Systematic review of the effect of fusion added to discectomy compared with discectomy alone for lumbar disk prolapse surgery

Ricardo Vieira Botelho¹,² Thiago Lusvarghi Cardoso² Wanderley Marques Bernardo³,⁴
¹Neurosurgical Service, Hospital do Servidor Público do Estado de São Paulo; ²Faculdade de Medicina, Universidade Cidade de São Paulo (UNICID); ³Faculdade de Medicina, Universidade Lusíadas (UNILUS), Santos; ⁴Hospital das Clínicas da Universidade de São Paulo, São Paulo, Brazil

Introduction: Lumbar disk herniation is the pathologic condition most commonly responsible for radicular pain, and the condition for which lumbar surgery is performed most frequently. Fusion in the lumbar spine, which is associated with surgical treatment of disk prolapse, has been discussed since the beginning of the contemporary era of lumbar spine surgery. It is questionable whether or not fusion would address the potential effects in the evolutive degenerative process in the years following the surgery. A systematic review of controlled studies was conducted to clarify the effect of fusion.

Methods: An evidenced-based practice systematic method called the PICO was used to search the best evidence in MEDLINE using PubMed tools.

Results: Only retrospective comparative studies were found, and they revealed superior long-term results in outcomes such as satisfactory rates, recurrence of disk herniation, and chronic low back rates for a fusion group of patients.

Discussion: Several papers have concluded that there is conflictive evidence of the role for fusion in surgery for lumbar herniated disks. For the first time, this conflictive evidence has been clarified by means of systematic review and meta-analysis.

Conclusion: There were substantial statistical heterogeneous results favoring spinal fusion in surgery for lumbar disk herniation, producing conflictive evidence to support fusion as an option.

Keywords: intervertebral disk displacement, lumbar vertebrae, surgical procedures, discectomy, spinal fusion, treatment outcome

Introduction
In 1934, Mixter and Barr¹ published the classic monograph describing the hitherto little-known clinical syndrome related to rupture of the intervertebral disk with spinal canal invasion and its aftermath. Lumbar disk herniation is the pathologic condition most commonly responsible for radicular pain, and the condition for which lumbar surgery is performed most frequently.²

The introduction of microsurgery in the surgical armamentarium allowed smaller surgeries with better lighting and more minimally invasive procedures. However, disk degeneration and its consequences is increasing in prevalence with age, which reveals that the degenerative process continues its role in the pathophysiology of spine diseases until later in life.³

Arthrodesis (fusion) in the lumbar spine, which is associated with surgical treatment of disk prolapse, has been discussed since the beginning of the contemporary era of lumbar spine surgery.⁴ It is questionable whether or not fusion addresses potential effects in the evolutive degenerative process in the years following the surgery.
Fusion after treatment of disk prolapse with sciatica has been one of the major disagreements between spinal surgeons.5–10

The purpose of this review is to analyze the quality of published comparative trials related to fusion in surgery for lumbar herniated disk and to potentially clarify whether or not fusion increases the chance of a successful surgery.

Methods
A clinical question was formulated utilizing an evidence-based practice systematic method called the PICO question, which means: P: population of interest, I: intervention, C: comparison intervention, and O: outcome of interest. The PICO method guided the construction of a focused question: (P) intervertebral disk displacement, lumbar vertebrae (I) discectomy, (C) compared with spinal fusion, (O) treatment outcomes.11

The study population included adult patients of both sexes with sciatic pain and lumbar disk prolapse. There was no time limit for the search.

Studies primarily addressing the treatment of chronic low back pain, referenced by the terms “degenerative disk disease”, “discogenic pain”, “chronic low back pain”, “Black disk disease”, and “spondylosis”, were excluded from the research. Papers evaluating the surgeries for recurrences of disk herniations were, in the same way, excluded.

Types of intervention
A comparison was carried out between the results of surgical treatment of lumbar disk prolapse with and without arthrodesis.

Types of outcomes studied
Types of outcomes studied include (1) the rate of surgical success or failure, defined as satisfactory or unsatisfactory result; (2) the rate of recurrence of disk prolapse; and (3) the rate of residual low back pain after surgery.

The search for papers was carried out by two independent reviewers.

Design of studies included
Only comparative studies (controlled) between the two techniques were included and subsequently classified according to their strength of recommendation.

Databases consulted
The papers described were in the MEDLINE and CENTRAL Cochrane databases for randomized trials in the English language.12

Search strategy
The MEDLINE-PubMed search strategy involved a systematic review using the electronic search tool in the PubMed database (www.pubmed.com) by structured question using “PICO” according to the following descriptors and in a manual search in the references of the selected papers:

(“Intervertebral disk displacement” AND “lumbar vertebrae” AND (diskectomy OR “spinal fusion”) AND (“clinical trials as topic” [MeSH Terms] OR “Controlled Clinical Trial” [Publication Type] OR trial [Text Word]).

The CENTRAL Cochrane of randomized trials search strategy was done under a broad base encompassing the terms “lumbar spine” AND “intervertebral disk displacement” AND discectomy OR “spinal fusion”.

Statistics
The individual and pooled results were described as risk, risk difference (increasing or decreasing absolute risk), risk ratio, and the number of patients needed to be treated to obtain a benefit (NNT) or to obtain damage (harm) (NNH). The significance of the result was expressed as a confidence interval or as probability of type 1 error ($P$).13 The statistical heterogeneity of results was evaluated with the Cochran $Q$ test and $I^2$ test. The software used in meta-analysis was Biostat’s Comprehensive Meta-Analysis (Englewood, NJ). Data were analyzed based on a randomic model.

Results
Evidence recovered: first quality analysis of resulting papers
Initially, 14,409 papers were selected from MEDLINE. From these, 1070 were described as clinical trials in humans. Papers related to the cervical spine, fractures and traumatic conditions, biomechanical studies, artificial disks, ankle, chronic low back pain, scoliosis, deformities, difference between autologous and heterologous grafts, tuberculosis, stenosis, or spondylolisthesis were excluded. We also excluded studies evaluating recurrent herniated disk treatment outcome. Of the remaining papers, 684 papers conducted clinical trials in humans and were published in English. Of these, we selected 34 potential papers by the title related to the study. These papers were reviewed by title and abstract, and seven studies comparing surgical treatments of lumbar disk prolapse, with and without arthrodesis (controlled), were selected for final review.

The Cochrane CENTRAL database of randomized trials recovered 560 papers. This search did not reveal any additional related paper to the MEDLINE search.
Evidence recovered: second quality analysis of papers published

Two papers were recorded by analyzing the related papers and cross-referenced between the described papers. The CENTRAL search revealed one potential paper for inclusion in the final analysis. The paper by Karabekir et al studied 29 patients with degenerative disk disease (eight patients had recurrent disk herniations), 40 patients who underwent simple discectomy, and another group of patients with degenerative disk disease, facet hypertrophy, and lateral spinal canal stenosis. The techniques compared were simple discectomy and two techniques of intersomatic fusion. We consider the compared groups to be heterogeneous and pooled together with other comparative papers.

Seven comparative papers were published (Figure 1, Table 1). None were randomized trials. The most recent article is from 2000, and the oldest is from 1967. All papers described patients who underwent discectomy (for disk prolapse and nerve root compression), with or without spinal fusion.

Although White et al had described their paper as prospective, the final analysis revealed it to be retrospective. The seven studies reviewed were retrospective case series with a controlled branch, classified as level of evidence 2C, grade of recommendation B, or retrospective cohort or outcome research, according to the Oxford Centre for Evidence Based Medicine classification, (available at http://www.cebm.net).

The level of existing evidence was also classified according to van Tulder et al as limited or conflicting evidence, provided by “only one randomized trial or by inconsistent findings”.

Description of recorded papers

Of the seven studies, six used posterior or posterolateral fusion with autograft, and only one used adjuvant instrumentation with Knodt rod (distraction). The results of these seven studies initially described 1241 patients (Table 1). The evaluation of the studies identified 600 patients treated with fusion (experimental group) and 605 in the nonfusion group (control).

The description of the sample consisted of patients with clinical pictures that were very similar between the studies.

The mean follow-up time was 6.46 years (Table 1).

Results for satisfactory or unsatisfactory rates outcomes

All studies assessed the rate of satisfactory and unsatisfactory treatment. The level of satisfaction was mainly evaluated by the same or a similar scale to that used by Webster and Smiley. In this, the results are classified as excellent, good, fair, or poor. In the excellent and good results, patients return to full work with no limitation and no need for medication (excellent) or only occasionally (good). Results worse than these were considered to be unsatisfactory.

Takeshima et al used the scale from the Japanese Orthopaedic Association to obtain results, but the description provided similar satisfactory and unsatisfactory results.

All papers provided the number of patients (or proportion) with satisfactory and unsatisfactory results. We calculated the risk of an unsatisfactory event in the experimental group (fusion) and the control group (discectomy alone) (Table 2), the relative risk of failure between the groups, the risk reduction, the NNT and NNH, and confidence intervals (CI).

The risk of an unsatisfactory outcome with fusion was 18% (CI 95%: 0.09 to 0.33), and without fusion (discectomy) the risk was 25% (CI 95%: 0.16 to 0.47) (Table 1). The relative risk of failure between the experimental (fusion) and control groups was 70% (CI 95%: 0.566 to 0.875) (Table 2). NNT was 13 patients (CI 95%: 8 to 34).

![Figure 1](https://www.dovepress.com/submit-your-manuscript.png)

**Figure 1** Forest plot showing risk difference of unsatisfactory results between seven papers.

**Note:** With the inclusion of White et al’s result (instrumentation), the diamond (overall result) touches the zero line (nonsignificant result).

**Abbreviations:** favors A, favors fusion; favors B, favors discectomy.
Table 1 Description of selected papers

<table>
<thead>
<tr>
<th>Author</th>
<th>Fusion technique</th>
<th>Publication year</th>
<th>N</th>
<th>X seg</th>
<th>Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaughan et al</td>
<td>Intert</td>
<td>1976</td>
<td>121</td>
<td>7.3</td>
<td>29.8%</td>
</tr>
<tr>
<td>Eie</td>
<td>Smith-Petersen modified</td>
<td>1978</td>
<td>244</td>
<td>6</td>
<td>6%</td>
</tr>
<tr>
<td>Takeshima et al</td>
<td>Mcnab and Dall</td>
<td>2000</td>
<td>95</td>
<td>6.6</td>
<td>17%</td>
</tr>
<tr>
<td>LaMont et al</td>
<td>Intert</td>
<td>1976</td>
<td>125</td>
<td>4</td>
<td>45%</td>
</tr>
<tr>
<td>White et al</td>
<td>Intert</td>
<td>1987</td>
<td>69</td>
<td>5</td>
<td>56%</td>
</tr>
<tr>
<td>Mixter and Barr</td>
<td>H-graft</td>
<td>1967</td>
<td>380</td>
<td>10</td>
<td>41%</td>
</tr>
<tr>
<td>Frymoyer et al</td>
<td>H-graft and Hibbs</td>
<td>1978</td>
<td>207</td>
<td>13.7</td>
<td>21%</td>
</tr>
</tbody>
</table>

Abbreviations: intert, intertransverse; loss, segment loss in the operated group; X seg, average follow-up time in years.

There was substantial statistical heterogeneous results (Q value = 19.20; DF[Q] = 5; P = 0.002; I² = 73.96). The analysis of data was performed by a random effect model.

Of the seven studies, only one used instrumentation as adjunct to arthrodesis. All other authors used arthrodesis without instrumentation. The instrumentation system used by White et al was composed of a rod distraction, now recognized as harmful because of the possibility of a loss of lordosis and the possible creation of lumbar kyphosis. The results were reassessed excluding the work of White.

Analysis of the six papers without instrumentation (Figure 2)

The use of instrumentation lowered the difference favoring fusion against only discectomy. With the exclusion of instrumented results, the average of unsatisfying results in the fusion group was 18.03% versus 25.61% in the group without fusion. The reduction in absolute risk of unsatisfactory results was 7.5% and NNT of 13 patients (CI 95%: –0.219 to –0.026). The difference of results can be observed comparing the significance level of meta-analysis in Figures 1 and 2.

Results of the outcome “recurrence of disk herniation at the same level or in another level” (for sake of simplification here described simply as recurrence). Four studies compared the rate of recurrence between the experimental and control groups: Vaughan et al, Eie, Takeshima et al, and Mixter and Barr (Table 3, Figure 3). These outcomes were assessed as a combined risk for recurrence for comparison between papers.

Vaughan et al reported the numbers of progressive degenerative disk disease without a precise definition of this entity in both groups beyond the rates of recurrent disk protrusion. For this paper the results were pooled together as a recurrence of disk herniation.

Frymoyer et al described five patients with degeneration at adjacent levels in the experimental group but included relapsed and recurrent disks at other levels in the unsatisfactory results, which did not allow for comparison.

Table 2 Risk of unsatisfactory result with fusion group or discectomy only group (disk)

<table>
<thead>
<tr>
<th>Author</th>
<th>Treat</th>
<th>Uns</th>
<th>Sat</th>
<th>χ²</th>
<th>P</th>
<th>RR (CI 95%)</th>
<th>RD (CI 95%)</th>
<th>NNT (CI 95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaughan et al</td>
<td>Fusion</td>
<td>5</td>
<td>28</td>
<td>18.83</td>
<td>&lt;0.00001</td>
<td>0.246 (0.107–0.568)</td>
<td>0.46 (0.284–0.644)</td>
<td>2 (2–4)</td>
</tr>
<tr>
<td></td>
<td>Disk</td>
<td>32</td>
<td>20</td>
<td>1.797</td>
<td>0.18</td>
<td>0.627 (0.305–1.715)</td>
<td>0.09 (&lt;0.017–0.197)</td>
<td>11 (NNH = 5-NNT = 59)</td>
</tr>
<tr>
<td>Eie</td>
<td>Fusion</td>
<td>10</td>
<td>56</td>
<td>0.773</td>
<td>0.37</td>
<td>0.647 (0.301–1.39)</td>
<td>0.096 (&lt;0.072–0.264)</td>
<td>10 (NNH = 4-NNT = 14)</td>
</tr>
<tr>
<td></td>
<td>Disk</td>
<td>43</td>
<td>135</td>
<td>0.452</td>
<td>0.501</td>
<td>0.739 (0.375–1.456)</td>
<td>0.068 (&lt;0.079–0.215)</td>
<td>15 (NNH = 5-NNT = 13)</td>
</tr>
<tr>
<td>Takeshima et al</td>
<td>Fusion</td>
<td>9</td>
<td>42</td>
<td>0.42</td>
<td>0.192</td>
<td>1.6 (0.057–3.108)</td>
<td>A increased risk</td>
<td>NNHarm (NNH) 5.0</td>
</tr>
<tr>
<td></td>
<td>Disk</td>
<td>13</td>
<td>20</td>
<td>1.702</td>
<td>0.327</td>
<td>0.613 (0.304–1.238)</td>
<td>0.018 (&lt;0.042–0.409)</td>
<td>26 (NNH = 11-NNT = 56)</td>
</tr>
<tr>
<td>LaMont et al</td>
<td>Fusion</td>
<td>13</td>
<td>203</td>
<td>1.396</td>
<td>0.237</td>
<td>0.802 (0.536–1.2)</td>
<td>0.074 (&lt;0.066–0.215)</td>
<td>13 (NNH = 5-NNT = 15)</td>
</tr>
<tr>
<td></td>
<td>Disk</td>
<td>16</td>
<td>147</td>
<td>1.001</td>
<td>0.371</td>
<td>0.802 (0.536–1.2)</td>
<td>0.076 (0.029–0.122)</td>
<td>13 (8–34)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; NNH, number needed to treat to cause harm; NNT, number needed to treat to obtain benefit; RD, risk difference; RR, relative risk; Sat, number of satisfactory results; Treat, treatment type; Uns, number of unsatisfactory results; χ², value obtained with the χ² test.
The average risk of recurrence in the fusion group was 1.8% and in the discectomy group was 12.125 (risk difference = 10% [CI 95%: 0.07 to 0.135], NNT = 10 [CI 95%: 7 to 14]).

The relative risk of recurrence in the fusion group in relation to discectomy group was 0.156 (Figure 3).

**Results for the outcome “back pain after surgery”**

Vaughan et al\(^a\) and Takeshima et al\(^a\) described the number of patients with low back pain in the compared groups (Figure 4).

Vaughan et al\(^a\) described that 18 patients (34%) in the nonfusion group (N = 52) developed progressive degenerative disk disease with chronic back pain, whereas one patient (3%) in the fusion group (N = 33), among five patients with unsatisfactory results, developed progressive degenerative disk disease with chronic back pain.

Takeshima et al\(^a\) reported that all patients reported low back pain before surgery in both groups. In the nonfusion group, postoperative low back pain was noted in 27 among 51 fusion patients (53%) and in 18 among 51 fusion patients (35%) at follow-up. Many of them reported stiffness or dullness in the lower back.

Takeshima et al\(^a\) also described the severity of postoperative low back pain as evaluated by Japanese Orthopaedic Association score (in the nonfusion group, 2.3 ± 0.6; in the fusion group, 2.6 ± 0.6 [P = 0.007]). The overall risk ratio of developing chronic low back pain between the fusion and nonfusion groups in these two studies was 0.56 (95% CI: 0.34 to 0.807; P = 0.03) (Figure 4). The risk difference in favor of the fusion group was 0.297 (CI: 0.18 to 0.41; P < 0.05). The NNT was 3.

**Discussion**

Surgery for lumbar sciatic pain due to lumbar disk prolapse aims primarily at root decompression and the treatment of associated low back pain.

Disk herniation is the final phase of the degenerative process. The herniated disk loses its functions of mechanical load support in the spine. Studies on the determinant effect of disk degeneration have shown the predominance of genetic determinism rather than mechanical and postural factors.\(^7\) The degenerative process begins in the lower part

---

**Table 3 Risk of disk herniation recurrence**

<table>
<thead>
<tr>
<th>Author</th>
<th>Treat</th>
<th>Rec</th>
<th>NR</th>
<th>(\chi^2)</th>
<th>P</th>
<th>RR (CI 95%)</th>
<th>RD (CI 95%)</th>
<th>NNT (CI 95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaughan et al(^a)</td>
<td>Fusion</td>
<td>1</td>
<td>32</td>
<td>18.43</td>
<td>&lt;0.0001</td>
<td>0.061 (0.009–0.42)</td>
<td>0.47 (0.32–0.618)</td>
<td>2 (2–3)</td>
</tr>
<tr>
<td></td>
<td>Disk</td>
<td>26</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eie(^b)</td>
<td>Fusion</td>
<td>2</td>
<td>66</td>
<td>1.91</td>
<td>0.16</td>
<td>0.327 (0.078–1.36)</td>
<td>0.061 (0.005–0.116)</td>
<td>16 (9–200)</td>
</tr>
<tr>
<td></td>
<td>Disk</td>
<td>19</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takeshima et al(^a)</td>
<td>Fusion</td>
<td>2</td>
<td>49</td>
<td>3.698</td>
<td>0.054</td>
<td>0.216 (0.048–0.963)</td>
<td>0.143 (0.017–0.268)</td>
<td>7 (4–59)</td>
</tr>
<tr>
<td></td>
<td>Disk</td>
<td>8</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixter and Barr(^a)</td>
<td>Fusion</td>
<td>2</td>
<td>215</td>
<td>0.593</td>
<td>0.441</td>
<td>0.376 (0.07–2.026)</td>
<td>0.015 (0.012–0.042)</td>
<td>65 (NNH = 24–NNT = 83)</td>
</tr>
<tr>
<td></td>
<td>Disk</td>
<td>4</td>
<td>159</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Fusion</td>
<td>7</td>
<td>362</td>
<td>29.27</td>
<td>&lt;0.0001</td>
<td>0.156 (0.072–0.339)</td>
<td>0.102 (0.07–0.135)</td>
<td>10 (7–14)</td>
</tr>
<tr>
<td></td>
<td>Disk</td>
<td>57</td>
<td>413</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** CI, confidence interval; NNH, number needed to treat to cause harm; NNT, number needed to treat to obtain benefit; NR, number of patients without recurrence; RD, risk difference; RR, relative risk; Rec, recurrence number of patients; Treat, treatment type; \(\chi^2\), value obtained with the \(\chi^2\) test.
of the lumbar spine and rises upward continuously during life, toward the upper lumbar spine.

Although it is not possible to prevent degenerative disease, its consequences are potentially treatable. There is controversy about the best treatment for long-term disk prolapse surgery. In any surgical treatment in various medical specialties, there will always be a continuing search for treatment techniques that are less invasive. Microsurgery for lumbar disk herniation is an established surgery with excellent results. The evolution of the degenerative process could decrease the rate of long-term satisfaction of surgery for disk herniation. The potential benefit or risk of long-term fusion in the treatment of disk prolapse remains controversial.

Most of the papers describe the nosologies regarding degenerative lumbar spine diseases in a more broad and unspecific fashion, mostly with the term “degenerative disk disease”.

There was no prospective randomized clinical trial comparing both treatments. All of the seven recovered papers analyzed were retrospective. The average follow-up was at least 6.4 years after the surgery. The studies still had large average segment loss (>20%). The papers were not recent, ranging in publication between 1967 and 2000. The clinical characteristics samples were composed of homogeneous groups of patients with disk herniation and sciatica in all groups. They determined the level of evidence as 2 and grade of recommendation as B. The seven papers provided information sufficient to calculate the number of patients with satisfactory and unsatisfactory results, for isolated and pooled analysis of data, and for determining the treatment risks and benefits.

Patients treated with arthrodesis had more satisfactory results and fewer relapses and recrudescence. Patients not undergoing fusion had a higher risk of recurrence and progression of the degenerative process.

Low back pain was evaluated only in two manuscripts, and the risk associated with development of chronic low back pain was small in the fusion group. The evaluation of results revealed that there was a statistically significant difference favoring the fusion group between both kinds of treatments.

The works recovered did not study the effect of instrumentation, but White et al did so with an old distraction system, which is now obsolete in the lumbar spine. Therefore, their results were worse in the instrumented group than the results with simple discectomy (Figures 1 and 2). All other authors used in situ arthrodesis (Table 1). The role of modern instrumentation remains unclear.

The results of this analysis were based on papers with a limited or conflictive level of evidence and a low level of recommendation. Therefore, this review reveals the quality of published data and the gap in literature concerning this
well-defined pathology in the last decades. Though it is not possible to reveal a high degree of recommendation in favor of arthrodesis, but rather conflicting evidence of the fusion effect, the published papers were very homogeneous in design and reproduced familiar, reproducible, and validated outcomes to be pooled: satisfactory rate, recurrence rate, and residual low back pain rate with each procedure. Therefore, until randomized trials, arthrodesis remains an option in the treatment of lumbar disk prolapse surgery, because the conflicting evidence is not sufficient to recommend fusion as a routine treatment.

Of course, there are methodological limitations when performing meta-analysis with retrospective data, and this must be done only in the absence of randomized data and when the outcome results permit.

**Conclusion**

In conclusion, the treatment of lumbar disk prolapse with sciatica has not been studied since 2000, and there is no randomized trial evaluating outcomes. Nowadays, surgery has been changed to a minimally invasive model and, at this time, microdiscectomy remains the standard approach against which any other surgery must be compared. Future research should state with adequate methodology the role of fusion in protecting the spine for evolving degenerative process in the following decades in patients harboring lumbar disk prolapse.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**