 55.3 min;  
340 ± 287 ml  
P: 149 ± 38.7 min;  
188 ± 92.1 ml |
| Shibuya et al. [28] | BO: A: 8.6 ± 2.9, P: 7.9 ± 2.4  
              1 years PO: A: 55.5 ± 25.3 %, P: 61.4 ± 21.2 %  
              5 years PO: A: 49.3 ± 29.3 %, P: 41.0 ± 26.6 %  
              12 years PO: A: 52.4 ± 28.1 %, P: 50.9 ± 25.9 % | 1 years PO: A: 59.9 ± 27.4 %, P: 49.5 ± 25.8 % | – | – | A: 10/34(6 for pseudoarthrosis and 4 for adjacent deterioration), P: 0 | A: 3-levels  
1818 ± 1607 ml  
2-levels 1292 ± 942 ml  
1-level 662 ± 553 ml  
P: 404 ± 426 ml  
A: 3-levels  
371 ± 89 min  
2-levels 334 ± 73 min  
1-level 265 ± 51 min  
P: 175 ± 60 min |

A anterior surgery group, P: posterior surgery group, BO before operation, PO post operation
studies. The major baseline characteristics of participants in each study (inclusion and exclusion criteria, demographics, preoperative JOA scores, surgical segments, and surgical procedures) were similar (Table 1). The quality of studies included was assessed using the Newcastle–Ottawa quality assessment scale for non-randomized case-controlled studies and cohort studies (www.ohri.ca/programs/clinical_epidemiology/oxford.htm) [29]. Of the studies, seven scored 8 points and one scored 7 points; hence, the studies were of a relatively high quality (Table 3).

Surgical approaches

The eight studies included a total of 530 patients; 245 who underwent anterior surgery, 192 who underwent corpectomy, and 53 who underwent multi-level discectomy with fusion and fixation. Of the 285 patients who underwent posterior surgery, 83 underwent laminectomy with fusion and fixation and 202 underwent laminoplasty.

Clinical outcome

Five studies used the JOA score to assess the clinical outcome, all of which provided a preoperative JOA score (n = 351 patients; 162 in the anterior surgery group and 189 in the posterior surgery group). The preoperative JOA score was similar between the two groups [P < 0.05, WMD = 0.00 (−0.56, 0.56); heterogeneity: \( \chi^2 = 6.04, df = 4, P = 0.20 \); I² = 34 %, fixed-effects model, Fig. 2]. There was no significant difference in the preoperative JOA score between the anterior surgery group and the posterior group in either subgroup A [P < 0.05, WMD = 0.17 (−0.85, 0.51); heterogeneity: \( \chi^2 = 4.23, df = 2, P = 0.12 \); I² = 53 %, fixed-effects model] or subgroup B [P < 0.05, WMD = 0.34 (−0.64, 1.32); heterogeneity: \( \chi^2 = 1.10, df = 1, P = 0.29 \); I² = 9 %, fixed-effects model]. Four of five studies provided a postoperative (last follow-up) JOA score (n = 268 patients; 128 in the anterior surgery group and 140 in the posterior surgery group). The postoperative JOA score was significantly higher in the anterior surgery group compared with the posterior surgery group [P < 0.05, WMD = 0.79 (0.16, 1.42); heterogeneity: \( \chi^2 = 5.09, df = 3, P = 0.17 \); I² = 41 %, fixed-effects model, Figs. 3]. There was a

### Table 3 Quality assessment according to the Newcastle–Ottawa scale

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Selection</th>
<th>Comparability</th>
<th>Exposure</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yonenobu et al. [21]</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Wada et al. [22]</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Edwards et al. [23]</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Kristof et al. [24]</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Liu et al. [25]</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Ghogawala et al. [26]</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Hirai et al. [27]</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Shibuya et al. [28]</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

Fig. 2 Weight mean difference of preoperative JOA score between anterior surgery group and posterior surgery group

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Total</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hirai et al. [27]</td>
<td>9.9</td>
<td>3.1</td>
<td>39</td>
<td>9.7</td>
<td>2.9</td>
</tr>
<tr>
<td>Liu et al. [25]</td>
<td>8.16</td>
<td>3.41</td>
<td>25</td>
<td>8.59</td>
<td>2.98</td>
</tr>
<tr>
<td>Shibuya et al. [28]</td>
<td>8.6</td>
<td>2.9</td>
<td>34</td>
<td>7.9</td>
<td>2.4</td>
</tr>
<tr>
<td>Wada et al. [22]</td>
<td>7.9</td>
<td>1.8</td>
<td>23</td>
<td>7.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Yonenobu et al. [21]</td>
<td>8.2</td>
<td>2.2</td>
<td>41</td>
<td>9.3</td>
<td>3</td>
</tr>
</tbody>
</table>

Total (95% CI) 162 189 100.0% -0.00 [-0.56, 0.56]

Heterogeneity: \( \text{Chi}^2 = 6.04, df = 4 (P = 0.20); I^2 = 34 \%

Test for overall effect: Z = 0.00 (P = 1.00)

Fig. 3 Weight mean difference of postoperative JOA score between anterior surgery group and posterior surgery group

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Total</th>
<th>Total</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hirai et al. [27]</td>
<td>15</td>
<td>2.3</td>
<td>39</td>
<td>13.5</td>
<td>2.6</td>
<td>36.9% 1.50 [0.46, 2.54]</td>
</tr>
<tr>
<td>Liu et al. [25]</td>
<td>13.2</td>
<td>2.72</td>
<td>25</td>
<td>13.67</td>
<td>2.7</td>
<td>18.2% -0.47 [-1.94, 1.00]</td>
</tr>
<tr>
<td>Wada et al. [22]</td>
<td>13.4</td>
<td>2.8</td>
<td>23</td>
<td>12.2</td>
<td>3</td>
<td>14.4% 1.20 [0.46, 2.86]</td>
</tr>
<tr>
<td>Yonenobu et al. [21]</td>
<td>13.3</td>
<td>2.6</td>
<td>41</td>
<td>12.8</td>
<td>2.7</td>
<td>30.5% 0.50 [-0.64, 1.64]</td>
</tr>
</tbody>
</table>

Total (95% CI) 128 140 100.0% 0.79 [0.16, 1.42]

Heterogeneity: \( \text{Chi}^2 = 5.09, df = 3 (P = 0.17); I^2 = 41 \%

Test for overall effect: Z = 2.47 (P = 0.01)
significant difference in the postoperative JOA score between the anterior surgery group and the posterior group in subgroup A \( \chi^2 = 14.80, df = 1, P = 0.0001; \hat{I}^2 = 93 \% \). In one study [27], the recovery rate was significantly higher in the anterior surgery group compared with the posterior surgery group (72.9 ± 28.3 % vs 50.2 ± 26.6 %, \( P < 0.05 \)). In the other study in subgroup A [21], the recovery rate was similar between the anterior and posterior surgery groups (44.9 ± 26.2 % vs 55.3 ± 30.2 %, \( P > 0.05 \)). There was no statistically significant difference in recovery rate between the anterior and posterior surgery groups in subgroup B \( \chi^2 = 1.65, df = 2, P = 0.44; \hat{I}^2 = 0 \% \), fixed-effects model]. Two studies used Nurick grade to assess the clinical outcome, but neither provided the standard deviation; hence, meta-analysis could not be performed.

Complications and reoperation rate

The study records of postoperative complications were relatively inconsistent between studies and the definition of complications varied. Some studies provided all complications, whereas some provided the overall complications rate. Seven studies (\( n = 447 \) patients; 211 in the anterior surgery group and 236 in the posterior surgery group) provided a list of the postoperative complications or the incidence of postoperative complications (Table 1). The postoperative complication rate was significant higher in the anterior surgery group compared with the posterior surgery group \( \chi^2 = 9.66, df = 6, P = 0.14; \hat{I}^2 = 38 \% \), fixed-effects model, Fig. 5]). The postoperative complication rate was significantly higher in the anterior surgery group compared with the posterior surgery group in subgroup A \( \chi^2 = 8.91, df = 5, P = 0.11; \hat{I}^2 = 44 \% \), fixed-effects model]. Subgroup B comprised only one study; hence, meta-analysis could not be performed.

![Fig. 4](image_url) Weight mean difference of recovery rate between anterior surgery group and posterior surgery group. There was significant heterogeneity between studies (\( \chi^2 = 15.65, P = 0.001, \hat{I}^2 = 81 \% \))

![Fig. 5](image_url) Odds ratio of complication rates between anterior surgery group and posterior surgery group
Of the 245 patients from 8 studies who received anterior surgery, 21 (8.57 %) required reoperation; 13 (61.9 %) for pseudoarthrosis/non-union of the graft; 7 (33.3 %) for adjacent deterioration; and 1 (4.8 %) for fixation loosening. Of the 285 patients who received posterior surgery, only 1 (0.3 %) required reoperation for radiculopathy due to new disc herniation. The reoperation rate was significantly higher in the anterior surgery group compared with the posterior surgery group ($P < 0.001$).

Blood loss and operation time

Six studies reported the intraoperative blood loss and operation time; however, two of these studies were excluded from the meta-analysis (one did not provide the standard deviation of intraoperative blood loss and operation time, whereas the other only provided the blood loss and operation time for the subgroups). A total of 236 patients from 3 studies (104 patients in the anterior surgery group and 132 patients in the posterior surgery group) were included in the comparison of blood loss and operative time for subtotal corpectomy vs laminoplasty/laminectomy. Blood loss and operation time were significantly higher in the anterior subtotal corpectomy group compared with the posterior surgery laminectomy/laminoplasty group [$P < 0.05$, WMD = 150.10 (63.53, 236.66); heterogeneity: $\chi^2 = 4.29$, $df = 2$, $P = 0.12$; $I^2 = 53$ %, fixed-effects model; and $P < 0.05$, WMD = 59.17 (45.69, 72.66); heterogeneity: $\chi^2 = 3.63$, $df = 2$, $P = 0.16$; $I^2 = 45$ %, fixed-effects model, Figs. 6, 7]. All three studies were included in subgroup A; hence, subgroup analysis was not performed. One study, which compared blood loss and operation time for anterior multilevel discectomy and posterior laminoplasty, was not included in the meta-analysis.

**Discussion**

Several review articles have been published regarding decision making in the treatment of CSM [30, 31]. Cunningham et al. [30] published a systemic review of cohort studies comparing surgical treatment for CSM, including single segment CSM. In this article, the authors’ focused on comparing the clinical outcomes with different surgical approaches (ACDF, corpectomy, laminectomy and laminoplasty). The literature suggests that decision making in the treatment of multi-segmental CSM can be challenging for treating physicians. Blood loss, operation time, the incidence of postoperative adjacent degeneration, and the incidence of complications all increase significantly with the number of involved anterior surgical segments. Liu et al. [31] published a systemic review on surgical approaches (anterior or posterior) for multilevel myelopathy, including that caused by cervical spondylosis and OPLL. Over the last 2 years, a series of comparative studies on surgical approaches for the treatment of multi-segmental CSM have been published; therefore, we designed a systemic review and meta-analysis to compare surgical approaches (anterior or posterior) for the treatment of multilevel CSM.

Selection of the surgical approach for the treatment of multi-level CSM remains controversial. Ghogawala et al. [32] performed a study assessing eligibility for...
randomization to surgical approaches in the treatment of CSM. In the study, the authors sent 20 images associated with actual cases to 239 surgeons and asked the surgeons to provide demographic information, their preferred surgical approach, and eligibility for randomization for 10 cases. Of the 20 cases, 12 were considered to be potentially eligible for randomization. Although no RCT studies were included in our study, all selected studies were of high quality and the baseline variables (such as age, gender ratio, preoperative JOA score, surgical segments, and surgical procedures) were similar; hence, we considered the included studies to be comparable.

Two clinical outcome end points were selected in our meta-analysis. In the meta-analysis of JOA scores, there was no significant difference between the anterior surgery group and the posterior surgery group compared with the posterior surgery group. Postoperative JOA scores were better in the anterior surgery group compared with the posterior surgery group. Subgroup analysis was consistent with the overall analysis and the heterogeneity was low to median. These findings indicate that the two groups had similar baseline neural function, but that postoperative neural function condition was better in the anterior surgery group compared with the posterior surgery group. However, meta-analysis of recovery rate, which reflects postoperative improvement in neural function condition (JOA scores), there was significant heterogeneity between groups ($\chi^2 = 15.65$, $df = 3$, $P = 0.001$; $I^2 = 81\%$).

Subsequent subgroup analysis revealed low heterogeneity and no significant differences in subgroup B. In contrast, there was significant heterogeneity in the one study in subgroup A, with a significantly higher recovery rate in the anterior surgery group compared with the posterior group ($72.9 \pm 28.3\%$ vs $50.2 \pm 26.6\%$, $P < 0.05$). In this article [27], the authors suggested that residual anterior compression of the spinal cord after LAMP was the cause of the lower recovery rate in the posterior surgery group. Interestingly, the recovery rate without residual anterior compression was similar between the posterior surgery and anterior surgery groups ($P > 0.05$). When this study was excluded, there was no significant difference in the recovery rate between the two groups ($P > 0.05$, WMD = $-3.14$ ($-10.47$, 4.18); heterogeneity: $\chi^2 = 2.17$, $df = 2$, $P = 0.34$; $I^2 = 8\%$). Although postoperative neural function condition was better in the anterior surgery group compared with the posterior surgery group, there was no difference in the recovery rate. There seemed no difference of clinical effects between anterior approach and posterior approach for the treatment of multilevel CSM.

We selected the reoperation rate for descriptive analysis and the complication rate for meta-analysis in the evaluation of complication-related outcomes. In the meta-analysis of complication rate, we found a significantly higher incidence of complications in the anterior surgery group compared with the posterior surgery group. Subgroup analysis findings were similar. This indicates that anterior approaches for the treatment of multilevel CSM are associated with a higher incidence of complications. Considering the most of the complications were pseudarthrosis or non-fusion on the anterior approaches, this seemed to be due to technical reasons and quality of the bone grafts. We also assessed reoperation rate as another complication-related measure. We found that the reoperation rate was significantly higher in the anterior surgery group compared with the posterior surgery group ($P < 0.05$). Although the indications for reoperation between studies were not the same, anterior approach for the treatment of multilevel CSM seemed to have a high risk of reoperation.

In the evaluation of surgical trauma, operation time and blood loss were selected for meta-analysis. We compared studies in which corpectomy was performed with studies in which laminoplasty/laminectomy was performed. Both overall and subgroup analyses revealed that blood loss and operation time were significantly higher in the corpectomy group compared with the laminoplasty/laminectomy group. This indicates that, in the treatment of multilevel CSM, the surgical trauma associated with corpectomy is higher than that associated with laminoplasty/laminectomy. Meta-analysis of surgical trauma between multilevel ACDF surgery and laminoplasty/laminectomy was not performed because only one relevant study was identified.

Our study has a number of limitations that warrant mention. Firstly, none of the studies included in the meta-analysis were RCTs. Secondly, there was variability among the studies in the choice of indicators to evaluate the postoperative clinical effect. This clearly reflects the lack of a gold standard outcome measure. Thirdly, most of the studies focused on the evaluation of neurological function improvement (i.e., JOA score and recovery rate), but neglected to evaluate overall quality of life using instruments such as the SF-36 scale and the HR-QOL scale. Finally, follow-up time varied between the studies and thus may have influenced our results.

**Conclusion**

In conclusion, although the anterior approach was associated with better postoperative neural function than the posterior approach in the treatment of multilevel cervical spondylotic myelopathy, there was no apparent difference in the neural function recovery rate. The complication and reoperation rates were significantly higher in the anterior group compared with the posterior group. The surgical trauma associated with corpectomy was significantly higher than that associated with laminoplasty/laminectomy.
The comparison of multilevel ACDF with laminoplasty/ laminectomy requires further investigation.

Conflict of interest None.

References