FINAL APPRAISAL DOCUMENT

SPINAL FUSION FOR CERVICAL DEGENERATIVE DISC DISEASE

February 21, 2013

Daniel A. Ollendorf, MPH, ARM
Jennifer A. Colby, PharmD
Christopher Cameron, MSc
Swetha Sitaram, MS
Steven D. Pearson, MD, MSc, FRCP

Chief Review Officer
Sr. Research Associate
Decision Scientist
Research Associate
President
# CONTENTS

ABOUT ICER ................................................................................................................................. 3  
EXECUTIVE SUMMARY .................................................................................................................. 4  
ICER INTEGRATED EVIDENCE RATING – Fusion v Conservative Management .. 27  
ICER INTEGRATED EVIDENCE RATING – Fusion v Discectomy/Foraminotomy . 29  
FULL APPRAISAL REPORT .............................................................................................................. 36  
  
  Background .................................................................................................................................. 38  
  The Alternative Treatment Strategies ......................................................................................... 39  
  Clinical Guidelines ....................................................................................................................... 46  
  Medicare and Representative Private Insurer Coverage Policies ............................................ 49  
  Previous Systematic Reviews/Tech Assessments ...................................................................... 50  
  Ongoing Clinical Studies ............................................................................................................ 51  
  The Evidence ............................................................................................................................... 53  
  Clinical Benefits .......................................................................................................................... 66  
  Potential Harms ........................................................................................................................... 73  
  Differential Effectiveness & Safety of Cervical Fusion in Key Patient Subgroups .................... 78  
  Decision Analytic Model ............................................................................................................. 82  
  Recommendations for Future Research ......................................................................................... 99  
  
REFERENCES .................................................................................................................................. 101  
  
APPENDIX A .................................................................................................................................. 118  
APPENDIX B.................................................................................................................................. 126
ABOUT ICER

The Institute for Clinical and Economic Review (ICER), based at the Massachusetts General Hospital’s Institute for Technology Assessment (ITA), provides independent evaluation of the clinical effectiveness and comparative value of new and emerging technologies. ICER’s mission is to lead innovation in comparative effectiveness research through methods that integrate evaluations of clinical benefit and economic value. By working collaboratively with patients, clinicians, manufacturers, insurers and other stakeholders, ICER develops tools to support patient decisions and medical policy that share the goals of empowering patients and improving the value of healthcare services.

ICER’s academic mission is funded through a diverse combination of sources; funding is not accepted from manufacturers or private insurers to perform reviews of specific technologies. Since its inception, ICER has received funding from the following sources:

- Aetna Foundation
- The Agency for Healthcare Research & Quality (AHRQ)
- America’s Health Insurance Plans (AHIP)
- Amgen, Inc.
- Blue Cross Blue Shield of Massachusetts
- Blue Shield of California Foundation
- Greater Boston Chamber of Commerce
- Harvard Pilgrim Health Care
- HealthPartners
- The John W. Rowe Family Foundation
- Johnson & Johnson
- Kaiser Permanente
- Merck & Co.
- The National Pharmaceutical Council
- Philips Healthcare
- Robert Wood Johnson Foundation
- United Health Foundation
- The Washington State Health Care Authority

More information on ICER’s mission and policies can be found at www.icer-review.org.
EXECUTIVE SUMMARY

Introduction
Chronic neck pain is a prevalent and costly disorder. Approximately 15-20% of adults report at least one episode of neck pain during a given year, and nearly half of these individuals seek care (Carroll, 2008). On an annual basis, it is estimated that 11-14% of workers will have some limitation in their activities due to neck pain (Côté, 2008). While no recent studies have been conducted in the US on the economic burden of neck pain specifically, the combined burden of neck and back disorders in this country has been estimated to total $86 billion (Deyo, 2008).

One of the common causes of chronic neck pain is the progression of degenerative disc disease (DDD) of the cervical spine, a natural consequence of aging that results in gradual deterioration of cervical intervertebral discs (Emery, 2001). As the ability of these discs to absorb the shock and stress of vertebral motion declines, they become inelastic and cause a settling of the spinal column structure and abnormal spinal motion patterns. This process may in turn cause the development of abnormal bony growths and/or spurs (spondylosis), osteoarthritis, and/or herniation of one or more cervical discs. All of these conditions may in turn cause radiculopathy, or peripheral nerve root impingement. Symptoms of cervical radiculopathy include neck and arm pain, and weakness, tingling, or numbness in the upper extremities (Mayo Clinic, 2012). Less commonly, cervical DDD progression and its sequelae may directly compress parts of the spinal cord (myelopathy), affecting gait and balance in addition to causing arm and/or leg weakness and numbness.

A variety of management options are available to treat cervical neck pain and related symptoms arising from degenerative disc disease. These options are described in detail in the main body of the report. Briefly, the major options include:

- **Conservative Treatment**
  - Physical Therapy
  - Cervical Collar Immobilization
  - Spinal Manipulation (i.e., chiropractic or manual physical therapy)
  - Medication (e.g., analgesics, muscle relaxants, opioids)
  - Alternative Treatments (e.g., yoga, acupuncture)
  - Self-Care (e.g., educational materials, home stretching)

- **Spinal Injections**
  - Steroids
  - Nerve Blocks
  - Chemonucleolysis
  - Other (e.g., Botulinum toxin)
• **Minimally-Invasive Procedures**
  - Radiofrequency Denervation
  - Coblation Nucleoplasty

• **Surgical Procedures**
  - Spinal Fusion
  - Discectomy
  - Foraminotomy
  - Laminectomy/Laminoplasty

The most common surgical procedure employed in the U.S. for patients with symptomatic cervical DDD is spinal fusion (Cowan, 2006). The rate of spinal fusion has increased dramatically in recent years; an analysis of U.S. hospital discharge data from 1990-2004 showed an 8-fold increase in the utilization of anterior fusion procedures, even while the overall rate of hospital admissions for cervical DDD remained steady (Marawar, 2010). The cost of these procedures, as well as questions regarding the short- and long-term outcomes associated with fusion, have raised considerable interest in understanding the evidence on the relative effectiveness of this procedure in comparison to other management options.

**Appraisal Scope**
This appraisal sought to evaluate the comparative clinical effectiveness and comparative value of spinal fusion and its alternatives in patients with cervical degenerative disc disease (DDD). The scope of this appraisal is summarized in the Analytic Framework figure below.

**Analytic Framework: Management Options for Cervical Degenerative Disc Disease**
There are limited data directly demonstrating the impact of most cervical DDD management strategies on summary measures of “treatment success” or “successful clinical outcome”, so judgments about the effectiveness of these interventions must rest primarily upon consideration of multiple and potentially overlapping measures (e.g., pain, function, quality of life) as well as evaluation of treatment-associated risks. In addition, various stakeholders will by nature be more interested in certain outcomes than others. For example, payers and employers may be most interested in functional improvement and/or return to work, while clinicians and patients may focus more on relief of symptoms and spinal stability.

The focus of this appraisal was on adults (>17 years of age) with cervical DDD symptoms, including neck pain, arm pain, and/or radiculopathic symptoms (e.g., numbness, tingling); these symptoms could occur with or without the presence of spondylosis. We did not focus on studies that consisted primarily or in total of patients whose primary complaint was cervical myelopathy, as this is generally considered a neurologic emergency in all but the mildest cases and patients typically proceed directly to surgical intervention (McCormick, 2003); included studies may have involved a minority subgroup of patients with myelopathy, however. In all cases, the target population was focused on patients whose symptoms have persisted despite an initial short course (i.e., 4-6 weeks) of self-care and conservative management.

Evidence was sought to answer the following key questions:

1. What is the comparative clinical effectiveness of cervical fusion for DDD relative to that of conservative management approaches, minimally-invasive procedures, and other forms of surgery?

2. What are the adverse events and other potential harms associated with cervical fusion compared to conservative management approaches, minimally-invasive procedures, and other forms of surgery?

3. What is the differential effectiveness and safety of cervical fusion according to factors such as age, sex, race or ethnicity, pre-existing conditions (e.g., smoking history), neuromuscular disease states (e.g., Parkinsonism), measurable spinal instability, technical approach to fusion, insurance status (e.g., worker’s compensation vs. other), and treatment setting (e.g., inpatient vs. ambulatory surgery)?

4. What are the costs and potential cost-effectiveness of cervical fusion relative to alternative approaches?

Outcomes of interest included measures of pain, function, health-related quality of life, and employment status. In addition, information was obtained on standardized or study-specific measures of “treatment success” or “successful clinical outcome”. Potential harms of interest included perioperative (occurring during the operative episode or the 30 days following) mortality, major and minor complications, and procedure revision, as well as
long-term mortality, adverse events, and requirements for subsequent surgery or additional treatment.

Information was obtained from all randomized controlled trials (RCTs) and comparative cohort studies that compared fusion to one of the comparators described above. Importantly, studies that only compared different variants of fusion (e.g., according to instrumentation or graft type employed) were excluded from consideration, as numerous systematic reviews have not found evidence distinguishing these approaches. Studies that compared anterior to posterior anatomic approaches to fusion as well as single-level, 2-level, and multi-level fusion were included, however, as evidence suggests that rates of mortality and certain complications may differ between these approaches (Shamji, 2008; Riley, 2010). Additionally, the Washington HCA was interested in examining whether fusion outcomes differed by surgical setting (i.e., inpatient vs. ambulatory/outpatient).

Data on harms and/or subgroups of interest were also obtained from large (>50 patients), long-term (≥12 months of follow-up) case series evaluating cervical fusion.

**Study Quality**

We used standardized criteria specific to previous systematic reviews in back pain to rate the quality of each included RCT; these criteria have also been widely adopted for use in studies of neck pain (Nikolaidis, 2010). The criteria, which related to issues of study design, reporting, and minimization of bias, are presented in Appendix A. RCTs meeting a majority of criteria (i.e., 5 of 9) were deemed to be “higher quality”. We used general criteria to assess the quality of comparative cohort studies, using the categories “good”, “fair”, or “poor”, based on the criteria employed by the U.S. Preventive Services Task Force (AHRQ, 2008) to assess these studies in terms of comparability of study populations, retention of patients during follow-up, use of standardized and/or validated outcomes, and level of attention paid to confounders.

Study quality was not assessed for case series, as the focus of quality ratings was on the level of bias in assessing the comparative impact of fusion vs. alternative treatments on measures of effectiveness and harm.

Data on costs and potential cost-effectiveness were obtained primarily from a newly-developed decision-analytic model that simulated and compared multiple treatment pathways in patients with symptomatic cervical DDD, as described in the “Comparative Value” section on page 20.

**Evidence on Comparative Clinical Effectiveness (KQ 1)**

**Overview of Evidence & Data Quality**

Of the 14 RCTs identified (total N=1,209), nearly all (13) focused entirely on patients with symptoms and radiographic evidence of cervical radiculopathy; 1 included patients with evidence of disc herniation. Most available RCTs limited patients to those with single- or 2-level disease. **Importantly, we found no RCT evidence for spinal fusion procedures in**
patients with only generalized neck pain. Sample sizes were generally small, ranging from 10-50 patients per treatment arm.

The 7 comparative cohort studies included 929 patients evaluated in single- and multicenter studies, as well as nearly 100,000 patients assessed in a retrospective evaluation of the Nationwide Inpatient Sample database maintained by the U.S. Agency for Healthcare Research and Quality (AHRQ) (Shamji, 2008). Of the 7 studies, 1 was prospective.

Study quality is presented in Table ES1 below by study type and fusion comparator. A total of 10 of the 14 RCTs were identified as higher quality, including 1 comparing fusion to conservative management, 3 with minimally-invasive surgery as a comparator, and 6 comparing fusion to alternative open surgical approaches. Among the 7 comparative cohort studies, none were identified as good quality. Four studies were deemed to be of “fair” quality, including 1 comparing fusion to conservative management, 1 comparing fusion to foraminotomy, 1 comparing fusion procedures performed in inpatient vs. outpatient settings, and 1 comparing anterior to posterior fusion techniques.

Table ES1. Studies of cervical fusion: study quality, by type of study and patient population.

<table>
<thead>
<tr>
<th>Study Type</th>
<th>Comparator</th>
<th>Higher</th>
<th>Lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCT</td>
<td>Physical Therapy/Cervical Collar</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Discectomy</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Discectomy/Foraminotomy</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Microdiscectomy</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Endoscopic discectomy</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Endoscopic foraminotomy</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Comparative Cohort</td>
<td>Interdisciplinary Rehabilitation</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Laminoplasty</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Foraminotomy</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Fusion (in- vs. outpatient)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Fusion (anterior vs. posterior)</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Findings are organized by type of comparator to fusion in the sections that follow. No comparative data were available comparing fusion to minimally-invasive nonsurgical management options such as spinal injections, radiofrequency denervation, or coblation nucleoplasty.
Spinal Fusion vs. Conservative Treatment
A total of 2 studies, an RCT (Persson, 2001) and a cohort study (Mayer, 2002) compared cervical spinal fusion to conservative management approaches. These studies are summarized by outcome and population in the sections below.

“Treatment Success”
Neither of the studies comparing fusion to conservative treatment included measures of treatment success.

Pain and Function
A higher-quality RCT comparing outcomes for patients with cerviobrachial pain of > 3 months’ duration and nerve root compression due to spondylitic spurs randomized 81 patients into equal groups receiving anterior discectomy with fusion, physical therapy, or cervical collar immobilization; patients were followed for 16 months (Persson, 2001). Pain was evaluated on a 100 mm VAS scale and measured at baseline, 4 months, and 16 months. As noted in Figure ES1 below, pain was reduced at 4 months among fusion patients vs. those receiving either form of conservative management; for cervical collar immobilization, this difference was statistically significant. By month 16, however, the gap in pain scores between conservative management and fusion had narrowed, and VAS scores did not differ statistically at this timepoint. (NOTE: 5-10 point changes on VAS scores represent the minimum change that would be considered “clinically important”, and changes ≥ 30% from baseline represent those that would involve significant improvements for patients).

Figure ES1. Self-rated current pain on visual analogue scale, fusion vs. physical therapy vs. cervical collar.

*: p<.01, fusion vs. collar; all other comparisons not statistically significant
Data on pain and function were also available from the comparative cohort study prospectively evaluating the effects of anterior fusion in combination with an interdisciplinary rehabilitation program compared to interdisciplinary rehabilitation alone in 202 patients filing worker’s compensation claims for cervical spinal disorders (Mayer, 2002). Rehabilitation consisted of medically-supervised exercise, psychological counseling, workplace and vocational services, and case management. Patients were followed for 12 months. No statistically-significant differences between groups were noted during follow-up in measures of pain, including a 10 cm VAS scale and a questionnaire-based VAS scale known as the Million VAS (Anagnostis, 2003). In addition, no statistically-significant differences were observed in either mean cumulative scores for physical function or the percentage of patients with total cumulative functional scores demonstrating a “significant degree of pathology” (Mayer, 2002).

Quality of Life
In the Persson RCT, no statistically-significant differences between fusion and physical therapy were observed in Sickness Impact Profile or Mood Adjective Check List scores at any point during follow-up (Persson, 2001). Quality of life was also evaluated in the cohort study comparing fusion and interdisciplinary rehabilitation to rehabilitation alone (Mayer, 2002); however, such evaluation focused only on mean Beck Depression Inventory scores at 12 months without controlling for baseline levels or change from baseline.

Return to Work
Work-related outcomes were not evaluated in the Persson RCT. In the Mayer comparative cohort study of fusion vs. interdisciplinary rehabilitation, no statistically-significant differences were observed between treatment arms in the percentages of patients returning to full or modified work, returning to work at the same employer, or filing additional claims for recurrent injury at 12 months (Mayer, 2002).

Spinal Fusion vs. Discectomy or Foraminotomy
A total of 13 RCTs (9 higher-quality) examined the effects of fusion compared to discectomy or foraminotomy for cervical DDD. In 10 of these, the comparator was discectomy, microdiscectomy, or endoscopic discectomy alone. In one RCT, separate treatment arms receiving discectomy or foraminotomy were included (Wirth, 2000), and in another, comparator patients received a combination of discectomy and foraminotomy (Martins, 1976). Patients in a third RCT received endoscopic foraminotomy (Ruetten, 2008). Characteristics of all RCTs comparing fusion to these alternative surgical approaches are summarized in Table ES2 on the following page. Fusion was compared to foraminotomy in a single comparative cohort study of 292 patients who were treated for radicular symptoms and followed for 6 years (Korinth, 2006).

“Treatment Success”
Treatment success was evaluated in 6 of the 9 higher-quality RCTs. In one, 3 types of discectomy with fusion (autograft, polymethylmethacrylate [PMMA] graft, and titanium cage) were compared to microdiscectomy alone in 125 patients with single-level cervical disc disease and radiculopathy who were treated at a single institution in Switzerland and
followed for 12 months (Barlöcher, 2002). Based on Odom’s criteria, a greater percentage of patients undergoing fusion with titanium cage had “excellent” or “good” results vs. microdiscectomy alone at 12 months (94.4% vs. 75.5%, p<.02); differences were not statistically-significant for the other fusion groups compared to microdiscectomy. No statistically-significant differences were observed for measures of treatment success in the other higher-quality RCTs.

The retrospective comparative cohort study also evaluated outcome based on Odom’s criteria. In this study, an assessment of 292 patients receiving either PMMA fusion or posterior foraminotomy (Korinth, 2006), long-term outcome was assessed after a mean of 6 years. The number of patients reporting excellent or good outcome was statistically-significantly greater in the fusion group (93.6% vs. 85.1%, p<.05).

Table ES2. Characteristics of RCTs comparing fusion to discectomy or foraminotomy.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Comparator(s)</th>
<th>Sample Size</th>
<th>Study Duration (months)</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abd-Alrahman</td>
<td>1999</td>
<td>Discectomy</td>
<td>90</td>
<td>15</td>
<td>Lower</td>
</tr>
<tr>
<td>Barlöcher</td>
<td>2002</td>
<td>a--Microdiscectomy b--PMMA Fusion c--Fusion w/Cage</td>
<td>125</td>
<td>12</td>
<td>Higher</td>
</tr>
<tr>
<td>Dowd</td>
<td>1999</td>
<td>Discectomy</td>
<td>84</td>
<td>54</td>
<td>Lower</td>
</tr>
<tr>
<td>Hauerberg</td>
<td>2008</td>
<td>Discectomy</td>
<td>86</td>
<td>24</td>
<td>Higher</td>
</tr>
<tr>
<td>Martins</td>
<td>1976</td>
<td>Discectomy and foraminotomy</td>
<td>51</td>
<td>12</td>
<td>Higher</td>
</tr>
<tr>
<td>Oktenoglu</td>
<td>2007</td>
<td>Microdiscectomy</td>
<td>20</td>
<td>18</td>
<td>Higher</td>
</tr>
<tr>
<td>Rosenørn</td>
<td>1983</td>
<td>Discectomy</td>
<td>63</td>
<td>12</td>
<td>Higher</td>
</tr>
<tr>
<td>Ruetten</td>
<td>2009</td>
<td>Endoscopic discectomy</td>
<td>120</td>
<td>24</td>
<td>Higher</td>
</tr>
<tr>
<td>Ruetten</td>
<td>2008</td>
<td>Endoscopic foraminotomy</td>
<td>200</td>
<td>24</td>
<td>Lower</td>
</tr>
<tr>
<td>Savolainen</td>
<td>1998</td>
<td>a--Discectomy b--Fusion w/Plating</td>
<td>91</td>
<td>36</td>
<td>Higher</td>
</tr>
<tr>
<td>van den Bent</td>
<td>1996</td>
<td>Discectomy</td>
<td>81</td>
<td>24</td>
<td>Higher</td>
</tr>
<tr>
<td>Wirth</td>
<td>2000</td>
<td>a--Discectomy b--Foraminotomy</td>
<td>72</td>
<td>60</td>
<td>Lower</td>
</tr>
<tr>
<td>Xie</td>
<td>2007</td>
<td>a -- Discectomy b -- Instrumented Fusion</td>
<td>45</td>
<td>24</td>
<td>Higher</td>
</tr>
</tbody>
</table>

We conducted a random-effects meta-analysis examining the likelihood of treatment success for fusion vs. discectomy based on data from 2 studies using Odom’s criteria to
assess treatment success. No statistically-significant difference was observed, as noted in Figure ES2 below (Rate ratio [RR]: 0.98, 95% CI: 0.81, 1.18; p=.84).

**Figure ES2. Meta-analysis of treatment success, based on Odom’s criteria.**

<table>
<thead>
<tr>
<th>Study/Subgroup</th>
<th>Fusion* Events Total</th>
<th>Discectomy Events Total</th>
<th>Total Weight</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barlöcher 2002</td>
<td>24 30 25 33 50.5%</td>
<td>1.06 [0.81, 1.37]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>van den Bent 1996</td>
<td>28 40 30 39 49.5%</td>
<td>0.91 [0.70, 1.19]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td><strong>70 72 100.0%</strong></td>
<td></td>
<td></td>
<td><strong>0.98 [0.81, 1.18]</strong></td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>52 55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Tau² = 0.00; Chi² = 0.61, df = 1 (P = 0.43); I² = 0%
Test for overall effect: Z = 0.20 (P = 0.84)

*Anterior discectomy and fusion

**Pain and Function**
Information on pain was available from 6 of the 9 higher-quality RCTs comparing fusion to alternative surgical procedures. No significant effects of treatment on pain were observed in 4 of these RCTs. In the van den Bent RCT comparing discectomy with PMMA fusion to fusion alone (van den Bent, 1996), the percentage of patients reporting relief of neck pain was statistically-significantly greater in the fusion group at 6 weeks (78% vs. 43% for discectomy, p=.04). Pain relief improved in the discectomy group thereafter, however, and differences were no longer statistically-significant for the remainder of the 24-month follow-up. The percentage of patients with any improvement in VAS neck or radicular pain was assessed at 2, 6, and 12 months in the RCT comparing 3 types of fusion to microdiscectomy alone (Barlöcher, 2002). At month 12, there was a statistically-significant difference in favor of fusion with titanium cage (97.3% vs. 81.9% for microdiscectomy alone, p<.05); no other measures differed statistically at any timepoint.

**Quality of Life**
Data on quality of life were found in the Xie RCT comparing fusion with or without instrumentation to discectomy (Xie, 2007). Quality of life was assessed via the SF-36 instrument; in repeated-measures analyses, no statistically-significant differences were noted between groups in SF-36 total scores as well as scores for individual domains.

**Return to Work**
Data on return to work were available from 4 higher-quality RCTs. The proportion of patients returning to work did not statistically differ at any timepoint between treatment groups in the previously-described Hauerberg and Xie RCTs. In a third RCT examining 63 patients undergoing anterior cervical discectomy with and without fusion who were followed for 12 months (Rosenørn, 1983), a statistically-significantly greater percentage of fusion patients had returned to work at 2, 3, 4, 6, and 9 weeks postoperatively; differences were no longer significant when measured at weeks 12, 26, and 52, however. Finally, the percentage of working individuals not yet able to return to work was assessed in the Barlocher RCT (Barlöcher, 2002). At 6 months, a statistically-significantly lower percentage of patients undergoing fusion with titanium cage were not yet able to work as compared to
microdiscectomy alone (5.5% vs. 18.1%, p<.05). However, differences were nonsignificant when evaluated at 12 months.

We also assessed return to work in a random-effects meta-analysis comparing fusion to discectomy in 4 higher-quality RCTs. It is important to note that we used timepoints for return-to-work data that were uniform across all available RCTs (i.e., 12-24 months); the “early” benefits of fusion as noted in some studies above were therefore not considered in this analysis. As shown in Figure ES3 below, the pooled estimate directionally favored discectomy in terms of return to work at 12-24 months, but this difference was not statistically significant. A second analysis focused return to work at 6 months following surgery in two of these RCTs (Barlöcher, 2002; Xie, 2007) also did not yield statistically-significant results.

Figure ES3. Meta-analysis of likelihood of return to work at 12-24 months, fusion vs. discectomy.

<table>
<thead>
<tr>
<th>Study/Subgroup</th>
<th>Fusion* Events</th>
<th>Discectomy Events</th>
<th>Total Events</th>
<th>Total Weight</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barlöcher 2002</td>
<td>25</td>
<td>30</td>
<td>55</td>
<td>41.8%</td>
<td>0.95 [0.77, 1.16]</td>
</tr>
<tr>
<td>Hauerberg 2007</td>
<td>11</td>
<td>36</td>
<td>47</td>
<td>5.0%</td>
<td>0.66 [0.37, 1.18]</td>
</tr>
<tr>
<td>Rosenorn 1983</td>
<td>24</td>
<td>31</td>
<td>55</td>
<td>39.5%</td>
<td>0.83 [0.67, 1.02]</td>
</tr>
<tr>
<td>Xie 2007</td>
<td>12</td>
<td>15</td>
<td>27</td>
<td>13.6%</td>
<td>0.96 [0.67, 1.37]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>72</td>
<td>89</td>
<td>161</td>
<td>100.0%</td>
<td>0.88 [0.77, 1.01]</td>
</tr>
</tbody>
</table>

Heterogeneity: Tau² = 0.00; Chi² = 2.43, df = 3 (P = 0.49); I² = 0%
Test for overall effect: Z = 1.85 (P = 0.06)

* Anterior discectomy and fusion

Potential Harms (KQ 2)

Cervical spinal fusion is associated with a number of harms, some of which are common to both fusion and other surgical alternatives, and others of which are unique to the fusion procedure itself (e.g., pseudarthrosis). Relevant harms are presented on Table ES3 on the following page, described in the sections that follow, and categorized according to their timing.

Information from the observational studies examined in this review suggests that risks of surgical interventions may be higher than reported in RCTs. For example, in the 14 RCTs examined for this appraisal, only 1 provided any data on perioperative mortality (Xie, 2007; rates were 0% in both treatment groups). In contrast, rates of in-hospital and 30-day mortality from large database studies, while <1%, were certainly nonzero (Shamji, 2008; Wang, 2007). Other harms that may not be apparent until after hospital discharge, such as pneumonia or venous thrombosis, appear relatively rarely in observational studies but are not reported at all in available RCTs.
Table ES3. Reported ranges of rates of potential harms from RCTs and comparative cohort studies, by type of study and comparator.

<table>
<thead>
<tr>
<th>Type of Harm</th>
<th>Fusion</th>
<th>Conservative Rx</th>
<th>Surgical Approaches</th>
<th>No. of studies reporting harms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RCT</td>
<td>CC</td>
<td>RCT</td>
<td>RCT</td>
</tr>
<tr>
<td><strong>Perioperative Events</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality</td>
<td>0</td>
<td>0-0.05</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Complications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Hemorrhage</td>
<td>NR</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>o Hematoma</td>
<td>1-6.6</td>
<td>0-0.8</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>o Nerve Damage*</td>
<td>2.5-8</td>
<td>0.8-6</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>o Paralysis</td>
<td>NR</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>o Infection</td>
<td>0-13</td>
<td>0-0.02</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>o Hoarseness</td>
<td>5-20</td>
<td>1.6</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>o Dysphagia</td>
<td>3-17.5</td>
<td>0-10</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>o Thrombosis</td>
<td>NR</td>
<td>0.02</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>o CSF Leak</td>
<td>NR</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Return to OR</td>
<td>NR</td>
<td>0</td>
<td>NA</td>
<td>NR</td>
</tr>
<tr>
<td><strong>Long term Events†</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Chronic pain</td>
<td>4.8</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>o ASD</td>
<td>6.9-16.6</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>o Pseudarthrosis</td>
<td>8</td>
<td>3.2</td>
<td>NA</td>
<td>NR</td>
</tr>
<tr>
<td>o Neurological Decline*</td>
<td>3-23.3</td>
<td>0</td>
<td>14.2</td>
<td>NR</td>
</tr>
<tr>
<td>o Myelopathy</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>o Muscle weak</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>o Paresthesia</td>
<td>14.2</td>
<td>3</td>
<td>8.2</td>
<td>NR</td>
</tr>
<tr>
<td>Subsequent Rx</td>
<td>0-5.21.7</td>
<td>0-3.2</td>
<td>13.8</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| NR: Not reported; NA: Not Applicable; ASD: Adjacent segment disease
| ♦: Conservative treatment = Physiotherapy; ♦♦: Conservative treatment = Interdisciplinary treatment; *: Nerve damage includes numbness, weakness and nerve palsy. **: Neurological decline includes sensory loss and neurological deficit; † Rates are annualized.
Data from RCTs and Comparative Cohorts

As mentioned previously, certain types of complications and adverse outcomes were not reported in any available RCT or comparative cohort study. For example, data on perioperative paralysis or hemorrhage were not available in any RCT or comparative cohort study, and information on thrombosis was only reported in comparative cohort studies of fusion. In addition, there was a significant degree of overlap between treatment alternatives in reported rates of complications, with no clear or discernible pattern of differential rates for most complication types.

As described previously, perioperative mortality was rarely identified or reported in any RCT or comparative cohort study. Among perioperative complications, the most frequently reported for fusion included dysphagia, hoarseness, and infection. Not surprisingly, generally higher rates of dysphagia and hoarseness were reported for fusion, given the increasing use of anterior surgical approaches for this procedure (Shamji, 2008); rates were similar to those reported for discectomy in available RCTs, however. The upper end of the range of infection rates was higher with fusion relative to other procedures, which was related to 2 cases of donor site infection in a small RCT (Xie, 2007). Leakage of cerebrospinal fluid was rarely reported for any type of procedure.

Rates of longer-term events were annualized to account for differential follow-up across studies. Mortality was again rarely reported in RCTs or comparative cohort studies with a significant degree of overlap by intervention. In some fusion studies, relatively high rates of adjacent segment disease and pseudarthrosis were reported. High rates of neurological decline were also reported in 1 RCT comparing fusion to discectomy, but these included measures of unchanged or worsened sensory loss in 1 RCT (Persson, 2001). Rates of repeat surgery or subsequent therapy were more commonly reported across study types. Again, ranges of reported rates overlapped significantly between them. The highest rate of reoperation was reported in the Persson RCT, where 8 of 30 patients underwent a second surgery by 16 months of follow-up (Persson, 1997); 6 of these had surgery adjacent to the initial surgical site.

Data from Fusion Case Series

Long-term data on harms were reported in 46 reports of case series, describing events in nearly 6,000 patients. Follow-up ranged from 1 to 21 years in these studies. The most frequently-documented events included reoperation (n=23), pseudarthrosis (n=19), and adjacent segment disease (n=13), with ranges similar to those reported in RCTs and comparative cohorts. Mortality data were reported in 9 studies; rates ranged from 0-2.6% on an annualized basis.

In an effort to create an exhaustive list of long-term adverse events reported in patients undergoing cervical fusion, all reported harms from these studies are listed in Table ES4 on the following page. Recent data on risks associated with any surgical procedure are also provided for additional context within the main body of the report.
Table ES4. Reported ranges of annualized rates of potential harms from fusion case series.

<table>
<thead>
<tr>
<th>Type of Harm</th>
<th>Number of studies</th>
<th>% of patients with event (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>9</td>
<td>0-2.6</td>
</tr>
<tr>
<td>Complications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjacent segment disease</td>
<td>13</td>
<td>0-27</td>
</tr>
<tr>
<td>Arm pain</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>Donor site pain</td>
<td>1</td>
<td>8.7</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>2</td>
<td>3.6-33</td>
</tr>
<tr>
<td>Dysphonia</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>Hepatitis C infection</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>Hoarseness</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>Laryngeal paresis</td>
<td>2</td>
<td>0.5-1.4</td>
</tr>
<tr>
<td>Neck pain</td>
<td>3</td>
<td>0.3-2.5</td>
</tr>
<tr>
<td>New-onset radiculopathy</td>
<td>2</td>
<td>0.8-5.0</td>
</tr>
<tr>
<td>Paresis</td>
<td>1</td>
<td>1.3</td>
</tr>
<tr>
<td>Pseudarthrosis</td>
<td>19</td>
<td>0-8.5</td>
</tr>
<tr>
<td>Ptosis</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Tetraparesis</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>Upper limb numbness</td>
<td>2</td>
<td>0.3-2.2</td>
</tr>
<tr>
<td>Upper limb weakness</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>Worsening headache pain</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>Worsening lower extremity function</td>
<td>1</td>
<td>1.7</td>
</tr>
<tr>
<td>Worsening sensory function/strength</td>
<td>2</td>
<td>0.7-2.7</td>
</tr>
<tr>
<td>Worsening upper extremity function</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>Re-operation</td>
<td>23</td>
<td>0-6.8</td>
</tr>
<tr>
<td>Inability to return to work</td>
<td>2</td>
<td>0.5-4.3</td>
</tr>
</tbody>
</table>

**Differential Effectiveness/Safety for Key Patient Subgroups (KQ 3)**

Data examining the differential effects of cervical fusion in key patient subpopulations were obtained from RCTs, comparative cohort studies and case series where available. Findings are reported below by type of study and subgroup in the sections that follow.

**Randomized Controlled Trials**

**Single- vs. 2-level Surgery**

In an RCT comparing anterior cervical discectomy with fusion to discectomy alone in 51 patients (Martins, 1976), the percentage of patients with excellent or good results (complete relief or minimal persistence of preoperative symptoms and abnormal signs improved or unchanged) was compared by level of surgery. Among patients undergoing single-level procedures, the percentage of those with excellent or good results was higher in the fusion group (82% vs. 66% for discectomy), while the rate for 2-level surgery was lower for fusion (26% vs. 63%). Rates were not tested statistically.
Smoking Status
In the RCT comparing fusion to physical therapy and cervical collar immobilization (Persson, 2001), the improvement in VAS pain among those undergoing surgery was found to be better among nonsmokers vs. smokers (p<.05), although the actual data on VAS changes among these subgroups were not provided.

Gender
In an RCT comparing anterior discectomy with and without fusion in 63 patients (40 men and 23 women) (Rosenorn, 1983), a higher percentage of males (n=16) undergoing fusion had excellent or good results (defined as return to previous occupation with no or minimal symptoms) at 12 months (15 [94%] vs. 19 [86%] for discectomy), while a lower percentage of females (n=13) had excellent or good results (5 [38%] vs. 8 [89%] for discectomy). While statistical testing was not performed for the stratified analysis across treatment groups, the percentage of patients with treatment success in the fusion group was greater (p<.005) among males vs. females.

Comparative Cohort Studies
Data were available from 4 comparative cohort studies; the focus of attention in the descriptions below is on the 2 studies that were fair-quality. In these studies, the subgroups of interest defined the study comparators (i.e., inpatient vs. outpatient fusion, posterior vs. anterior fusion).

Inpatient vs. Outpatient Fusion
The effects of fusion performed in inpatient vs. ambulatory care settings were assessed in a single fair-quality comparative cohort study. In this study, a comparison of 50 consecutive day-surgery patients to 53 retrospectively-analyzed inpatient controls (Silvers, 1996), no statistically-significant differences by setting were noted for functional outcomes, VAS pain, performance of activities of daily living, or return to work or normal activities. The rate of reoperation was numerically higher in the inpatient group (9.4% vs. 4.0% for outpatient), but this difference was not statistically tested.

Anterior vs. Posterior Fusion
Comparisons of anterior vs. posterior fusion techniques were performed in 1 fair-quality retrospective cohort study. This study examined differences between anterior and posterior fusion techniques using the U.S. Nationwide Inpatient Sample (NIS) database (AHRQ, 2012), and assessed hospital outcomes among nearly 100,000 patients undergoing anterior or posterior fusion for cervical DDD between 1988 and 2003 (Shamji, 2008). Patients were further stratified by whether their record included a diagnosis of myelopathy (approximately 75% of patients did not have such a diagnosis). On an unadjusted basis, patients undergoing posterior fusion experienced higher rates of death and major complications (p≤.001), regardless of the presence of myelopathy on the record. Length of stay and inflation-adjusted cost was also significantly increased, as shown in Table ES5 on the following page (data on myelopathy patients are displayed for context). After adjustment for differences in patient characteristics between groups, patients undergoing posterior fusion had significantly higher rates of all complications except for thrombophlebitis when compared to anterior fusion.
Table ES5. Outcomes of anterior vs. posterior fusion surgery, stratified by neurological status (n=96,773).

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Anterior</th>
<th>Posterior</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Myelo</td>
<td>Myelo</td>
<td>No Myelo</td>
</tr>
<tr>
<td>Death (%)</td>
<td>0.05</td>
<td>0.52</td>
<td>0.36</td>
</tr>
<tr>
<td>Pulmonary embolism (%)</td>
<td>0.02</td>
<td>0.06</td>
<td>0.14</td>
</tr>
<tr>
<td>Pneumonia (%)</td>
<td>0.14</td>
<td>0.62</td>
<td>1.04</td>
</tr>
<tr>
<td>Transfusion (%)</td>
<td>0.34</td>
<td>1.02</td>
<td>3.33</td>
</tr>
<tr>
<td>Thrombophlebitis (%)</td>
<td>0.02</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>Infection %</td>
<td>0.02</td>
<td>0.10</td>
<td>0.36</td>
</tr>
<tr>
<td>Length of stay (mean days)</td>
<td>1.95</td>
<td>3.42</td>
<td>4.37</td>
</tr>
<tr>
<td>Inflation-adjusted cost ($)</td>
<td>20,639</td>
<td>28,581</td>
<td>34,963</td>
</tr>
</tbody>
</table>

Source: Shamji et al., J Neurosurg: Spine, 2008;9:10-16
Myelo: Myelopathy diagnosis on inpatient record

While the results of this study suggest higher rates of complications with posterior fusion, it is often the case that contemporary posterior techniques are reserved for patients with greater disability and/or spinal instability who require multi-level intervention (Caridi, 2011). Information from administrative databases is understandably limited in controlling for differences in clinical presentation.

While not considered a “comparative cohort” study per se, an earlier study analyzing data from the NIS also showed higher rates of complications and mortality for posterior vs. anterior fusion (Wang, 2007); in multiple logistic regression analyses, posterior fusion was significantly associated with a higher risk of in-hospital complications (but not mortality).

**Fusion Case Series**
A total of 28 fusion case series stratified data according to key patient subgroups of interest; findings are presented in Appendix C, Table C23 and described in further detail below.

**Single- vs. Multi-Level Surgery**
Subgroup analyses of patients undergoing single- and multi-level fusion procedures were analyzed in 17 case series. In most of these studies, increases in the number of levels involved were associated with increased rates of pseudarthrosis, although the statistical significance of any observed differences was often not tested. Reoperation rates and development of adjacent segment disease were assessed in 3 studies (Matsumoto, 2009; Heidecke, 2000; Bishop, 1996); no statistically-significant differences in these rates according to number of levels involved were observed. In one series, rates of dysphagia were reported for patients undergoing 1-, 2-, and 3+ level anterior fusion; these rates increased according to the number of levels involved (11% vs. 24% vs. 43%, significance not tested) (Riley, 2005). Findings from a later systematic review by the same author were similar in nature (Riley, 2010).
Smoking Status
Six case series examined the impact of pre-operative smoking status on adverse events and clinical outcomes. While 1 study described statistically-significantly fewer cases of pseudarthrosis among non-smokers (20% vs. 50% for smokers, p=.001) (Goldberg, 2002), 2 found no correlation between smoking status and development of pseudarthrosis or adjacent segment disease (Matsumoto, 2009; Bindal, 2007). In terms of clinical outcomes, 3 series evaluated the effect of smoking status on treatment success using Odom’s criteria. In one, data from a series of 144 patients indicated that smokers had a significantly (p=.008) higher rate of fair or poor outcomes (Jensen, 2009), although actual percentages were not reported. Another study (n=190) found non-smokers to have a significantly higher rate of excellent outcomes (43.0% vs. 27.3%, p<.03) (Hilibrand, 2001). A third smaller series (n=66) found no statistically-significant differences in this measure (Samartzis, 2005).

Gender
Clinical outcomes and adverse events did not statistically differ according to patient gender in 4 case series (Chen, 2009; Matsumoto, 2009; Bindal, 2007; Javid, 2001). Data from an additional series did find statistically-significantly greater rates of dysphagia (41.2% vs. 22.9% for women vs. men, p<.05) and dysphonia (28.2% vs. 8.6%, p<.05) among women (Yue, 2005b).

Anterior vs. Posterior Fusion
Fusion approach was evaluated in a single case series of 120 patients undergoing revision procedures for pseudarthrosis (Carreon, 2006). Over 3-4 years of follow-up, the rate of subsequent reoperation was lower among patients undergoing posterior revision (2.2% vs. 44.4%), although this difference was not tested statistically.

Duration of Symptoms
Four case series analyzed the impact of symptom duration on clinical prognosis and long-term outcomes. While 1 study found no correlation (Chen, 2009), 3 case series demonstrated significantly better outcomes in patients with shorter duration of symptoms (Kadoya, 2003; Hamburger, 2001; Heidecke, 2000). In one study, patients with a symptom duration of less than 3 months had a statistically-significantly higher rate of excellent outcomes based on Odom’s criteria as compared to those with symptoms > 12 months (48.9% vs. 33.3%, p<.03) (Hamburger, 2001). In a study of anterior microdiscectomy with fusion, no significant difference in the percentage of patients with a self-reported “good” outcome was seen in patients with radiculopathy symptoms; however, patients with myeloradiculopathy symptoms for < 1 year had a significantly higher rate of good outcome (78.9% vs. 50.0%, p<.01) (Heidecke, 2000). Finally, a significant (p=.01) correlation was observed between outcome scores derived from the Neurological Cervical Spine Scale and duration of symptoms. Those with symptom duration of 6 months or less had the highest mean score (3.3); scores declined with increasing symptom duration, culminating with score of 1.2 among patients with symptom duration > 4 years (Kadoya, 2003).
Age
Seven case series provided data on patient subgroups based on age. While 3 studies described no correlation between age and clinical outcomes such as the Neck Disability Index and the Hirabayashi recovery rate (Chen, 2009; Matsumoto, 2009; Bindal, 2007), 3 other case series demonstrated statistically-significant differences based on patient age in rates of adverse events and neurologic outcomes (Cabraja, 2011; Kadoya, 2003; Heidecke, 2000). For example, the rate of neurologic improvement significantly declined with increasing age, from 71.0% among those age < 40 years to 11.1% among patients age 70 years or older (p=.014) (Kadoya, 2003). An additional case series found a greater incidence of dysphagia in younger (mean age 48 years) vs. older (mean age 55 years) patients (Yue, 2005b); differences were not tested statistically, however.

Workers’ Compensation
Goldberg et al. evaluated outcomes in 80 patients with and without worker’s compensation (Goldberg, 2002). Patients underwent anterior discectomy and fusion with a mean follow-up of 4 years. There were no statistically-significant differences in outcome based on Odom’s criteria; incidence of donor site pain and the development of pseudarthrosis also did not differ. Seventy percent of patients with workers’ compensation returned to work without restriction vs. 80% of patients without workers’ compensation, although this difference was not tested.

In a case series involving 66 patients undergoing anterior discectomy and fusion, a subgroup analysis of work- and non-work-related injuries demonstrated no significant differences in outcome based on Odom’s criteria at 22 months of follow-up (Samartzis, 2005).

Analysis of Comparative Value (KQ 4)
The comparative value of cervical fusion was assessed through a review of existing literature on cost-effectiveness and the development of a de novo decision analytic model and subsequent performance of cost-utility analyses. Decision analysis is a systematic quantitative approach for assessing the relative value of one or more different decision options. To structure the analysis a simulation model is constructed. In the kind of model used for our analysis, called a Markov model, patients with the condition in question are assumed to be in one of a limited number of distinct clinical “states”, e.g. “without symptoms”. Each state is assigned a “utility”, which represents an estimate of the relative quality of life of patients in that state. The probability of patients transitioning from one state to another is estimated from the clinical literature and/or expert input. The model is then used to compare clinical outcomes and costs over time across different “pathways” of care using different clinical interventions. One pathway is termed the “reference” pathway as it serves as the general comparator for all other interventions. Comparisons can be made of cumulative clinical events, such as complications; of cumulative mortality and quality of life; and of costs.
The basic outline of the decision model designed for this evaluation is shown in Figure ES4 below. The model was structured so that, as time passes, the clinical status of patients with cervical DDD is represented by one of the three mutually exclusive states shown. The model shifted patients between the different model states at three-month intervals over a 1-3 year time horizon.

**Figure ES4. Markov Disease State Diagram for Cervical Degenerative Disc Disease**

SAE=serious adverse event.

Anterior cervical discectomy and fusion (ACDF) was chosen as the primary approach of spinal fusion for the analysis. The reference case analysis compared spinal fusion with conservative management and assumed that fusion would initially have more pronounced effects versus conservative management but the effects for conservative therapy “catches up” over time (fusion 40% better at 6 months; 20% better at 12 months, 10% better at 24 months; equal at 48 months), based on data from a randomized study judged to be the most reliable source for this purpose (Persson, 2001). Baseline transition probabilities for spinal fusion were derived using data from an RCT which compared fusion with cervical arthroplasty (Sasso, 2011). Physical therapy was the chosen conservative management modality in the reference case. Comparisons of spinal fusion with other potential treatment alternatives of interest are also reported, as follow below:

- Manual therapy with spinal manipulation
- Epidural steroid injections
- Posterior laminoforaminotomy
- Anterior discectomy without fusion

**Outcome Measures**
This evaluation assessed key clinical outcomes related to the diagnosis of cervical DDD including the proportion of patients who had resolved cervical pain symptoms, the proportion of patients who had cervical pain symptoms, and the occurrence of rare but important adverse events such as perioperative complications and longer-term complications such as adjacent segment disease, and death. Costs related to treatments, total costs to the payer, and the impact of different treatment pathways on quality of life (as reflected by QALYs) were also calculated.

**Key Assumptions**

Listed in Table ES6 below are the key assumptions made in designing the model for this evaluation. Our model was based to some degree on past decision models associated with the management of cervical DDD (Carreon, 2012; Van der Velde, 2008) as well as on clinical studies comparing spinal fusion with conservative treatment or cervical arthroplasty (Persson, 2001; Sasso, 2011).

<table>
<thead>
<tr>
<th>Table ES6. Key Model Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior to entering the model patients have had an initial trial of conservative management lasting between 6-12 weeks which did not resolve symptoms.</td>
</tr>
<tr>
<td>The gap in clinical benefit between spinal fusion and conservative treatment narrows over time as patients with conservative treatment reach similar pain and function levels, consistent with observations from clinical studies.</td>
</tr>
<tr>
<td>All forms of conservative management (e.g., physical therapy, spinal manipulation) have approximately equal clinical effectiveness.</td>
</tr>
<tr>
<td>Patients who do not resolve symptoms of cervical pain will have a reduction in health related quality of life and will incur costs equivalent to approximately $1,983 per cycle (~ equal to cost of ongoing physical therapy).</td>
</tr>
</tbody>
</table>

**Summary Model Results**

**Spinal Fusion vs. Conservative Treatment**

Table ES7 on the following page provides the results for adult patients with severe cervical pain who have failed an initial conservative therapy course of 6-12 weeks’ duration. A greater percentage of patients using spinal fusion resolved their cervical pain symptoms than patients treated with conservative therapy, although conservative therapy “catches up” over time; the absolute difference in this percentage was 14.8% after 1 year, but only 3.6% after 3 years.
Table ES7. Clinical Results from Reference Case Analyses

<table>
<thead>
<tr>
<th>Time Horizon</th>
<th>Symptoms of Cervical Pain</th>
<th>Resolution of Symptoms of Cervical Pain</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spinal Fusion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>25.8%</td>
<td>74.0%</td>
<td>0.2%</td>
</tr>
<tr>
<td>2 year</td>
<td>26.3%</td>
<td>73.2%</td>
<td>0.5%</td>
</tr>
<tr>
<td>3 year</td>
<td>26.5%</td>
<td>72.9%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Conservative Treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>40.6%</td>
<td>59.2%</td>
<td>0.2%</td>
</tr>
<tr>
<td>2 year</td>
<td>33.7%</td>
<td>65.9%</td>
<td>0.5%</td>
</tr>
<tr>
<td>3 year</td>
<td>30.1%</td>
<td>69.2%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Absolute Difference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>-14.8%</td>
<td>14.8%</td>
<td>0.0%</td>
</tr>
<tr>
<td>2 year</td>
<td>-7.3%</td>
<td>7.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>3 year</td>
<td>-3.6%</td>
<td>3.6%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Following the pattern of clinical benefit, the incremental cost to achieve 1 additional treatment response (i.e., a patient with resolution of symptoms) increased from $174,515 in year 1 to $677,917 in year 3, as shown below in Table ES8.

Table ES8. Results of Reference Case Cost-Effectiveness Analysis

<table>
<thead>
<tr>
<th></th>
<th>Spinal Fusion</th>
<th>Conservative Therapy</th>
<th>Difference</th>
<th>Cost per Additional Responder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cost</td>
<td>% With Resolution of Symptoms</td>
<td>Cost</td>
<td>% With Resolution of Symptoms</td>
</tr>
<tr>
<td>1 year</td>
<td>$31,981</td>
<td>74.0%</td>
<td>$6,153</td>
<td>59.2%</td>
</tr>
<tr>
<td>2 year</td>
<td>$33,957</td>
<td>73.2%</td>
<td>$8,895</td>
<td>65.9%</td>
</tr>
<tr>
<td>3 year</td>
<td>$35,897</td>
<td>72.9%</td>
<td>$11,204</td>
<td>69.2%</td>
</tr>
</tbody>
</table>

From a cost-utility perspective, spinal fusion produced more QALYs than conservative therapy, albeit an increased cost (range $24,693 - $25,828 across horizons) (Table ES9 on the following page). The incremental cost per QALY gained for spinal fusion versus conservative therapy ranged from $347,473 to $579,428 depending on the time horizon considered.
Table ES9. Results of Reference Case Cost-Utility Analysis

<table>
<thead>
<tr>
<th></th>
<th>Spinal Fusion</th>
<th>Conservative Therapy</th>
<th>Difference</th>
<th>Cost per QALY, Spinal Fusion versus Conservative Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>$31,981</td>
<td>$6,153</td>
<td>$25,828</td>
<td>$579,428</td>
</tr>
<tr>
<td>QALYs</td>
<td>0.6609</td>
<td>0.6163</td>
<td>0.0446</td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost</td>
<td>$33,957</td>
<td>$8,895</td>
<td>$25,062</td>
<td>$401,306</td>
</tr>
<tr>
<td>QALYs</td>
<td>1.3060</td>
<td>1.2435</td>
<td>0.0625</td>
<td></td>
</tr>
<tr>
<td>2 year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost</td>
<td>$35,897</td>
<td>$11,204</td>
<td>$24,693</td>
<td>$347,473</td>
</tr>
<tr>
<td>QALYs</td>
<td>1.9303</td>
<td>1.8592</td>
<td>0.0711</td>
<td></td>
</tr>
<tr>
<td>3 year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Spinal Fusion vs. Other Treatments
Table ES10 below and on the following page provides a summary of comparisons of spinal fusion with the other treatments. Spinal fusion was slightly more effective and slightly less expensive than laminoforaminotomy (based on Washington HCA reimbursement amounts), but was more expensive than discectomy. Fusion was most cost-effective when compared to initial therapy with epidural steroid injections, as we assumed that such injections would only be one-third as effective as fusion at a constant rate over time.

Table ES10. Comparisons of Spinal Fusion to Other Treatments

<table>
<thead>
<tr>
<th>Comparator</th>
<th>Incremental Cost for Fusion</th>
<th>Incremental QALYs for Fusion</th>
<th>Incremental Response (% Improvement) for Fusion versus Comparator</th>
<th>Incremental Cost per QALY Gained (ICER)</th>
<th>Incremental Cost per Responder for Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual therapy with spinal manipulation</td>
<td>$28,465</td>
<td>0.0711</td>
<td>3.6%</td>
<td>$400,544</td>
<td>$781,460</td>
</tr>
<tr>
<td>Laminoforaminotomy (relative risk (RR) of cervical pain resolution, 0.98 vs. spinal fusion, mean cost, $29,556)</td>
<td>-$328</td>
<td>0.0115</td>
<td>2.2%</td>
<td>Less expensive and more effective</td>
<td>Less expensive and more effective</td>
</tr>
<tr>
<td>Discectomy (RR of cervical pain resolution, 0.98 vs. spinal fusion, mean cost, $22,284)</td>
<td>$6,945</td>
<td>0.0115</td>
<td>2.2%</td>
<td>$603,558</td>
<td>$317,757</td>
</tr>
</tbody>
</table>
**Table ES10. Comparisons of Spinal Fusion to Other Treatments**

<table>
<thead>
<tr>
<th>Comparator</th>
<th>Incremental Cost for Fusion</th>
<th>Incremental QALYs for Fusion</th>
<th>Incremental Response (% Improvement) for Fusion versus Comparator</th>
<th>Incremental Cost per QALY Gained (ICER)</th>
<th>Incremental Cost per Responder for Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epidural steroid Injection (RR of cervical pain resolution, 0.39 vs. spinal fusion, mean cost, $443 and 2 per cycle)</td>
<td>$18,831</td>
<td>0.2340</td>
<td>44.4%</td>
<td>$80,488</td>
<td>$42,375</td>
</tr>
</tbody>
</table>

**Sensitivity and Alternative Analyses**

Due to the uncertainty around clinical evidence, a substantive number of sensitivity analyses were conducted to explore the impact of varying parameter values and assumptions within the model. Across a range of assumptions and variation of key model parameters, including discount rate, inclusion of work-loss costs, assumptions regarding treatment effects, subsequent treatment, and specific consideration of side effects, the cost-utility values for spinal fusion exceeded thresholds of $100,000 per QALY. As one example, when the employer perspective is taken and estimated productivity losses/gains are included for the 3-year time horizon, the incremental cost-effectiveness of spinal fusion versus conservative therapy was $322,429 indicating that considering productivity has relatively small impact on estimates of the incremental cost-effectiveness of surgery vs. conservative management.

**Limitations**

There were considerable gaps in available clinical evidence. As a result, findings from the economic evaluation should be interpreted with caution. There was a limited body of clinical evidence comparing spinal fusion with alternative treatments other than surgery. Further, there was considerable variation in patient populations, study design, and outcome definitions across studies, which limits the comparability of evidence. As a result, effect estimates are based on assumptions around the clinical effects of conservative treatment versus spinal fusion, using data derived from the Persson, 2001 study. Because of the underlying uncertainty around clinical effects of fusion versus conservative treatment, we conducted numerous sensitivity analyses where efficacy for surgery versus conservative treatments was varied. Nonetheless, to obtain more robust comparative clinical and cost-effectiveness estimates, more clinical studies are needed comparing fusion with alternative conservative therapies after patients have had a trial of conservative therapy.
Comparison to Other Economic Studies of ACDF
Carreon et al. conducted a cost-effectiveness analysis of single level ACDF in the United States (Carreon, 2012). Carreon reported cost per QALY estimates of $106,256 at year 1, $54,622 at year 2, $38,836 at year 3, $29,454 at year 4 and $24,479 at year 5, and concluded that single-level instrumented ACDF is both effective and durable resulting in QALY gained as compared to other widely accepted healthcare interventions. The results from Carreon differ from those reported in our study largely because the Carreon study did not compare fusion’s costs and effects to any alternative treatment strategy. Standard guidelines for economic evaluation recommend that a comparator be used when conducting economic evaluations when they are available. Further, Carreon applied a cost of $15,714 which is almost half the cost of fusion in the Washington State Health Care Authority. If the cost of surgery (~$30,000) in the Washington State Health Care Authority were applied in the Carreon analysis, the cost per QALY estimates in the Carreon study would approximately double.

Although the Carreon study applied similar utility gains for resolving cervical pain to those applied in this report (0.18), it is important to note that assumed utility gains are equivalent to gaining 67.5 days of perfect health each year. These estimates exceed those reported for severe complications such as stroke or myocardial infarction (Sullivan, 2006). Consequently, the utility gains in the Carreon analysis (and our reference case analysis) may not be generalizable to patients with less severe forms of cervical DDD, where utility gains may be less pronounced. These aforementioned issues (no comparator, assumed cost of fusion of $15,714, and assumed utility gain of 0.18) limit the generalizability of findings from the Carreon analysis to patient populations with more severe forms of cervical DDD in which conservative treatments are not an option. Indeed, we were able to produce findings similar to those reported in the Carreon analysis (~$40,000 per QALY gained at year 3) when we: 1) did not use a comparator (i.e., assumed conservative treatment costs $0 and fusion is 100% more effective), 2) assumed fusion costs $15,714, and 3) applied utility gains of 0.185 for resolution of cervical pain symptoms.

Our analysis is more generalizable to the Washington State Health Care Authority – we consider alternative treatment strategies; use Washington State Health Care Authority specific costs; and our analysis considers multiple sources of utilities, some of which may be more generalizable to broader patient populations such as those less severe forms of disease where an alternative conservative treatment remains an option despite a previous trial of conservative treatment. Nonetheless, further research is needed exploring how gains in health related quality of life vary by disease severity.

The only other study identified compared the cost-effectiveness of various types of fusion (i.e., plating and allograft vs. autograft) (Angevine, 2005). Because this type of comparison was outside the scope of our systematic review, it is not discussed in detail here.
Summary ICER Integrated Evidence Ratings

The ICER integrated evidence rating matrix is shown below; a detailed explanation of the methodology underpinning this rating system can be found beginning on page 30. Separate ratings are provided for cervical fusion in comparison to conservative management (i.e., physical therapy or cervical collar immobilization) and discectomy or foraminotomy respectively.

ICER Integrated Evidence Rating™:
Spinal Fusion vs. Conservative Management (Physical Therapy/Cervical Collar)
for Patients with Cervical Degenerative Disc Disease

<table>
<thead>
<tr>
<th>Rationale for ICER Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>As noted previously, there were no randomized controlled trials focusing only on patients with generalized neck pain only. In fact, most comparative cohort studies also required some documentation of radiculopathic or myelopathic symptoms. Therefore, the evidence for fusion’s comparative clinical effectiveness in relation to conservative management in patients without radiculopathy is “Insufficient”.</td>
</tr>
</tbody>
</table>
Even for patients with clinical symptoms of radiculopathy and radiographic evidence of nerve root compression there is not a large evidence base comparing outcomes between spinal fusion and conservative management. We identified only 1 RCT and 1 comparative cohort study, neither of which stood out for their methodologic rigor, size, or generalizability.

Nevertheless, despite variability in study design, entry criteria, and outcomes measured, findings were reasonably consistent. Specifically, spinal fusion appeared to provide faster relief of pain and symptoms than conservative management (i.e., physical therapy or cervical collar immobilization) in the short term. Over time, however, these differences diminished and no material differences in outcome were observed by 12 months after intervention. Because of this, and because spinal fusion may cause relatively rare but significant complications, we deemed the overall comparative clinical effectiveness of fusion to conservative management “Comparable”. In some patients, however, neck pain and related symptoms may be so severe and disabling that the faster relief potentially afforded by fusion surgery would also allow a quicker return to work and other normal activities. For such patients, fusion might in fact be considered “Incremental” in comparison to ongoing conservative management.

However, fusion is a high-cost intervention; as illustrated by our decision analysis, even the greater short-term clinical and return-to-work benefits assumed for fusion cannot offset its much higher costs relative to conservative management, particularly because these benefits wane over longer time horizons; as such, fusion is associated with high cost-effectiveness ratios and costs per treatment responder at 1 year that only increase over time. As such, the comparative value of fusion vs. conservative management is deemed to be “Low”.
**ICER Integrated Evidence Rating™:**

Spinal Fusion vs. Discetomy or Foraminotomy for Patients with Cervical Degenerative Disc Disease

With Radiculopathic Symptoms

- Vs. Discetomy: [Yellow Square]
- Vs. Foraminotomy: [Green Square]

Without Radiculopathic Symptoms: [Red Square]

### Comparative Clinical Effectiveness

<table>
<thead>
<tr>
<th>Superior: A</th>
<th>Incremental: B</th>
<th>Comparable: C</th>
<th>Inferior: D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aa</td>
<td>Ba</td>
<td>Ca</td>
<td>Da</td>
</tr>
<tr>
<td>Ab</td>
<td>Bb</td>
<td>Cb</td>
<td>Db</td>
</tr>
<tr>
<td>Ac</td>
<td>Bc</td>
<td>Cc</td>
<td>Dc</td>
</tr>
</tbody>
</table>

### Comparative Value

- Superior: **A**
- Incremental: **B**
- Comparable: **C**
- Inferior: **D**
- Promising/Inconclusive: **P**
- Insufficient: **I**

#### Rationale for ICER Rating

As in comparisons to conservative management, there is insufficient evidence to evaluate the clinical benefits of spinal fusion relative to other major surgical procedures in patients with only generalized neck pain and no confirmatory evidence of radiculopathic symptoms.

In patients with clinical symptoms of radiculopathy and radiographic evidence of nerve root compression, a moderate level of RCT- and cohort-based evidence exists that suggests that clinical outcomes do not materially differ between these types of surgery, either over the short- or long-term. Differences in rates of complications and subsequent surgery have been observed in some studies, but findings have not been consistent. We therefore chose...
to rate the comparative clinical effectiveness of spinal fusion to discectomy or foraminotomy as “Comparable”.

However, data from the Washington HCA suggests that, while average payments for cervical fusion and foraminotomy are similar (approximately $30,000), payments for discectomy procedures are lower ($22,000). Given that equivalent clinical outcomes would be expected with all of these procedures, we chose to rate the comparative value of spinal fusion relative to foraminotomy as “Reasonable/Comparable”, and the comparative value of spinal fusion relative to discectomy as “Low”.
Methodology: ICER Integrated Evidence Rating™

The ICER Integrated Evidence Rating™ is constructed as a matrix, with a vertical axis denoting the possible categories for a rating of comparative clinical effectiveness, and the horizontal axis divided into 3 possible rating categories for comparative value (Ollendorf, 2010). It is important to note that these ratings are specified as comparing specific uses of medical interventions; that is, there may be different ratings for different uses of a test, treatment, or other intervention depending on the specified indication and patient population(s).

Level of Certainty in a Comparative Net Health Benefit

The underlying approach to ICER’s rating of comparative clinical effectiveness evolved from an earlier model developed by a multi-stakeholder workgroup convened in 2007 by America’s Health Insurance Plans (AHIP). The rating matrix, depicted in the graphic below, is used to rate the comparative clinical effectiveness of one or more interventions relative to a comparator of interest (Ollendorf, 2010, and is designed with the flexibility to compare multiple types of interventions, including drugs, devices, procedures, programs, and healthcare system processes. The rating matrix relies on a joint judgment of:

a) The **magnitude** of the difference between an intervention and its comparator in “net health benefit” – the balance between clinical benefits and risks and/or adverse effects (horizontal axis); AND

b) The level of **certainty** that one has in the best point estimate of net health benefit (vertical axis).

<table>
<thead>
<tr>
<th>High Certainty</th>
<th>D</th>
<th>C</th>
<th>B</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate Certainty</td>
<td>I</td>
<td>P/I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Certainty</td>
<td>I</td>
<td>I</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative Net Benefit</th>
<th>Comparable Net Benefit</th>
<th>Small Net Benefit</th>
<th>Substantial Net Benefit</th>
</tr>
</thead>
</table>
A = “Superior” - High certainty of a moderate-large net health benefit
B = “Incremental” - High certainty of a small net health benefit
C = “Comparable” - High certainty of a comparable net health benefit
D = “Inferior” - High certainty of an inferior net health benefit
P/I = “Promising but Inconclusive” - Moderate certainty of a small or moderate-large net health benefit
I = “Insufficient” - The evidence does not provide high certainty that the net health benefit of the technology is at least comparable to that provided by the comparator(s).

The term comparative “net” health benefit is used because of the importance attached to an explicit judgment of the overall balance of benefits and risks between an intervention and its selected comparator(s). The rating of net health benefit on the horizontal axis of the Comparative Clinical Effectiveness Matrix represents the best conceptual “point estimate” ICER can make given its interpretation of the existing evidence. There is no set definition of the boundaries between “comparable,” “small,” and “substantial” comparative net health benefit. For example, if the results of the appraisal include an estimate of a small lifetime quality-adjusted life year (QALY) advantage for one intervention compared with another, balanced against known greater short-term risks, whether or not these findings should be judged as conferring a comparative net health benefit will depend on many features of the relative certainty of the benefits and harms, as well as value judgments of the importance to patients of small QALY gains over a lifetime.

Despite the variability that will attend these judgments, presenting a categorical judgment of net health benefit serves an important goal: it enhances understanding of the underlying evidence by forcing the review team to justify its rating. The review team must describe more concretely than they might otherwise their view of how the disparate findings of a systematic review and decision model sum up. The review team’s justification can be debated and disagreed with, but in all cases it will give decision makers a more clear insight into the key issues they should consider when summing up the evidence and applying it to particular clinical actions or policies.

The vertical axis of the comparative clinical effectiveness matrix rates the level of certainty that the evidence provides in the precision of the net health benefit. There are 3 categories: high, moderate, and low. While the vertical axis represents a judgment of certainty, the horizontal axis of the Comparative Clinical Effectiveness Matrix displays gradients of the estimated net health benefit provided by a health intervention compared with the net health benefit of the selected comparator intervention. The categories for comparative net health benefit begin at the far left with “negative”; as the estimate of net health benefit increases, the rating moves to “comparable,” then to “small net benefit,” and culminates with a rating of “substantial” comparative net health benefit.

When assigning a level of certainty, it may be useful to consider that “conceptual confidence intervals” around a point estimate of comparative net health benefit that do not extend beyond a single box on the matrix represent a “high” level of certainty. Thus, if the point estimate of comparative net health benefit is “comparable,” and you feel that the reasonable bounds of your conceptual confidence interval do not extend into either “negative” or “small,” then you have high certainty in a “comparable” net health benefit.
At the other end of the spectrum, no matter where your point estimate of comparative net health benefit is, if you feel that the conceptual confidence interval extends across multiple boxes such that there is a reasonable chance that the comparative net health benefit could be “negative,” then there is “low” certainty. Finally, “moderate” certainty is used to imply that the conceptual confidence interval could extend one, or perhaps even two boxes in either direction. Essentially, for both low and moderate certainty, the possibility exists that the introduction of new evidence might move the estimate of comparative net health benefit in either direction.

Summary Rating of Comparative Clinical Effectiveness

As shown in the figure above, the Comparative Clinical Effectiveness Matrix maps the 3 categories of certainty upon the categories of comparative net health benefit to define a summary rating of comparative clinical effectiveness. Here, the relationship between level of certainty and magnitude of net health benefit comes into sharper relief. With a high level of certainty, the point estimate of net health benefit in one category is relatively assured, and therefore each cell in the matrix on the row of high certainty has a distinct label. A technology whose evidence base provides high certainty of a moderate-to-high net health benefit is rated to have “superior” comparative clinical effectiveness. As the net health benefit diminishes, the rating of comparative clinical effectiveness shifts to “incremental,” then “comparable,” and finally “inferior.”

When the level of certainty in the point estimate is only moderate, however, the summary ratings change to reflect that the conceptual confidence interval stretches 1-2 boxes in either direction. The “P/I” (promising but inconclusive) category describes an intervention with evidence suggesting that it provides either a small or substantial net benefit over the comparator (i.e., the “point estimate”). This point estimate, however, is relatively uncertain – further evidence may change the estimate of comparative net health benefit, and the conceptual confidence interval may extend one box into the “comparable” net health benefit box, and even includes a small but not unreasonable chance that the true comparative net health benefit is “inferior.” The P/I rating is a particularly important one for emerging interventions with some evidence of benefit over alternative treatments, but limitations in the evidence, particularly on longer-term safety and effectiveness, usually reduce certainty in the true magnitude of comparative net health benefit.

The final rating category is “I” (insufficient). This is used in two situations: (a) when there is moderate certainty that the best point estimate of an intervention’s comparative net health benefit is comparable or inferior; and (b) any situation in which the level of certainty in the evidence is low, indicating that no matter what the point estimate is for comparative net health benefit, limitations in the body of evidence are so severe that the conceptual confidence interval extends across multiple categories in the ICER matrix, representing a reasonable likelihood that the true net benefit is inferior.
Rating Comparative Value
The rating of comparative clinical effectiveness can stand alone, to be discussed and applied by decision makers, but it also forms the first of the 2 parts of the ICER Integrated Evidence Rating. The second component is a rating of “comparative value.” ICER rates the use of interventions for particular patient populations as having “high,” “reasonable or comparable,” or “low” comparative value.

ICER does not employ a single measure of cost effectiveness, such as the incremental cost-effectiveness ratio, for assignment of a rating of comparative value, and therefore does not rely on a formal cost-effectiveness threshold. Instead, the rating of comparative value is informed by multiple measures of potential economic impact.

To determine a final rating of “high,” “reasonable/ comparable,” or “low” value, ICER considers all of the economic findings, including the relative uncertainty of model findings as explored through multiple deterministic sensitivity analyses and a probabilistic sensitivity analysis. To aid transparency, ICER provides general guidance that incremental cost per QALY ratios of less than approximately $50,000 will often be considered as indicative of a “high” value intervention; incremental cost per QALYs from about $50,000 to $150,000 would often fit within a designation as “reasonable” values; and incremental cost per QALYs above $150,000 would be more likely to suggest “low” value interventions. This general guidance is based upon previous academic work benchmarks modified by ICER’s interpretation of evidence on the role medical inflation and societal willingness to pay should have in creating cost-effectiveness thresholds (Braithwaite, 2008; King, 2005). While there is a limited normative or empiric basis for the loose boundaries ICER presents, these boundaries also reflect input from stakeholders in today’s health care system on how best to present incremental cost-effectiveness ratios within broad categories that can be widely understood, gain relative consensus, and be actionable.

Integrated Ratings
The ICER Integrated Evidence Rating™, as presented on the following page, combines the individual ratings given for comparative clinical effectiveness and comparative value. The overall purpose of the integrated ratings is to highlight the separate considerations that go into each element but to combine them for the purposes of conveying that clinical benefits provided by technologies come at varying relative values based on their cost and their impact on the outcomes of care and the health care system (Ollendorf, 2010).
The ICER Integrated Evidence Rating™

<table>
<thead>
<tr>
<th>Comparative Clinical Effectiveness</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior: A</td>
<td>Aa</td>
<td>Ab</td>
<td>Ac</td>
<td></td>
</tr>
<tr>
<td>Incremental: B</td>
<td>Ba</td>
<td>Bb</td>
<td>Bc</td>
<td></td>
</tr>
<tr>
<td>Comparable: C</td>
<td>Ca</td>
<td>Cb</td>
<td>Cc</td>
<td></td>
</tr>
<tr>
<td>Inferior: D</td>
<td>Da</td>
<td>Db</td>
<td>Dc</td>
<td></td>
</tr>
<tr>
<td>Promising/Inconclusive: P</td>
<td>Pa</td>
<td>Pb</td>
<td>Pc</td>
<td></td>
</tr>
<tr>
<td>Insufficient: I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td></td>
</tr>
</tbody>
</table>

Comparative Value

- a: High
- b: Reasonable/Comparative
- c: Low
APPRAISAL REPORT

SPINAL FUSION FOR CERVICAL DEGENERATIVE DISC DISEASE
Final Scope

Chronic neck pain due to the progression of cervical degenerative disc disease (DDD) is a prevalent condition that results in high rates of disability and work loss as well as substantial costs to the U.S. medical system. This appraisal focuses on cervical spinal fusion procedures and their major alternatives, including conservative treatments (e.g., physical therapy), spinal injections and other minimally-invasive procedures, and other surgical procedures. The final scope of this appraisal, described using the Populations, Interventions, Comparators, Outcomes, Timing, and Setting (PICOTS) format (Counsell, 1997), is described in detail in the sections that follow. Patients receiving cervical fusion are generally stratified according to the presence of specific symptoms:

- Radiculopathy (pain due to nerve root compression)
- Myelopathy (pain due to spinal cord compression)
- Axial neck pain (musculoskeletal or soft tissue pain due to muscle strain, whiplash, or non-specific reasons)

These symptoms can be present in patients with or without spondylosis, the presence of abnormal growths or bony spurs on the vertebrae caused by the progression of cervical DDD and the development of osteoarthritis.

Objective and Methods:
The objective of this report is to appraise the comparative clinical effectiveness and comparative value of spinal fusion for cervical DDD. To support this appraisal we report the results of a systematic review of published randomized controlled trials, systematic reviews, and observational studies as well as the findings from a de novo decision analysis.

Key Questions:

1. What is the comparative clinical effectiveness of cervical fusion for DDD relative to that of conservative management approaches, minimally-invasive procedures, and other forms of surgery?

2. What are the adverse events and other potential harms associated with cervical fusion compared to conservative management approaches, minimally-invasive procedures, and other forms of surgery?

3. What is the differential effectiveness and safety of cervical fusion according to factors such as age, sex, race or ethnicity, pre-existing conditions (e.g., smoking history), neuromuscular disease states (e.g., Parkinsonism), measurable spinal instability, technical approach to fusion, insurance status (e.g., worker’s compensation vs. other), and treatment setting (e.g., inpatient vs. ambulatory surgery)?

4. What are the costs and potential cost-effectiveness of cervical fusion relative to alternative approaches?
1. Background

1.1 The Condition

Chronic neck pain is a prevalent and costly disorder. Approximately 15-20% of adults report at least one episode of neck pain during a given year, and nearly half of these individuals seek care (Carroll, 2008). On an annual basis, it is estimated that 11-14% of workers will have some limitation in their activities due to neck pain (Côté, 2008). While no recent studies have been conducted in the US on the economic burden of neck pain specifically, the combined burden of neck and back disorders in this country has been estimated to total $86 billion (Deyo, 2008).

One of the common causes of chronic neck pain is the progression of degenerative disc disease (DDD) of the cervical spine. DDD is not in fact a disorder, but a natural consequence of aging that results in gradual deterioration of cervical intervertebral discs (Emery, 2001). As the ability of these discs to absorb the shock and stress of vertebral motion declines, they become inelastic and cause a settling of the spinal column structure and abnormal spinal motion patterns. This process may in turn cause the development of abnormal bony growths and/or spurs (spondylosis), osteoarthritis, and/or herniation of one or more cervical discs. All of these conditions may in turn cause radiculopathy, or peripheral nerve root impingement. Symptoms of cervical radiculopathy include neck and arm pain, and weakness, tingling, or numbness in the upper extremities (Mayo Clinic, 2012). Less commonly, cervical DDD progression and its sequelae may directly compress parts of the spinal cord (myelopathy), affecting gait and balance in addition to causing arm and/or leg weakness and numbness.

A variety of options are available to manage chronic neck pain associated with cervical DDD. Many patients benefit from conservative management, which may include pain medication, immobilization, exercise and/or physical therapy, or spinal manipulation (Korinth, 2008). Patients not responding to these initial measures may receive therapeutic injections of anesthetic, nerve blocks, or steroids, as well as other minimally-invasive procedures such as radiofrequency ablation. Patients with ongoing neurologic symptoms not responding to initial management or minimally-invasive options are considered possible candidates for surgical treatment (Korinth, 2008).

The most common surgical procedure employed in the U.S. for patients with symptomatic cervical DDD is spinal fusion (Cowan, 2006). The rate of spinal fusion has increased dramatically in recent years; an analysis of U.S. hospital discharge data from 1990-2004 showed an 8-fold increase in the utilization of anterior fusion procedures, even while the overall rate of hospital admissions for cervical DDD remained steady (Marawar, 2010). The cost of these procedures, as well as questions regarding the short- and long-term outcomes associated with fusion, have raised considerable interest in understanding the evidence on the relative effectiveness of this procedure in comparison to other management options.
2. The Alternative Management Strategies

A plethora of management options is available for patients with neck pain due to cervical DDD, from conservative approaches, including cervical immobilization and physical therapy, to spinal injections and other minimally-invasive procedures, to several surgical options. The management options of primary interest for this appraisal are described in detail in the sections that follow.

2.1 Conservative Treatment

Conservative treatment for chronic neck pain consists of a number of pharmacological and non-pharmacological therapies as well as self-care interventions. Frequently used as an initial treatment strategy for patients presenting with neck pain, the individual options can take many forms.

The most frequently-studied forms of conservative management for neck pain include cervical immobilization, physical therapy, and spinal manipulation. Immobilization can be accomplished using a range of devices, including soft and hard collars, body jackets, and braces. Cervical immobilization is intended to provide short-term pain relief sufficient for the patient to attempt other forms of therapy, and are typically used for no longer than 1-2 weeks (Muzin, 2008). Potential disadvantages of immobilization include muscle atrophy after long-term use, as well as restricted breathing and aspiration risk in patients using hard collars (Muzin, 2008).

Physical therapy typically includes both exercise-based approaches to strengthen supporting muscle groups as well as postural support, with the ultimate goal of increased range of motion and relief of pain. Chiropractic spinal manipulation, which is known as “manual therapy” when performed by physical therapists, involves both thrust and non-thrust techniques to adjust the cervical spine and correct any postural abnormalities (Childs, 2008). The intensity and duration of both types of therapy can vary based on patient symptoms, but most courses of therapy involve 1-3 visits per week followed by evaluation at 4 weeks; therapy can then be continued if pain and symptoms persist (University of Maryland Medical Center, 2012). Risks of physical therapy and spinal manipulation include emergence or re-emergence of pain or other radicular symptoms; manipulation techniques in particular also carry small but nonzero risks of serious complications, including vertebral artery dissection, dural tear, phrenic nerve injury, and stroke (Ernst, 2007). Other forms of conservative management may include, but are not limited to:

Self-Care
- Books, handouts
- Ice and/or heat
- Gentle stretching
- Positional support during work and/or sleep
Pharmacologic Therapy

- Acetaminophen
- NSAIDs
- Tricyclic antidepressants
- Benzodiazepines
- Tramadol, opioids

Non-pharmacologic Therapy

- Massage
- Acupuncture
- Yoga
- Cognitive-behavioral therapy

2.2 Minimally-Invasive Procedures

Spinal Injections

Spinal injections deliver medication to the anatomic location that has been identified as the likely source of pain (Falco, 1998). Several types of spinal injections are used in practice today. They can be classified as either intraspinal injections or injections outside the spine. Intraspinal injections are further categorized into one of three categories: steroid injections, nerve blocks, or chemonucleolysis, as shown below (Chou, 2009).

- Steroid injections
  - Epidural steroid injection
  - Facet joint steroid injection
  - Intradiscal steroid injection

- Nerve blocks
  - Medial branch blocks
  - Sympathetic nerve blocks
  - Selective nerve root blocks

- Chemonucleolysis

Epidural steroid injections deliver the steroid into the epidural space, the space between the dura and the spine. The injection typically includes both a long-lasting steroid and a local anesthetic. Additional types of steroid injections have other anatomic targets. Facet joint steroid injections deliver corticosteroids into the facet joints, joints that are located between and behind adjacent vertebrae. Intradiscal steroid injections involve injecting a corticosteroid into an intervertebral disc to treat what is believed to be “discogenic” pain.

Nerve block injections include an anesthetic and may also include a corticosteroid. These injections are intended to target specific areas thought to be the source of pain, temporarily blocking pain signals. Most commonly, these injections target the medial
branch nerves, which emanate from the facet joints and in turn carry pain signals from these joints. Nerve-blocking injections may also target the sympathetic nervous system, which control some of the body’s involuntary functions. Nerve blocks may target selective nerve roots. These injections are intended primarily to diagnose the source of pain, not to treat it.

Chemonucleolysis uses a proteolytic enzyme, usually chymopapain, to dissolve the inner part of a herniated disc, in an effort to resolve radicular pain.

Injections outside the spine may include Botulinum toxin (Botox) injections, local injections, and prolotherapy. Botox injections are injected into the muscles of the neck to control muscle spasms. Local injections utilize a local anesthetic, injected into the muscles or soft tissues of the neck to treat inflammation. Prolotherapy, which may also be referred to as sclerotherapy, is a procedure in which a chemical irritant is injected into the soft tissues of the neck. This promotes an inflammatory response, which is thought to lead to a natural healing that will strengthen the injured soft tissue and thus reduce neck pain.

Each type of injection procedure may last between 15 and 30 minutes. Spinal injections are often done under fluoroscopic (live X-ray) guidance, although controversy exists over whether such guidance improves outcomes (Murtagh, 2000). Once the needle is in the proper position, a contrast dye is injected to confirm the position of the needle. Following confirmation, the steroid/anesthetic solution is injected.

Risks associated with these procedures include misplacement of the needle (either advancing the needle too deeply or placing it in the wrong position). The outcomes of incorrect needle position include nerve damage, infection, bleeding, and headaches. Risks associated with the medications include elevated blood glucose, arthritis, stomach ulcers, and weight gain. Chemonucleolysis may also cause anaphylactic reactions in some patients. One risk specifically associated with epidural steroid injections is wet tap, in which the needle penetrates the spinal sac and enters the cerebrospinal fluid. This causes the fluid to leak, resulting in severe headaches. Other rare complications associated with epidural steroid injections include epidural hematoma and abscess. In addition, a recent outbreak of fungal meningitis has been tied to nonsterile preparation of methylprednisolone for spinal injection by a large compounding pharmacy in the Northeast; the most recent data available indicate 656 confirmed cases in 19 U.S. states, 39 of which have been fatal (Centers for Disease Control, 2012).

**Radiofrequency Denervation**

Radiofrequency denervation (also known as radiofrequency neurotomy) is a type of procedure that uses heat to cauterize the affected nerve(s) thought to be associated with neck pain (Niemisto, 2003). This procedure attempts to interrupt pain signals from these nerves, thereby reducing pain perception by the brain.

The physician uses fluoroscopy to help advance the placement of the needle into the desired location. A small amount of current is passed through the needle to ensure that it is next to the target nerve; this may briefly cause facet joint pain. The nerves are then
numbed to minimize joint pain while the lesion is being created. The process is repeated for up to 1-5 additional nerves. The entire procedure can last between 30 and 90 minutes and is performed in an outpatient setting. Patients are usually able to resume their normal activities in a short period. Risks associated with this procedure include pain or discomfort around the injection site, worsened joint pain, permanent nerve pain, infection, and bleeding.

Coblation Nucleoplasty
Coblation nucleoplasty (also known as percutaneous disc decompression) is a relatively new minimally-invasive procedure used to treat cervical and lumbar disc herniation. The procedure uses radiofrequency energy to create small channels within the herniated disc, which is then thermally treated, producing an area of thermal coagulation. Channels are then formed in the nucleus in order to decompress the herniated discs (Sim, 2011).

Similar to radiofrequency denervation, this procedure usually takes 20 to 30 minutes and is performed in an outpatient setting. Patients are typically able to resume normal activity within a short time after the procedure.

2.3 Surgical Procedures

In order to better visualize the effects of different surgical procedures for cervical DDD, the cervical spinal anatomy is displayed in Figure 1 on the following page.

Spinal Fusion
Spinal fusion may be performed in addition to discectomy or laminectomy (see below) in order to achieve adequate decompression of the nerve root or the spinal cord or in patients with significant spinal instability.

The spine is stabilized by fusing two or more vertebrae together, using bone grafts from the patient or bone bank; in some cases, bone-related products such as bone morphogenetic proteins (BMP) or synthetic products such as polymethylmethacrylate (PMMA) may be used as graft material instead (American Academy of Orthopaedic Surgeons, June 2010). Fusion also may or may not use instrumentation such as screws, plates, or cages. Instrumentation is generally used as an internal or external splint to hold the vertebrae together while the grafts heal. Fusion may be performed on one, two, or multiple disc levels. The surgical approach may be from the back (posterior) or front (anterior); anterior approaches have become the predominant form of fusion surgery in recent years due to concerns of nerve or spinal cord injury with posterior fusion.

During the operation, the surgeon removes the lamina to help relieve the pressure on the nerve. The surgeon then removes any additional bone that may impinge upon the affected nerve. Grafts are then added to the spine; these will eventually fuse with the spine to form a solid union. Patients usually stay in the hospital for at least 3-4 days post-procedure. Completion of the fusion process takes some time to achieve and the healing process varies from person to person. Because of this, assessment of fusion
“success” via radiographic images is not attempted until at least 6 weeks following the procedure. Patients are often asked to limit their activities before fusion is deemed to be complete so as not to prolong the process (Mayo Clinic, 2012). Return to work tends to be discouraged during this period, as even sedentary work may affect recovery (Medical Disability Guidelines, 2012). Following the completion of the fusion process, the surgeon may recommend a rehabilitation program to strengthen the muscles around the fused area (Mayo Clinic, 2012).

Figure 1. Illustration of cervical spinal anatomy.

Source: http://www.spineinfo.co.uk

Risks associated with spinal fusion include nerve root damage, cerebrospinal fluid leakage, bleeding, and infection. While the major risks are relatively rare, the odds of injury increase with increasing complexity of surgical approach and use of instrumentation (Deyo, 2010). Other complications, common to all types of major surgery, may include blood clots, myocardial infarction, pulmonary embolism, and pneumonia. Patients undergoing cervical fusion may have reduced range of neck motion due to relative inflexibility of the fusion material in comparison to a cervical disc. Finally, there are concerns that fusion can cause long-term adverse effects, most prominently adjacent segment disease, in which discs adjacent to the fusion location degenerate and cause a return of symptoms, necessitating further surgery (American Association of Neurological Surgeons, 2012).

**Discectomy**
Cervical discectomy is a surgical procedure to remove part of a bulging or herniated disc in an attempt to alleviate pressure on the surrounding nerve roots (American Academy of Orthopaedic Surgeons, June 2010). Open discectomy involves making a small incision in
the skin over the spine, removing some of the ligament and bone to access the disc, and removing some of the disc material; discectomy can be performed alone or in combination with cervical fusion.

Open discectomy is performed under general anesthesia and typically requires a one-day hospital stay. The surgeon makes an incision in the skin over the affected area of the spine. The muscle is removed from the bone. Once the surgeon can visualize the lamina, disc, and other surrounding structures, he or she will remove the section of the disc that is protruding from the disc wall. No material is used to replace the removed disc. After the procedure, patients should avoid strenuous activity and heavy lifting for some time. Sedentary work may be resumed within 1-2 weeks.

In addition to the open procedure, there are several alternative approaches to discectomy. Microdiscectomy is a form of discectomy where only the ruptured portion of the disc is removed. The surgeon makes a very small incision in the neck over the problem disc. A small portion of the vertebra is removed. An X-ray is used to help guide the surgeon to the right disc. Once the bony material has been removed, the surgeon locates the area near the pinched nerve root. With the aid of a microscope or endoscope, the ruptured portion of the disc is removed as well as any disc fragments that have broken off in the process.

Discectomy is generally a safe procedure but it is associated with some risks. These risks include infection, bleeding, injury to surrounding blood vessels or nerves, leaking cerebrospinal fluid, and injury to the dura mater, the outer layer of the spinal cord.

**Laminectomy and Laminoplasty**
Laminectomy and laminoplasty are surgical procedures used to alleviate pain that is believed to be caused by neural impingement (Mayo Clinic, 2012). Laminectomy involves the removal of the lamina bone, a thin bony layer that covers and protects both the spinal canal and spinal cord. Surgeons may also remove bone spurs from the facet joints during laminectomy; this also helps to remove pressure from the spinal nerves. In laminoplasty, the lamina bone and its associated membranes are preserved, but freed from positions that impinge on the nerves and repositioned (Medscape, 2012). Both procedures are typically performed under general anesthesia. Hospital stays may range from one to 3 days.

Activities such as lifting and bending should be avoided for a few weeks after these procedures. Procedural complications of laminectomy and laminoplasty are similar to those of other cervical procedures. In addition, manipulation of the lamina can cause nerve root deficits, and over the long term, the consequences of this surgery may result in atrophy of the posterior cervical muscles, which may in turn cause spinal deformity and instability (Korinth, 2008). Preservation of the laminae with laminoplasty was thought to prevent some of these complications, but long-term follow-up studies suggest comparable outcomes with these procedures (Nurboja, 2012).
**Foraminotomy**

Cervical foraminotomy is performed in patients with radiculopathy symptoms and focuses on widening the area where the spinal nerve roots exit the spinal column in order to relieve pressure on the affected nerve roots. Foraminotomy is considered less likely to be effective in patients with significant spondylosis or disc protrusion (Korinth, 2008). Foraminotomy may be performed on one side of the spinal column at single or multiple levels, on both sides at single and multiple levels, or in combination with laminectomy or laminoplasty (Epstein, 2002). In contrast to discectomy and laminectomy/laminoplasty, foraminotomy is an alternative to cervical fusion rather than a procedure that can be performed with or without fusion.

This surgery is usually done with through a posterior approach. The surgeon may use a small, rotary cutting tool (a burr) to shave the inside edge of the facet joint. This opens up the outer rim of the neural foramen. The burr is sometimes used to shave a small section of the bony ring on the back of the vertebra above and below the affected nerve root. Small cutting instruments are used to carefully remove soft tissues within the neural foramen. The surgeon takes out any small disc fragments that are present and scrapes off nearby bone spurs. In this way, tension and pressure are taken off the nerve root (Orthogate.org, 2012).

As with discectomy, foraminotomy may be performed using an open approach or employing microscopic and/or endoscopic techniques. Foraminotomy is generally a safe and straightforward procedure, with most patients able to leave the hospital after an overnight stay. Risks are similar to other forms of surgery for cervical DDD (Korinth, 2008).
3. Clinical Guidelines

Major guideline statements regarding cervical fusion can be found in the sections that follow below.

- **North American Spine Society (NASS, 2010)**
  

  Anterior cervical discectomy with fusion (ACDF) is recommended in the treatment of 1-level cervical radiculopathy from degenerative disorders and is considered a comparable treatment strategy to anterior cervical discectomy (ACD) based on long-term follow-up. ACDF or posterior laminoforaminotomy (PLF) are recommended for the treatment of 1-level cervical radiculopathy secondary to foraminal soft disc herniation, while ACDF is recommended over PLF in patients with 1-level disease from central and paracentral nerve root compression and spondylotic disease. Evidence suggests that ACDF results in comparable short-term success relative to ACD, PLF, and reconstruction with total disc replacement.

- **American Association of Neurological Surgeons/Congress of Neurological Surgeons Joint Section on Disorders of the Spine and Peripheral Nerves (AANS/CNS 2009)**
  

  For patients with cervical spondylotic myelopathy (CSM) or ossification of the posterior longitudinal ligament (OPLL), cervical laminectomy with fusion is recommended as an equivalent strategy to laminectomy or laminoplasty and is associated with postoperative neurological improvement. Laminectomy and fusion consistently results in ventral and dorsal cord decompression.


  ACD and ACDF produce equivalent clinical outcomes for patients with 1-level cervical disc degeneration. ACDF is recommended over ACD to reduce risk of kyphosis and increase fusion rate for patients with 1-level disease. ACDF is also considered superior to ACD in achieving quicker relief of neck or arm pain, though functional outcomes may be similar.

  Anterior cervical plating (ACDFI) does not improve long-term outcomes in patients with level-1 disease but is considered superior to ACDF in improving arm pain for patients with 2-level cervical disc degeneration. Plating does not improve other clinical outcomes with respect to 2-level disease. For patients with 1-level cervical degeneration, plating is recommended to reduce risk of pseudarthrosis, incidence of graft-related complications, and improve cervical lordosis, but not to improve clinical outcomes alone. Plating may increase surgical blood loss.
Anterior surgical nerve root decompression via ACDF is recommended with patients with cervical radiculopathy for fast relief (3–4 months) of arm or neck pain and/or sensory loss over physical therapy (PT) or immobilization with a cervical collar. Anterior surgical nerve root decompression may also improve long-term functional outcomes relative to PT, including wrist extension, elbow extension, shoulder abduction, and internal rotation. However, recurrent symptoms are common.

- **American College of Occupational and Environmental Medicine (ACOEM, 2011)**
  
  Cervical discectomy and fusion is recommended to speed recovery in patients with chronic cervical radiculopathy or symptomatic spinal stenosis who continue to have significant functional limitations after 6 weeks of appropriate non-operative therapy. All forms of decompressive surgery, with or without fusion, are recommended in patients with symptoms of cervical myelopathy. Cervical fusion is recommended in patients with degenerative spondylolisthesis or in patients undergoing discectomy for this condition if during the same operative episode as the discectomy.

  Cervical fusion is not recommended for chronic non-specific cervical pain.

- **Work Loss Data Institute (WLDI, 2011)**
  
  Anterior cervical fusion procedures are considered an option for a variety of chronic neck conditions. Posterior fusion remains under study and is not specifically recommended. Multi-level corpectomy with fusion is considered equivalent to other procedures in patients with cervical myelopathy, although the complication rate with fusion may be somewhat higher. Patients undergoing fusion at the C1-C2 level should refrain from returning to any activity with a risk of reinjury.

- **UpToDate (2012)**
  
  ACDF and other decompressive procedures should be considered in patients with (1) signs and symptoms of radiculopathy; (2) MRI or CT myelographic evidence of nerve root compression; and (3) persistence of radicular pain despite conservative management of at least 6-12 weeks’ duration. There is little convincing evidence that any one surgical option is superior to another, or that any improve upon the natural history of the condition.
Surgical consultation is warranted in patients presenting with cervical myelopathy and disabling neurologic deficits, or in patients with mild symptoms who are at risk of neurologic deterioration. There is no evidence to distinguish the relative benefits and risks of fusion techniques, laminoplasty, laminectomy, or corpectomy in patients with cervical myelopathy.
4. Medicare and Representative Private Insurer Coverage Policies

Centers for Medicare and Medicaid Services (CMS): Medicare does not have a National Coverage Determination (NCD) for any form of fusion surgery. Local coverage decisions (LCDs) are limited to the use of spinal fusion for lumbar degenerative disc disease only.

Similarly, few private payers have explicit coverage policies in place for cervical fusion surgery; those that were found are summarized below.

Humana: Patients are eligible for cervical fusion surgery when radiographic evidence confirms any of the following:

1. As a concurrent stabilization procedure with corpectomy or laminectomy;
2. Degenerative spinal segment adjacent to a previously decompressed or fused spinal segment with symptomatic myelopathy or radiculopathy corresponding to the adjacent level;
3. Disc herniation with radiculopathy with failure of conservative treatment and unremitting radicular pain secondary to nerve root compression;
4. Multilevel spondylotic myelopathy or radiculopathy;
5. Ossification of the posterior longitudinal ligament;
6. Symptomatic pseudoarthrosis from a prior procedure;
7. Symptomatic spondylosis with instability

Aetna: Cervical laminectomy (with or without anterior fusion) is considered medically necessary for individuals with a herniated disc or other causes of spinal cord or nerve root compression (osteophytic spurring, ligamentous hypertrophy) when all of the following criteria are met:

1. All other reasonable sources of pain have been ruled out; AND
2. Presence of neck or cervico-brachial pain with findings of weakness, myelopathy, or sensory deficit; AND
3. Imaging studies (e.g., CT or MRI) indicate nerve root or spinal cord compression at the level corresponding with the clinical findings; AND
4. Member has failed at least 6 weeks of conservative therapy (unless there is evidence of cervical cord compression, which requires urgent intervention); AND
5. Member has physical and neurological abnormalities confirming the historical findings of nerve root or spinal cord compression (e.g., reflex change, sensory loss, weakness) at or below the level of the lesion and may have gait or sphincter disturbance (evidence of cervical radiculopathy or myelopathy); AND
6. Member's activities of daily living are limited by persistent neck or cervico-brachial pain.
5. Previous Systematic Reviews/Technology Assessments

No recent technology assessments focusing on cervical spinal fusion were identified from national or international organizations other than an assessment from the National Institute for Health and Clinical Excellence that focused on a specific type of screw used during fusion. Two recent systematic reviews are available from the Cochrane Collaboration, however, and are summarized below.

- **Surgery for Cervical Radiculopathy or Myelopathy (Nikolaidis, 2010)**
  There is low quality evidence that surgery may provide pain relief faster than physiotherapy or hard collar immobilization in patients with cervical radiculopathy; but there is little or no difference in the long-term. There is very low quality evidence that patients with mild myelopathy feel subjectively better shortly after surgery, but there is little or no difference in the long-term.

- **Single or Double-Level Anterior Interbody Fusion Techniques for Cervical Degenerative Disc Disease (Jacobs, 2011)**
  Among discectomy alone, discectomy with interbody fusion, and discectomy with interbody fusion and anterior plating, there is no evidence that any one technique is better than another for clinically significant pain relief for patients with chronic cervical degenerative disc disease or disc herniation. There was no statistically-significant difference in complication rates in studies comparing discectomy alone to fusion procedures. There is moderate quality evidence that the use of a bone graft (bone transplanted from another part of the body) is more effective than discectomy alone in achieving fusion.
6. Ongoing Clinical Studies

Information on ongoing clinical studies that have been submitted to the U.S. National Institutes of Health’s registry of publicly- and privately-supported studies ([www.clinicaltrials.gov](http://www.clinicaltrials.gov)) is presented in the table below and on the following page.

<table>
<thead>
<tr>
<th>Title/Trial Sponsor</th>
<th>Design</th>
<th>Comparators</th>
<th>Patient Population</th>
<th>Primary Outcomes</th>
<th>Estimated Completion Date</th>
</tr>
</thead>
</table>
| Anterior vs. posterior procedures for cervical spondylotic myelopathy: prospective randomized clinical trial | RCT | Anterior decompression and fusion
Posterior cervical laminoplasty | • n = 100
• 40 to 80 years
• Patients with cervical spondylotic myelopathy requiring surgery of 1-3 levels | JOA score at 3, 6, 12 months, and 3, 5, 10 years | December 2020 |
| NCT00876603
The University of Hong Kong | | | | | |
| PureGen: a radiographic analysis of rate and quality of fusion in patients undergoing anterior cervical discectomy and fusion | Prospective single-arm cohort | N/A | • n=50
• 18 years and older
• Symptoms of cervical degenerative disc disease in 1-4 levels
• Persistent neck/arm pain
• Unresponsive to ≥6 weeks of conservative treatment | Fusion at 12 months | NR (study start date: February 2011) |
| NCT01291134
Alphatec Spine, Inc. | | | | | |

Source: [www.clinicaltrials.gov](http://www.clinicaltrials.gov)
<table>
<thead>
<tr>
<th>Title/ Trial</th>
<th>Design</th>
<th>Comparators</th>
<th>Patient Population</th>
<th>Primary Outcomes</th>
<th>Estimated Completion Date</th>
</tr>
</thead>
</table>
| Outcome analysis for minimally invasive spine surgery | Retrospective single-arm cohort | N/A | • n=200  
• Patients with degenerative disc disease  
• Procedures include posterior cervical fusion and anterior discectomy and fusion using silicate-substituted calcium phosphate ceramic graft material | VAS up to 2 years | April 2013 |
| Bone graft materials observational registry (APPROACH-001) | Prospective single-arm cohort | N/A | • n=300  
• 18 years and older  
• Failed conservative treatment, and potential candidate for spinal fusion surgery | Evidence of successful radiographic fusion at 6, 12 & 24 months | December 2013 |
| Evaluation of DTRAX graft in patients with cervical degenerative disc disease | Prospective single-arm cohort | N/A | • n=100  
• 35 to 80 years  
• Patients with degenerative disc disease, including arm/shoulder pain and disc herniation  
• Single- or multi-level disease  
• Unresponsive to ≥6 weeks of conservative treatment  
• NDI ≥30 | Clinical outcome data (NDI, VAS and quality of life questionnaire) at baseline and up to 12 months | February 2015 |

Source: [www.clinicaltrials.gov](http://www.clinicaltrials.gov)
7. The Evidence

Objectives
The primary objectives of the systematic review were to:

- Evaluate and compare the published evidence on the effects of cervical spinal fusion procedures relative to comparator treatments on pain, function, and health-related quality of life in patients with neck pain and other symptoms (e.g., radiculopathy) arising from cervical degenerative disc disease (DDD) with or without spondylosis;

- Evaluate and compare the clinical benefits of these therapies in terms of other outcomes, including rates of return to work and rates of “successful” outcomes;

- Evaluate and compare the potential harms of these therapies, including procedure-related fatalities and major and minor complications as well as requirements for repeat surgery;

- Examine the differential effectiveness and safety of cervical fusion according to factors such as age, sex, race, ethnicity, pre-existing conditions (e.g., smoking history), neuromuscular disease states (e.g., Parkinsonism), measurable spinal instability, technical approach to fusion, insurance status (e.g., worker’s compensation vs. other), and treatment setting (e.g., inpatient vs. ambulatory); and

- Assess the costs and potential cost-effectiveness of cervical spinal fusion in multiple patient populations relative to alternative approaches.

As discussed in greater detail in the sections that follow, the target population for this appraisal was patients with chronic neck pain and related symptoms due to cervical DDD, implying that these patients were treated conservatively at initial presentation and continued to have persistent symptoms. Our focus was therefore on management options available to these patients following 4-6 weeks of conservative therapy. As described in further detail in Section 7.6, we focused attention on evidence for cervical spinal fusion and the management options to which it has been compared in randomized controlled trials (RCTs) or comparative cohort studies, and did not therefore evaluate comparisons of other management options to non-fusion controls (e.g., laminectomy vs. laminoplasty). Case series of cervical fusion were also abstracted for long-term data on harms as well as information on key patient subgroups. We did not include for evaluation studies comparing cervical disc arthroplasty to fusion, as the Washington HCA has previously evaluated this evidence (Washington Health Care Authority, 2008).

Our recording of data on potential harms of fusion and other surgical procedures included “peri-procedure” fatalities and complications occurring during the procedure or within 30 days following, as well as longer-term adverse events observed throughout the remainder of follow-up. We abstracted each type of complication in detail in order to understand the universe of specific complications reported for fusion and its comparators.
While not part of the systematic review, published studies of the economic impact of cervical fusion are summarized in Section 8 to provide additional context for the ICER decision analytic model.

Analytic Framework
The analytic framework for this review is shown in the Figure below. Note that the figure is intended to convey the conceptual links involved in evaluating outcomes of cervical fusion and its alternatives, and is not intended to depict a clinical pathway through which all patients would flow. This framework also does not represent the clinical pathways as they were constructed for the decision analytic model (see Section 8).

There are limited data directly demonstrating the impact of most cervical DDD management strategies on summary measures of “treatment success” or “successful clinical outcome”, so judgments about the effectiveness of these interventions must rest primarily upon consideration of multiple and potentially overlapping measures (e.g., pain, function, quality of life) as well as evaluation of treatment-associated risks. In addition, various stakeholders will by nature be more interested in certain outcomes than others. For example, payers and employers may be most interested in functional improvement and/or return to work, while clinicians and patients may focus more on relief of symptoms and spinal stability.

There is considerable debate about how much credence to place in comparisons across studies of multiple outcome measures for the management of cervical DDD. As has been observed in studies of both chronic neck pain and low back pain, patient populations may differ significantly in terms of baseline severity of their condition and degree of impairment, which can then in turn affect the sensitivity of measurement instruments to
detect clinically important differences, even when these instruments are standardized and validated (Carey, 2007). In addition, the primary research questions and goals of management may differ substantially by approach. For example, interventions with a goal of functional restoration may show little to no effects on pain and quality of life measures.

7.1 Patient Populations

The focus of this appraisal was on adults (>17 years of age) with cervical DDD symptoms, including neck pain, arm pain, and/or radiculopathic symptoms (e.g., numbness, tingling); these symptoms could occur with or without the presence of spondylosis. We did not focus on studies that consisted primarily or in total of patients whose primary complaint was cervical myelopathy, as this is generally considered a neurologic emergency in all but the mildest cases and patients typically proceed directly to surgical intervention (McCormick, 2003). These studies are nevertheless included in evidence tables (see Appendix C), shaded to distinguish them from the studies of primary interest.

It was also recognized that some studies enrolled a mix of patients that might have included those with myelopathic symptoms. We did not therefore exclude any study purely on the basis of the presence of a subgroup with myelopathy, as long as this subgroup constituted the minority of subjects.

Studies of patients receiving intervention for acute trauma or systemic disease affecting the spine (e.g., malignancy, autoimmune conditions) were excluded. It was also noted that surgical studies would likely include patients with a variety of prior attempts at treatment, including conservative management for varying lengths of time. While these factors were noted in summaries of study and patient characteristics, all studies were included regardless of the nature, duration, and intensity of prior treatment received.

Certain patient subpopulations were also identified as of interest in evaluating whether treatment effects and/or harms differed in these groups. These included subpopulations defined by demographic characteristics (e.g., age, sex, race/ethnicity), pre-existing conditions (e.g., smoking history), neuromuscular comorbidities (e.g., Parkinsonism), measurable spinal instability, anatomic approach to fusion (i.e., anterior vs. posterior fusion, single- vs. multi-level procedures), insurance status (e.g., worker’s compensation vs. other), treatment setting (e.g., inpatient vs. ambulatory), and other subgroups as defined in available studies.

7.2 Interventions

The interventions of interest included the major approaches to cervical fusion. All forms of fusion were considered, regardless of type of decompressive surgery coupled with fusion (e.g., discectomy, laminectomy), use of instrumentation and type of hardware utilized, or graft material; these factors were nevertheless used as procedure descriptors. In addition, fusion studies were categorized by anatomic approach (i.e., anterior vs. posterior) and
number of disc levels involved (single, 2-level, or >2-level) where feasible, as benefits and risks are thought to vary according to these aspects (Shamji, 2008; Riley, 2010). Studies were also sought comparing outcomes for fusion performed in inpatient vs. ambulatory/outpatient settings.

7.3 Comparators

Potential comparators of interest in this review included all management options compared to fusion in RCTs and comparative cohort studies. These included conservative management approaches such as physical therapy, spinal manipulation, immobilization (i.e., via a cervical collar or brace), medication, and other approaches; minimally-invasive procedures such as spinal injections, radiofrequency denervation, and percutaneous procedures; and other forms of surgery, including decompressive procedures such as discectomy or laminectomy without fusion, laminoplasty, and foraminotomy. Importantly, we did not include studies comparing only one type of fusion to another (e.g. with vs. without plating), as recent systematic reviews have concluded that data are not sufficient to distinguish the performance of these approaches (Jacobs, 2011; Nishizawa, 2012; Gebremariam, 2012) and the focus of this review was on comparing fusion to alternative treatment modalities. Exceptions to this rule included studies where the comparison was of anatomic approach or number of levels fused, as well as the setting in which fusion was performed (i.e., inpatient vs. ambulatory/outpatient).

7.4 Outcomes

In order to adequately compare effectiveness across management options within each patient population, we prioritized for abstraction data from the most widely-used and validated outcome instruments; if other measures represented primary or key secondary outcomes, however, we also abstracted these. For example, some studies may have captured self-reported improvements in pain or other outcomes using a measure developed specifically for that study. Outcome measures of interest are described in more detail on the following pages, by type. Outcomes recorded at multiple timepoints were abstracted separately for each timepoint.

Operative Outcomes
A variety of data was abstracted as reported during the procedure and related hospital stay, including procedure duration, anesthesia duration, blood loss, number of disc levels, and length of stay in hospital. Any harms or unexpected events during the procedure or 30 days following also were recorded (see “Potential Harms” below).

Pain
Pain outcomes were evaluated based on visual analogue (VAS), numeric, or Likert rating scales, as well as the North American Spine Society (NASS) cervical questionnaire, which examines a variety of domains of pain and neurogenic symptoms (Stoll, 2004). Data were abstracted as recorded, including repeated-measures means and standard deviations or
standard errors at multiple timepoints, “change scores” (i.e., mean or median change from baseline), and both univariate and multivariate measures of treatment effect. The statistical significance of all findings was also abstracted as reported. While data were abstracted as reported for the systematic review, data transformations were performed for meta-analysis and simulation modeling if warranted. For example, VAS results using a different scale (e.g., 100 mm) may have been converted to 10 mm scales for the purpose of consistency. When examining changes in VAS pain, it is important to consider that influential clinical guideline and consensus statements consider 5-10 point changes on a 100-point VAS to be the minimal level of “clinically important” improvement (Chou, 2007). Improvements of ≥30% from baseline levels are considered clinically-meaningful to patients (Dworkin, 2008).

**Functional Status**

Findings with regard to patient functional status were assessed from studies employing well-known indices for measuring function in cervical disorders, including the Neck Disability Index, a 50-point scale derived from the Oswestry Low Back Pain Disability Index (Vernon, 1991); and the 17-point Japanese Orthopaedic Association (JOA) score (Keller, 1993). Further information on these measures can be found in Appendix A. Results from generic disability scales such as the Disability Rating Index (Salén, 1994) were also considered, as were subjective measures of functional ability and timed tasks such as handling everyday items and walking. Data were abstracted as described above for pain. As above, transformations of data were considered in modeling and/or meta-analysis where warranted.

**Health-related Quality of Life**

We sought data on all reported quality-of-life measures, whether from generic instruments such as the SF-36 or Sickness Impact Profile, condition-specific measures correlated with neck pain (e.g., the Beck Depression Inventory), or study-specific measures. Data were abstracted as described above for pain and function above. We abstracted data for all reported domains on these instruments as well as summary scores where available.

**Successful Clinical Outcome**

The frequency or likelihood of “successful clinical outcome” or “treatment success” was abstracted where reported. A variety of standardized versions of these measures are available, including Odom’s criteria (Odom, 1958), Hilibrand’s criteria (Hilibrand, 1999), Hirabayashi’s recovery rate (Hirabayashi, 1981), as well as study-specific measures. While all data were abstracted, patients with either an “excellent” or “good” outcome on Odom’s or Hilibrand’s criteria were considered to be treatment successes, while those with a positive Hirabayashi value were categorized as having a successful outcome. Detailed descriptions of standardized measures can be found in Appendix A.

**Return to Work**

Multiple measures of return to work were recorded, including the frequency of successful return to work and total amount of sick leave/absenteeism. Where available, time-to-event measures of return to part- or full-time employment were also recorded.
Potential Harms

Peri-Procedure Mortality and Complications

Peri-procedure deaths were classified as those occurring during the procedure or within 30 days following. Procedure-related complications were recorded as described in order to obtain an exhaustive list of reported events. No attempt was made to further categorize the complications in terms of severity, as this information was typically lacking in most study reports. Examples of complications recorded are listed below.

- Nerve damage/palsy
- Wound infection
- Hemorrhage
- Cerebrospinal fluid leak
- Thrombosis
- Pneumonia
- Hardware failure
- New-onset numbness or weakness
- Dysphagia (difficulty swallowing)
- Hoarseness
- Donor site pain (for autologous bone grafts)

In addition, recent data on risks associated with any surgery, while not specific to fusion procedures, were nevertheless sought to provide further information on the spectrum of possible surgical risks.

Retreatment

Rates of repeat treatment were abstracted where reported, and categorized separately as follows. First, rates of return to the operating room during the initial hospital stay (e.g., for hardware failure, control of hemorrhage, etc.) were recorded. In addition, rates of repeat surgery with the same procedure following the peri-procedure period also were recorded. Finally, rates of repeat surgery with a different procedure at any point during follow-up (e.g., for adjacent segment disease) also were abstracted if recorded as such.

Longer-term Adverse Events

In addition to peri-procedure complications, complications and adverse events as well as death from any cause also were recorded over the remaining duration of follow-up. In addition to the complication types described above, examples of adverse events recorded following the peri-procedure period are listed below:

- New-onset radiculopathy or myelopathy
- Neurological decline
- Symptomatic pseudarthrosis (non-union of fusion)
- Adjacent segment degeneration or disease (ASD)
7.5 Timeframe

The timeframe for evaluation of clinical benefits and potential harms differed by study design (see Section 7.7 below). Data from RCTs and comparative cohorts were considered from baseline through a period of follow-up that typically did not extend beyond 2 years. Case series data were focused on outcomes recorded at timepoints 12 months or later following intervention.

7.6 Study Designs

Data from both RCTs and observational studies were considered. However, only data from RCTs and, secondarily, comparative cohort studies involving cervical fusion and another form of active treatment were used to evaluate measures of clinical effectiveness, given the significant risks of placebo effects inherent in other study designs. Case series of cervical fusion were also utilized based on the following criteria:

- Sample size >50 cases; AND
- Follow-up ≥12 months; AND
- Data available on repeat surgery and/or other harms; OR
- Study population defined by a patient subgroup of interest as described in Key Question 3 (e.g., patients in a worker’s compensation program)

7.7 Literature Search and Retrieval

Because previous systematic reviews and guideline statements have cited a dearth of available RCTs comparing cervical fusion to alternative treatments, we did not put explicit time limits on our RCT search. The general timeframe for literature search and retrieval for other study designs was January 1996 – September 2012. We focused on English-language reports only. As noted previously, RCTs and comparative cohort studies were limited to those comparing fusion to an alternative management option or comparing posterior to anterior or single- to multi-level fusion; there was no minimum sample size or duration of follow-up in these studies. Case series were limited by both sample size and duration of follow-up, as described in Section 7.6 above.

The electronic databases we searched as part of the systematic review included MEDLINE, EMBASE, and The Cochrane Library (including the Database of Abstracts of Reviews of Effects [DARE]) for health technology assessments (HTAs), systematic reviews, and primary studies. Reference lists of all eligible studies were also searched. The strategies used for MEDLINE, EMBASE, and The Cochrane Library are shown in Appendix B.

Studies were not further restricted by instrumentation, manufacturer, or treatment approach. Figures 1 and 2 on the following pages show flow charts of the results of all searches for RCTs (n=14), comparative cohorts (n=7), and case series (n=56).
Figure 1. PRISMA flow chart showing results of literature search (RCTs).

- Titles and abstracts identified through MEDLINE, Cochrane, EMBASE and DARE: n=750
- Additional records identified through other sources: n=6
- Records after duplicates removed: n=370
- Records screened: n=370
- Records excluded through title/abstract review: n=275
- Full-text articles assessed for eligibility: n=95
- Full-text articles excluded (n=80):
  - Letter to Editor (n=3)
  - Protocol only (n=3)
  - Unable to obtain articles (n=3)
  - Reviews (n=2)
  - Not in English (n=1)
  - Not a patient population of interest (n=10)
  - No comparators of interest (n=52)
  - No outcomes of interest (n=6)
- Articles included in analyses: n=15, evaluating 14 unique studies
Figure 2. PRISMA flow chart showing results of literature search (observational studies).

Titles and abstracts identified through MEDLINE
n=2,113

Records screened
n=2,113

Records excluded through title/abstract review
n=1,842

Case series review:
title/abstract review, full-text review
n=215

Articles included in analyses of case series
n=56

Full-text articles excluded (n=49)
- Not a patient population of interest (n=21)
- No comparators of interest (n=14)
- No outcomes of interest (n=1)
- <25 patients in fusion study arm (n=13)

Comparative cohort review

Full-text articles assessed for eligibility
n=56

Articles included in analyses of controlled cohorts
n=7
Study Quality
We used standardized criteria specific to previous systematic reviews in back pain to rate the quality of each included RCT; these criteria have also been widely adopted for use in studies of neck pain (Nikolaidis, 2010). The criteria, which related to issues of study design, reporting, and minimization of bias, are presented in Appendix A. RCTs meeting a majority of criteria (i.e., 5 of 9) were deemed to be “higher quality”. We used general criteria to assess the quality of comparative cohort studies, using the categories “good”, “fair”, or “poor”. Our methods were based on the criteria employed by the U.S. Preventive Services Task Force (AHRQ, 2008), as described below:

- **Good**: Comparable groups (for comparative studies) are assembled initially and maintained throughout the study (follow-up of at least 80%); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; important outcomes are considered; and appropriate attention to confounders in analysis.

- **Fair**: Generally comparable groups are assembled initially but some question remains whether minor/moderate differences occurred in follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for.

- **Poor**: Any of the following problems exist: (1) groups assembled initially are not close to being comparable or maintained throughout the study; (2) unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and (3) key confounders are given little or no attention.

Data from all retrieved studies were included in evidence tables regardless of study quality. However, the focus of attention in presentation of results was on higher-quality studies alone where available.

Study quality was not assessed for case series, as the focus of quality ratings was on the level of bias in assessing the comparative impact of fusion vs. alternative treatments on measures of effectiveness and harm.

Data Synthesis
Where feasible, estimates of treatment effect were synthesized using meta-analysis. Random-effects models were generated based on head-to-head data from available RCTs. Data were deemed to be sufficient if (a) the number of eligible higher-quality RCTs was 2 or more; (b) the measure of interest was reported using uniform methods; and (c) judgment of the clinical heterogeneity of the patient populations in candidate studies was judged to be low enough to attempt meta-analysis. For continuous variables such as pain or function ratings, the measure of choice for generating pooled estimates of effect was the standardized mean difference (SMD) at the latest reported timepoint. For dichotomous variables (e.g., likelihood of treatment success or return to work), the rate ratio (RR) was
used. Primary meta-analyses focused on comparisons of fusion to a uniform comparator (e.g., discectomy); sensitivity analyses also were conducted comparing fusion to any available control population. Additional, sensitivity analyses were conducted comparing fusion to any available control in all available RCTs, regardless of study quality. Finally, while cohort and case series studies were not candidates for meta-analyses of treatment effect, qualitative findings from these studies are described for each measure of interest. Detailed evidence tables are presented in Appendix C for all key outcomes and study designs evaluated in this review.

7.8 Results

Overview of Evidence and Quality Assessment
Of the 14 RCTs identified (total N=1,209), nearly all (13) focused on patients with radiculopathic symptoms and radiographic evidence of nerve root compression; 1 included patients with evidence of disc herniation. Most available RCTs limited patients to those with single- or 2-level disease (see Appendix C, Tables C1 & C2). Importantly, we found no RCT evidence for spinal fusion procedures in patients with only generalized neck pain. Of the RCTs conducted in patients with radiculopathy, 2 also required evidence of cervical spondylosis, and 4 were conducted in patients with a recorded attempt at prior conservative treatment. Sample sizes were generally small, ranging from 10-50 patients per treatment arm. In the majority of RCTs, the comparator was discectomy, endoscopic discectomy, or microdiscectomy. Foraminotomy alone or in combination with discectomy was evaluated in 3 RCTs. Fusion was also compared alternatively to physical therapy or cervical collar in 1 RCT.

The 7 comparative cohorts included 929 patients evaluated in single- and multicenter studies, as well as nearly 100,000 patients assessed in a retrospective evaluation of the Nationwide Inpatient Sample database maintained by the U.S. Agency for Healthcare Research and Quality (AHRQ) (Shamji, 2008). Of the 7 studies, 1 was prospective. Comparators were varied, and included discectomy or microdiscectomy, laminoplasty, foraminotomy, and interdisciplinary rehabilitation. One of these studies explicitly compared anterior and posterior fusion techniques, and 3 compared fusion conducted in inpatient vs. outpatient settings.

Study quality is presented in Table 1 on the following page by study type and fusion comparator. A total of 10 of the 14 RCTs were identified as higher quality, including 1 comparing fusion to conservative management, 3 with minimally-invasive surgery as a comparator, and 6 comparing fusion to alternative open surgical approaches. Among the 7 comparative cohort studies, none were identified as “good” quality. Four studies were deemed to be of “fair” quality, including 1 comparing fusion to conservative management, 1 comparing fusion to foraminotomy, 1 comparing fusion procedures performed in inpatient vs. outpatient settings, and 1 comparing anterior to posterior fusion techniques.
Table 1. Studies of cervical fusion: study quality, by type of study and patient population.

<table>
<thead>
<tr>
<th>Study Type</th>
<th>Comparator</th>
<th>Higher</th>
<th>Lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCT</td>
<td>Physical Therapy/ Cervical Collar</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Discectomy</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Discectomy/Foraminotomy</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Microdiscectomy</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Endoscopic discectomy</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Endoscopic foraminotomy</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Comparative Cohort</td>
<td>Interdisciplinary Rehabilitation</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Laminoplasty</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Foraminotomy</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Fusion (in- vs. outpatient)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Fusion (anterior vs. posterior)</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

While it might appear that the evidence base for cervical fusion in patients with radiculopathic symptoms is relatively robust, further investigation revealed several concerns with study design, entry criteria, and protocol. For one, the relatively small sample sizes in most RCTs were evaluated in single institutions, limiting the generalizability that might come from multicenter evaluations of treatment effect. In addition, studies were inconsistent with respect to whether the same or different groups of surgeons performed the surgical procedures being compared in the study. Studies also differed with respect to post-surgery protocol. For example, in some, patients wore a cervical collar for varying periods of time. In others, patients received advice on immobilization, and in still others, no restrictions were placed on mobilization after surgery.

It should be further noted that, despite our intent to focus on studies evaluating patients presenting for treatment after attempts at short-term (4-6 weeks) conservative management, symptom duration was much longer at baseline in most RCTs. Of the 14 RCTs identified, only 3 reported a mean duration of symptoms less than 3 months. Several RCTs randomized patients only if they had a minimum of 2-4 years of symptoms.

Finally, all of the RCTs and comparative cohorts identified compared fusion to other forms of surgery or some form of conservative treatment; as a result, we found no comparative studies of fusion and minimally-invasive therapies such as spinal injections, radiofrequency denervation, or coblation nucleoplasty. As such, any conclusions drawn regarding the comparative benefits and risks of these procedures would be indirect only. Even indirect comparisons might in fact be problematic, as patients evaluated in studies of minimally-

© 2013, Institute for Clinical and Economic Review
invasive procedures may have less severe pain and radicular symptoms than those assessed in surgical studies.

Due to the concerns described above, and because none of these studies stand apart from others based on a distinctive combination of size, duration, quality, and/or generalizability, we have not labeled any comparative study as “key” for the purposes of this appraisal. Important elements of specific studies are highlighted in the report where relevant, however.

Training Standards and Relationship to Outcomes
The benefits and harms associated with all procedures vary to some extent according to the skills of the operator. This is certainly true for spinal fusion procedures in patients with chronic neck pain. Unfortunately, the relative importance of training and experience on patient outcomes has been little studied. Moreover, whatever impact training and skill differences has on the outcomes reported in the published literature, it is likely that even broader variations are seen in general clinical practice. Despite this, there are few widely-accepted training standards for preoperative, operative, and postoperative management of patients undergoing cervical fusion procedures.

Studies examining the relation of procedure volume to outcome in patients with cervical disorders are relatively few in number. In an examination of Medicare inpatient data, Taylor and colleagues observed lower mortality rates at hospitals performing a high volume of back and neck procedures (Taylor, 1997). An examination of cervical fusion procedures performed in California in 2008 suggests a high degree of variation in hospital volume (2-169 procedures), rates of complications (0-12.5%), and implant cost ($2,053-$14,382), but made no attempt to evaluate the degree to which these factors were correlated (Berkeley Center for Health Technology, 2009). Other studies focusing specifically on factors associated with outcomes in patient cohorts undergoing cervical spinal procedures were conducted only in high-volume centers (Fehlings, 2012) or using databases that did not include data on procedure volume by setting (Wang, 2007).
Clinical Benefits (KQ 1)

Findings are organized by type of comparator to fusion in the sections that follow. As mentioned previously, no comparative data were available comparing fusion to minimally-invasive non-surgical management options such as spinal injections, radiofrequency denervation, or coblation nucleoplasty.

1. Spinal Fusion vs. Conservative Treatment

A total of 2 studies, an RCT (Persson, 2001) and a cohort study (Mayer, 2002) compared cervical spinal fusion to conservative management approaches. These studies are summarized by outcome and population in the sections that follow.

“Treatment Success”
Neither of the studies comparing fusion to conservative treatment included measures of treatment success.

Pain and Function
A higher-quality RCT comparing outcomes for patients with cerviobrachial pain of > 3 months’ duration and nerve root compression due to spondylitic spurs randomized 81 patients into equal groups receiving anterior discectomy with fusion, physical therapy, or cervical collar immobilization; patients were followed for 16 months (Persson, 2001) (Appendix C, Tables C4 and C5). Pain was evaluated on a 100 mm VAS scale and measured at baseline, 4 months, and 16 months. As noted in Figure 3 on the following page, pain was reduced at 4 months among fusion patients vs. those receiving either form of conservative management; in the case of cervical collar immobilization, this difference was statistically significant. By month 16, however, the gap in pain scores between conservative management and fusion groups had narrowed, and VAS scores did not differ statistically at this timepoint. (NOTE: as a reminder, 5-10 point changes on VAS scores represent the minimum change that would be considered “clinically important”, and changes ≥ 30% from baseline represent those that would involve significant improvements for patients.)
Data on pain and function were also available from the comparative cohort study (Appendix C, Tables C13 and C14) prospectively evaluating the effects of anterior fusion in combination with an interdisciplinary rehabilitation program compared to interdisciplinary rehabilitation alone in 202 patients filing worker’s compensation claims for cervical spinal disorders (Mayer, 2002). Rehabilitation consisted of medically-supervised exercise, psychological counseling, workplace and vocational services, and case management. Patients were followed for 12 months. No statistically-significant differences between groups were noted during follow-up in measures of pain, including a 10 cm VAS scale and a questionnaire-based VAS scale known as the Million VAS (Anagnostis, 2003). In addition, no statistically-significant differences were observed in either mean cumulative scores for physical function or the percentage of patients with total cumulative functional scores demonstrating a “significant degree of pathology” (Mayer, 2002).

Quality of Life
In the Persson RCT, no statistically-significant differences between fusion and physical therapy were observed in Sickness Impact Profile or Mood Adjective Check List scores at any point during follow-up (Persson, 2001) (Appendix C, Table C6). Quality of life was also evaluated in the cohort study comparing fusion and interdisciplinary rehabilitation to rehabilitation alone (Mayer, 2002) (Appendix C, Table C15); however, such evaluation focused only on mean Beck Depression Inventory scores at 12 months without controlling for baseline levels or change from baseline.
Return to Work
Work-related outcomes were not evaluated in the Persson RCT (Appendix C, Table C7). In the Mayer comparative cohort study of fusion vs. interdisciplinary rehabilitation (Appendix C, Table C16), no statistically-significant differences were observed between treatment arms in the percentages of patients returning to full or modified work, returning to work at the same employer, or filing additional claims for recurrent injury at 12 months (Mayer, 2002).

2. Spinal Fusion vs. Discectomy and Foraminotomy

A total of 13 RCTs (9 higher-quality) examined the effects of fusion compared to discectomy or foraminotomy for cervical DDD. In 10 of these, the comparator was discectomy, microdiscectomy, or endoscopic discectomy alone. In one RCT, separate treatment arms receiving discectomy or foraminotomy were included (Wirth, 2000), and in another, comparator patients received a combination of discectomy and foraminotomy (Martins, 1976). Patients in a third RCT received endoscopic foraminotomy (Ruetten, 2008). Characteristics of all RCTs comparing fusion to these alternative surgical approaches are summarized in Table 2 on the following page as well as in Appendix C, Table C1.

Fusion was compared to foraminotomy in a single fair-quality comparative cohort study of 292 patients who were treated for radicular symptoms and followed for 6 years (Korinth, 2006).

“Treatment Success”
Treatment success was evaluated in 6 of the 9 higher-quality RCTs (Appendix C, Table C3). In one, 3 types of discectomy with fusion (autograft, polymethylmethacrylate [PMMA] graft, and titanium cage) were compared to microdiscectomy alone in 125 patients with single-level cervical disc disease and radiculopathy who were treated at a single institution in Switzerland and followed for 12 months (Barlöcher, 2002). Based on Odom’s criteria, a greater percentage of patients undergoing fusion with titanium cage had “excellent” or “good” results vs. microdiscectomy alone at 12 months (94.4% vs. 75.5%, p<.02); differences were not statistically-significant for the other fusion groups compared to microdiscectomy.

No statistically-significant differences were observed for measures of treatment success in the other higher-quality RCTs. In one, 91 patients with single-level radicular symptoms were randomized to receive fusion with bone graft, fusion with plating, or discectomy alone and followed for 4 years (Savolainen, 1998). Clinical outcome was based on patient self-report of symptoms, with “good” classified as absence of symptoms, “fair” described as improved but still with some complaints, and “poor” defined as no better or worse than the preoperative state. The distribution of clinical outcome was not statistically different between the 3 groups at any timepoint including after 4 years of follow-up.
Table 2. Characteristics of RCTs comparing fusion to discectomy or foraminotomy.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Comparator(s)</th>
<th>Sample Size</th>
<th>Study Duration (months)</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abd-Alrahman</td>
<td>1999</td>
<td>Discectomy</td>
<td>90</td>
<td>15</td>
<td>Lower</td>
</tr>
<tr>
<td>Barlöcher</td>
<td>2002</td>
<td>a--Microdiscectomy b--PMMA Fusion c--Fusion w/Cage</td>
<td>125</td>
<td>12</td>
<td>Higher</td>
</tr>
<tr>
<td>Dowd</td>
<td>1999</td>
<td>Discectomy</td>
<td>84</td>
<td>54</td>
<td>Lower</td>
</tr>
<tr>
<td>Hauerberg</td>
<td>2008</td>
<td>Discectomy</td>
<td>86</td>
<td>24</td>
<td>Higher</td>
</tr>
<tr>
<td>Martins</td>
<td>1976</td>
<td>Discectomy and foraminotomy</td>
<td>51</td>
<td>12</td>
<td>Higher</td>
</tr>
<tr>
<td>Oktenoglu</td>
<td>2007</td>
<td>Microdiscectomy</td>
<td>20</td>
<td>18</td>
<td>Higher</td>
</tr>
<tr>
<td>Rosenørn</td>
<td>1983</td>
<td>Discectomy</td>
<td>63</td>
<td>12</td>
<td>Higher</td>
</tr>
<tr>
<td>Ruetten</td>
<td>2009</td>
<td>Endoscopic discectomy</td>
<td>120</td>
<td>24</td>
<td>Higher</td>
</tr>
<tr>
<td>Ruetten</td>
<td>2008</td>
<td>Endoscopic foraminotomy</td>
<td>200</td>
<td>24</td>
<td>Lower</td>
</tr>
<tr>
<td>Savolainen</td>
<td>1998</td>
<td>a--Discectomy b--Fusion w/Plating</td>
<td>91</td>
<td>36</td>
<td>Higher</td>
</tr>
<tr>
<td>van den Bent</td>
<td>1996</td>
<td>Discectomy</td>
<td>81</td>
<td>24</td>
<td>Higher</td>
</tr>
<tr>
<td>Wirth</td>
<td>2000</td>
<td>a--Discectomy b--Foraminotomy</td>
<td>72</td>
<td>60</td>
<td>Lower</td>
</tr>
<tr>
<td>Xie</td>
<td>2007</td>
<td>a – Discectomy b–Instrumented Fusion</td>
<td>45</td>
<td>24</td>
<td>Higher</td>
</tr>
</tbody>
</table>

Another RCT evaluated 86 patients with symptoms of nerve root compression, who were randomized to fusion with titanium cage or discectomy alone and followed for 24 months (Hauerberg, 2008). Treatment outcome was defined dichotomously, based on full recovery or improved symptoms (“good”) or no change or worse symptoms (“poor”). The percentage of patients with good outcome did not differ statistically between groups at months 3, 12, and 24. A third RCT assessed clinical outcome using Odom’s criteria in 81 patients randomized to discectomy with PMMA fusion or discectomy alone who were followed for 24 months (van den Bent, 1996). The percentage of patients with an excellent or good result at 24 months was over 70% in both groups and did not differ statistically.

Hilibrand’s criteria were used to assess treatment outcome in an RCT comparing anterior discectomy with fusion to endoscopic anterior discectomy alone in 120 patients with radiculopathy secondary to disc herniation who were treated at a spine center in Germany and followed for 24 months (Ruetten, 2009). Approximately 90% of patients in both arms
were categorized as having “excellent” or “good” outcome on this scale at 24 months. No statistically-significant differences were noted at any earlier timepoint as well.

A final, shorter-term RCT evaluated anterior fusion with bone graft vs. discectomy and foraminotomy in 51 patients who were followed for 6 months (Martins, 1976). A study-defined outcome scale rated treatment results as fair, good, or excellent based on the number of preoperative symptoms that were improved, or poor if signs and symptoms were unchanged. At 6 months, two-thirds of patients in each treatment group reported excellent or good results; differences were not statistically significant.

While 6 higher-quality RCTs were available, only 2 of these used the same instrument to measure treatment success (Odom’s criteria). We conducted a random-effects meta-analysis examining the likelihood of treatment success for fusion vs. discectomy based on data from these 2 studies. No statistically-significant difference was observed, as noted in Figure 4 below (Rate ratio [RR]: 0.98, 95% CI: 0.81, 1.18; p=.84).

Figure 4. Meta-analysis of treatment success, based on Odom’s criteria.

<table>
<thead>
<tr>
<th>Study/Subgroup</th>
<th>Fusion* Events</th>
<th>Discectomy Events</th>
<th>Total Events</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bärlacher 2002</td>
<td>24</td>
<td>30</td>
<td>25</td>
<td>1.06 [0.81, 1.37]</td>
</tr>
<tr>
<td>van den Bent 1996</td>
<td>28</td>
<td>40</td>
<td>30</td>
<td>0.91 [0.70, 1.19]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>70</td>
<td>72</td>
<td></td>
<td>0.98 [0.81, 1.18]</td>
</tr>
</tbody>
</table>

*Anterior discectomy and fusion

A sensitivity analysis also was conducted including data from a lower-quality RCT (Abd Al-Rahman, 1999). Findings were also not statistically-significant in this analysis (RR: 0.90; 95% CI: 0.79, 1.03; p=.11)(Appendix C, Figure C2).

The retrospective comparative cohort study also evaluated outcome based on Odom’s criteria (Appendix C, Table C12). In this study, an assessment of 292 patients receiving either PMMA fusion or posterior foraminotomy (Korinth, 2006), long-term outcome was assessed after a mean of 6 years. The number of patients reporting excellent or good outcome was statistically-significantly greater in the fusion group (93.6% vs. 85.1%, p<.05).

Pain and Function
Information on pain was available from 6 of the 9 higher-quality RCTs comparing fusion to alternative surgical procedures (Appendix C, Tables C4 and C5). In the van den Bent RCT comparing discectomy with PMMA fusion to discectomy alone (van den Bent, 1996), the percentage of patients reporting relief of neck pain was statistically-significantly greater in the fusion group at 6 weeks (78% vs. 43% for discectomy, p=.04). Pain relief improved in
the discectomy group thereafter, however, and differences were no longer statistically-significant for the remainder of the 24-month follow-up. In the Hauerberg RCT of discectomy with titanium cage fusion vs. discectomy alone (Hauerberg, 2008), subjective assessments of arm and neck pain on a 0-11 scale did not differ statistically between treatment groups at 3, 12, and 24 months of follow-up. Pain was also evaluated in the 3 lower-quality RCTs; differences between treatment groups were either not statistically-significant or not measured statistically. In a smaller RCT (n=45) comparing discectomy with fusion, discectomy with instrumented fusion, and discectomy alone (Xie, 2007), no statistically-significant differences were observed in the percentage of patients reporting arm or neck pain at 12 months or in 3 distinct scales of the McGill Pain Questionnaire (pain rating index, number of words chosen, and present pain intensity).

The percentage of patients with any improvement in VAS neck or radicular pain was assessed at 2, 6, and 12 months in the RCT comparing 3 types of fusion to microdiscectomy alone (Barlöcher, 2002). At month 12, a statistically-significant difference in favor of fusion with titanium cage (97.3% vs. 81.9% for microdiscectomy alone, p<.05) was seen for radicular pain; no other measures differed statistically at any timepoint. In the previously-described RCT comparing anterior discectomy with fusion to endoscopic anterior discectomy alone (Ruetten, 2009), no statistical differences were observed between groups on VAS measures of neck or arm pain as well as on the German version of the North American Spine Society (NASS) pain score. Finally, change in VAS arm and neck pain on a 10-point scale was also assessed in a small RCT comparing 20 Turkish patients receiving anterior cervical microdiscectomy with or without fusion (Oktenoglu, 2007); patients were followed for 12 months. Statistically-significant improvement in arm pain was seen for both groups, while improvement in neck pain was only statistically-significant for the fusion group. When groups were compared, however, no statistical differences were noted for either arm or neck pain at 12 months.

While VAS-based measures of arm and neck pain were available in several surgical studies, sufficient information on data variance (i.e., standard deviations or standard errors) was not. Meta-analyses of VAS pain were therefore not possible. We did, however, conduct “sensitivity” meta-analyses in which we assumed VAS data were normally distributed and standard deviations could therefore be derived. In these analyses, we transformed VAS scores to a 10-point scale where necessary. No statistically-significant treatment effects were observed in these analyses, which are presented in Appendix C, Figures C3-C8.

Quality of Life
Data on quality of life were found in the Xie RCT comparing fusion with or without instrumentation to discectomy (Xie, 2007) (Appendix C, Table C6). Quality of life was assessed via the SF-36 instrument; in repeated-measures analyses, no statistically-significant differences were noted between groups in SF-36 total scores as well as scores for individual domains at 12 months of follow-up.

Return to Work
Data on return to work were available from 4 higher-quality RCTs (Appendix C, Table C7). The proportion of patients returning to work did not statistically differ at any timepoint.
between treatment groups in the previously-described Hauerberg and Xie RCTs. In a third RCT examining 63 patients undergoing anterior cervical discectomy with and without fusion who were followed for 12 months (Rosenørn, 1983), a statistically-significantly greater percentage of fusion patients had returned to work at 2, 3, 4, 6, and 9 weeks postoperatively; differences were no longer significant when measured at weeks 12, 26, and 52, however. Finally, the percentage of working individuals not yet able to return to work was assessed in the Barlöcher RCT (Barlöcher, 2002). At 6 months, a statistically-significantly lower percentage of patients undergoing fusion with titanium cage were not yet able to work as compared to microdiscectomy alone (5.5% vs. 18.1%, p<.05). However, differences were nonsignificant when evaluated at 12 months.

We also assessed return to work in a random-effects meta-analysis comparing fusion to discectomy in 4 higher-quality RCTs; for the Barlöcher RCT, we used the inverse of the data on “not yet able to work”. As shown in Figure 5 below and in Appendix C, Figure C9, the pooled estimate directionally favored discectomy in terms of return to work at 12-24 months, but this difference was not statistically significant.

**Figure 5. Meta-analysis of likelihood of return to work at 12-24 months, fusion vs. discectomy.**

<table>
<thead>
<tr>
<th>Study/Subgroup</th>
<th>Fusion Events</th>
<th>Discectomy Events</th>
<th>Total Weight</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barlöcher 2002</td>
<td>25 30</td>
<td>29 33</td>
<td>41.8%</td>
<td>0.95 [0.77, 1.16]</td>
<td></td>
</tr>
<tr>
<td>Hauerberg 2007</td>
<td>11 36</td>
<td>20 43</td>
<td>5.0%</td>
<td>0.66 [0.37, 1.18]</td>
<td></td>
</tr>
<tr>
<td>Rosenørn 1983</td>
<td>24 31</td>
<td>30 32</td>
<td>39.5%</td>
<td>0.83 [0.67, 1.02]</td>
<td></td>
</tr>
<tr>
<td>Xie 2007</td>
<td>12 15</td>
<td>10 12</td>
<td>13.6%</td>
<td>0.96 [0.67, 1.37]</td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>112 120</td>
<td>100.0%</td>
<td></td>
<td>0.88 [0.77, 1.01]</td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>72 89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Tau² = 0.00; Chi² = 2.43, df = 3 (P = 0.49); I² = 0%
Test for overall effect: Z = 1.85 (P = 0.06)

* Anterior discectomy and fusion

A second analysis of return to work was conducted to examine the possible differential impact of these procedures on short-term recovery. The Xie and Barlöcher RCTs provided data on return to work at 6 months following surgery. No statistically-significant differences between fusion and discectomy were noted (RR: 0.77; 95% CI: 0.28, 2.11; p=.62)(Appendix C, Figure C10).

**Publication Bias**

We also assessed the possibility of publication bias in analyses of return-to-work data (the number of studies was too small to generate statistics for measures of treatment success). Statistical analyses using Egger’s regression suggested no significant asymmetry in findings. Results of testing can be found in Appendix C, Figure C11.
Potential Harms (KQ 2)

Cervical spinal fusion is associated with a number of different harms, some of which are common to both fusion and other surgical alternatives, and others of which are unique to the fusion procedure itself (e.g., pseudarthrosis). Relevant harms are presented on Table 3 on page 76, described in the sections that follow, and categorized according to their timing. Complications occurring during the “perioperative” period (i.e., during the operative episode through the 30 days following) are categorized separately from longer-term harms and adverse events tallied through the remainder of follow-up.

Note that repeat surgery, while not technically a patient harm, is also presented, as it represents the potential for additional clinical risk and inconvenience and plays an important role in patient and clinician decision-making. For the perioperative period, data were collected on procedure revisions and returns to the operating room. For longer-term follow-up, data on reoperation and/or need for subsequent treatment were tallied.

Information on other procedure-related measures, such as procedure duration, perioperative blood loss, and hospital length of stay, while not a point of focus in this section, is nevertheless available for review in Appendix C (Tables C8 and C17 for RCTs and comparative cohort studies respectively).

Data on harms are presented for cervical fusion and its relevant comparators, including non-invasive interventions. While reported complications were rare in these populations, data do exist on requirements for subsequent treatment. Among all interventions for cervical DDD, however, good evidence on the true rates of serious harms is not available from published RCTs. Individual studies are too small to capture reliable data on complications that occur infrequently, and the relatively low rate of serious complications has led to standards for research reporting that often do not include formal reports on all complications. Other contributing factors to the dearth of data on complications include the general exclusion from many RCTs of patients with significant disability or otherwise at high risk, possible publication bias that disfavors reports of unsuccessful outcomes, and the relative short-term nature of most studies, which can fail to detect adverse outcomes associated with surgical interventions that do not manifest until later years. Because of these factors, as well as the often variable level of detail on complications available in study reports, no attempt was made to meta-analyze data on harms.

Information from the observational studies examined in this review suggests that risks of surgical interventions may be higher than reported in RCTs. For example, in the 14 RCTs examined for this appraisal, only 1 provided any data on perioperative mortality (Xie, 2007; rates were 0% in both treatment groups). In contrast, rates of in-hospital and 30-day mortality from large database studies, while <1%, are certainly nonzero (Shamji, 2008; Wang, 2007). Other harms that may not be apparent until after hospital discharge, such as pneumonia or venous thrombosis, appear relatively rarely in observational studies but are not reported at all in available RCTs.
1. **Data from RCTs and Comparative Cohorts**

Information on harms is presented on the following page in Table 3 for RCTs and comparative cohort studies evaluating fusion and its major comparators. As mentioned previously, certain types of complications and adverse outcomes were not reported in any available RCT or comparative cohort study. For example, data on perioperative paralysis or hemorrhage were not available in any RCT or comparative cohort study, and information on thrombosis was only reported in comparative cohort studies of fusion. In addition, there was a significant degree of overlap between treatment alternatives in reported rates of complications, with no clear or discernible pattern of differential rates for most complication types.

As described previously, perioperative mortality was rarely identified or reported in any RCT or comparative cohort study. Among perioperative complications, the most frequently reported for fusion included dysphagia, hoarseness, and infection. Not surprisingly, generally higher rates of dysphagia and hoarseness were reported for fusion, given the increasing use of anterior surgical approaches to this procedure (Shamji, 2008); rates were similar to those reported for discectomy in available RCTs, however. The upper end of the range of infection rates was higher with fusion relative to other procedures, which was related to 2 cases of donor site infection in a small RCT (Xie, 2007). Leakage of cerebrospinal fluid was rarely reported for any type of procedure.

Rates of longer-term events were annualized to account for differential follow-up across studies. Mortality was again rarely reported in RCTs or comparative cohort studies, with a significant degree of overlap by intervention. In some fusion studies, relatively high rates of adjacent segment disease and pseudarthrosis were reported. High rates of neurological decline were also reported in 1 RCT comparing fusion to discectomy, but these included measures of unchanged or worsened sensory loss in 1 RCT (Persson, 2001). Rates of repeat surgery or subsequent therapy were more commonly reported across study types. Again, ranges of reported rates overlapped significantly between them. The highest rate of reoperation was reported in the Persson RCT, where 8 of 30 patients underwent a second surgery by 16 months of follow-up (Persson, 1997); 6 of these had surgery adjacent to the initial surgical site.
Table 3. Reported ranges of rates of potential harms from RCTs and comparative cohort studies, by type of study and comparator.

<table>
<thead>
<tr>
<th>Type of Harm</th>
<th>Fusion</th>
<th>Conservative Rx</th>
<th>Surgical Approaches</th>
<th>No. of studies reporting harms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RCT</td>
<td>CC</td>
<td>RCT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RCTw</td>
<td>CCw</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Discectomy</td>
<td>Foraminotomy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Laminoplasty</td>
<td>Foraminotomy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of Patients with Event</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perioperative Events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality</td>
<td>0</td>
<td>0-0.05</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RCT</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RCTw</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCw</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Discectomy</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Foraminotomy</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Laminoplasty</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Foraminotomy</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Hemorrhage</td>
<td>NR</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>o Hematoma</td>
<td>1-6.6</td>
<td>0-0.8</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td>o Nerve Damage*</td>
<td>2.5-8</td>
<td>0.8-6</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NA</td>
<td>0-8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RCT</td>
<td>9-14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RCTw</td>
<td>0.6</td>
</tr>
<tr>
<td>o Paralysis</td>
<td>NR</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td>o Infection</td>
<td>0-13</td>
<td>0-0.02</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td>o Hoarseness</td>
<td>5-20</td>
<td>1.6</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0-8</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9-14</td>
<td>0.6</td>
</tr>
<tr>
<td>o Dysphagia</td>
<td>3-17.5</td>
<td>0-10</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15-25</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
</tr>
<tr>
<td>o Thrombosis</td>
<td>NR</td>
<td>0.02</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td>o CSF Leak</td>
<td>NR</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td>Return to OR</td>
<td>NR</td>
<td>0</td>
<td>NA</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long term Events†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Chronic pain</td>
<td>4.8</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.6</td>
<td>NR</td>
</tr>
<tr>
<td>o ASD</td>
<td>6.9-16.6</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>o Pseudarthrosis</td>
<td>8</td>
<td>3.2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td>o Neurological Decline*</td>
<td>3-23.3</td>
<td>0</td>
<td>14.2</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>27.2</td>
<td>NR</td>
</tr>
<tr>
<td>o Myelopathy</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>o Muscle weakness</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>o Paresthesia</td>
<td>14.2</td>
<td>3</td>
<td>8.2</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>o Subsequent Rx</td>
<td>0.5-21.7</td>
<td>0-3.2</td>
<td>13.8</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.1-9.8</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.1</td>
</tr>
</tbody>
</table>

NR: Not reported; NA: Not Applicable; ASD: Adjacent segment disease
¥: Conservative treatment = Physiotherapy; ¥¥: Conservative treatment = Interdisciplinary treatment; *: Nerve damage includes numbness, weakness and nerve palsy. **: Neurological decline includes sensory loss and neurological deficit; † Rates are annualized.
2. **Data from Fusion Case Series**

Long-term data on harms were reported in 46 reports of case series, describing events in nearly 6,000 patients. Follow-up ranged from 1 to 21 years in these studies. The most frequently-documented events included reoperation (n=23), pseudarthrosis (n=19), and adjacent segment disease (n=13), with ranges similar to those reported in RCTs and comparative cohorts. Mortality data were reported in 9 studies; rates ranged from 0-2.6% on an annualized basis.

In an effort to create an exhaustive list of long-term adverse events reported in patients undergoing cervical fusion, all reported harms from these studies are listed in Table 4 below.

**Table 4. Reported ranges of annualized rates of potential harms from fusion case series.**

<table>
<thead>
<tr>
<th>Type of Harm</th>
<th>Number of studies</th>
<th>% of patients with event (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>9</td>
<td>0-2.6</td>
</tr>
<tr>
<td>Complications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjacent segment disease</td>
<td>13</td>
<td>0-27</td>
</tr>
<tr>
<td>Arm pain</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>Donor site pain</td>
<td>1</td>
<td>8.7</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>2</td>
<td>3.6-33</td>
</tr>
<tr>
<td>Dysphonia</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>Hepatitis C infection</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>Hoarseness</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>Laryngeal paresis</td>
<td>2</td>
<td>0.5-1.4</td>
</tr>
<tr>
<td>Neck pain</td>
<td>3</td>
<td>0.3-2.5</td>
</tr>
<tr>
<td>New-onset radiculopathy</td>
<td>2</td>
<td>0.8-5.0</td>
</tr>
<tr>
<td>Paresis</td>
<td>1</td>
<td>1.3</td>
</tr>
<tr>
<td>Pseudarthrosis</td>
<td>19</td>
<td>0-8.5</td>
</tr>
<tr>
<td>Ptosis</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Tetraparesis</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>Upper limb numbness</td>
<td>2</td>
<td>0.3-2.2</td>
</tr>
<tr>
<td>Upper limb weakness</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>Worsening headache pain</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>Worsening lower extremity function</td>
<td>1</td>
<td>1.7</td>
</tr>
<tr>
<td>Worsening sensory function/strength</td>
<td>2</td>
<td>0.7-2.7</td>
</tr>
<tr>
<td>Worsening upper extremity function</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>Re-operation</td>
<td>23</td>
<td>0-6.8</td>
</tr>
<tr>
<td>Inability to return to work</td>
<td>2</td>
<td>0.5-4.3</td>
</tr>
</tbody>
</table>
3. **Data on General Surgical Risks**

As previously mentioned, we also sought current information on the general risks associated with any surgical procedure. These should be considered to be broad-based estimates only, and not indicative of what might be expected for cervical spinal fusion specifically.

Recent data on complication risks associated with all surgical procedures are available from a retrospective review of the records of 1,442 surgical patients at Emory University Hospital from 2009-2011 (Kassin, 2012). Complications occurring during the surgical hospitalization or the 30-day period following were identified in 455 of these patients (31.6%); hospital readmission during this period occurred in 163 patients (11.3%). Data on the frequency of complications by type can be found in Table 5 below. The most frequent complications included need for transfusion (16.6%) and wound complications (superficial or deep wound infection, wound disruption; 12.4%).

<table>
<thead>
<tr>
<th>Complication Type</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfusion</td>
<td>239</td>
<td>16.6</td>
</tr>
<tr>
<td>Wound (superficial/deep wound infection, wound disruption)</td>
<td>179</td>
<td>12.4</td>
</tr>
<tr>
<td>Bacteremia/sepsis/shock</td>
<td>105</td>
<td>7.3</td>
</tr>
<tr>
<td>Pulmonary (ventilation &gt;48h, pneumonia, unplanned intubation)</td>
<td>85</td>
<td>5.9</td>
</tr>
<tr>
<td>Urinary tract infection</td>
<td>40</td>
<td>2.8</td>
</tr>
<tr>
<td>Renal (acute renal failure, progressive renal insufficiency)</td>
<td>28</td>
<td>1.9</td>
</tr>
<tr>
<td>Vascular (DVT, PE, graft/prosthesis failure)</td>
<td>25</td>
<td>1.7</td>
</tr>
<tr>
<td>Cardiac (MI, cardiac arrest)</td>
<td>13</td>
<td>0.9</td>
</tr>
<tr>
<td>Neurologic (Stroke, Coma &gt;24h, peripheral nerve injury)</td>
<td>12</td>
<td>0.8</td>
</tr>
</tbody>
</table>

DVT: Deep vein thrombosis; PE: Pulmonary embolism; MI: Myocardial infarction

Because one of the intentions of the study described above was to examine readmission rates, patients who died during the index hospitalization were excluded from analysis. Perioperative mortality nevertheless remains a significant concern with surgery. Findings from a recent study examining in-hospital mortality among nearly 47,000 adults undergoing non-cardiac surgery at 498 hospitals in 28 European countries during a 1-week period in 2011 indicated a 4% mortality rate (Pearse, 2012). While the rate ranged widely by country (i.e., from 1.2% in Iceland to 21.5% in Latvia), mortality in nations with developed health systems such as France, Germany, and the UK was higher than anticipated, ranging from 2.5-3.6%.
Differential Effectiveness & Safety of Cervical Fusion in Key Patient Subgroups (KQ 3)

Data examining the differential effects of cervical fusion in key patient subpopulations were obtained from RCTs, comparative cohort studies and case series where available. Findings are reported below by type of study and subgroup; detailed information is also available in Appendix C, Tables C19, C20, and C23.

1. **Randomized Controlled Trials**

A total of 3 higher-quality RCTs included data on specific patient subgroups, as described below (see Appendix C, Table C19).

**Single- vs. 2-level Surgery**
In an RCT comparing anterior cervical discectomy with fusion to discectomy alone in 51 patients (Martins, 1976), the percentage of patients with excellent or good results (complete relief or minimal persistence of preoperative symptoms and abnormal signs improved or unchanged) was compared by level of surgery. Among patients undergoing single-level procedures, the percentage of those with excellent or good results was higher in the fusion group (82% vs. 66% for discectomy), while the rate for 2-level surgery was lower for fusion (26% vs. 63%). Rates were not tested statistically.

**Smoking Status**
In the RCT comparing fusion to physical therapy and cervical collar immobilization (Persson, 2001), the improvement in VAS pain among those undergoing surgery was found to be better among nonsmokers vs. smokers (p<.05), although the actual data on VAS changes among these subgroups are not provided.

**Gender**
In an RCT comparing anterior discectomy with and without fusion in 63 patients (40 men and 23 women) (Rosenørn, 1983), a higher percentage of males (n=16) undergoing fusion had excellent or good results (defined as return to previous occupation with no or minimal symptoms) at 12 months (15 [94%] vs. 19 [86%] for discectomy), while a lower percentage of females (n=13) had excellent or good results (5 [38%] vs. 8 [89%] for discectomy). While statistical testing was not performed for the stratified analysis across treatment groups, the percentage of patients with treatment success in the fusion group was greater (p<.005) among males vs. females.

2. **Comparative Cohort Studies**

Data were available from 4 comparative cohort studies; while all studies are presented in detail in Appendix C, Table C20, the focus of attention in the descriptions below is on the 2 studies that were fair-quality. In these studies, the subgroups of interest defined the study comparators (i.e., inpatient vs. outpatient fusion, posterior vs. anterior fusion).
Inpatient vs. Outpatient Fusion
The effects of fusion performed in inpatient vs. ambulatory care settings were assessed in a single fair-quality comparative cohort study. In this study, a comparison of 50 consecutive day-surgery patients to 53 retrospectively-analyzed inpatient controls (Silvers, 1996), no statistically-significant differences by setting were noted for functional outcomes, VAS pain, performance of activities of daily living, or return to work or normal activities. The rate of reoperation was numerically higher in the inpatient group (9.4% vs. 4.0% for outpatient), but this difference was not statistically tested.

Anterior vs. Posterior Fusion
Comparisons of anterior vs. posterior fusion techniques were performed in 1 fair-quality retrospective cohort study. This study examined differences between anterior and posterior fusion techniques using the U.S. Nationwide Inpatient Sample (NIS) database (AHRQ, 2012), and assessed hospital outcomes among nearly 100,000 patients undergoing anterior or posterior fusion for cervical DDD between 1988 and 2003 (Shamji, 2008). Patients were further stratified by whether their record included a diagnosis of myelopathy (approximately 75% of patients did not have such a diagnosis). On an unadjusted basis, patients undergoing posterior fusion experienced higher rates of death and major complications (p≤.001), regardless of the presence of myelopathy on the record. Length of stay and inflation-adjusted cost was also significantly increased, as shown in Table 6 below (data on myelopathy patients are displayed for context). After adjustment for differences in patient characteristics between groups, patients undergoing posterior fusion had significantly higher rates of all complications except for thrombophlebitis when compared to anterior fusion.

Table 6. Outcomes of anterior vs. posterior fusion surgery, stratified by neurological status (n=96,773).

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Anterior</th>
<th>Posterior</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Myelo</td>
<td>Myelo</td>
<td>No Myelo</td>
</tr>
<tr>
<td>Death (%)</td>
<td>0.05</td>
<td>0.52</td>
<td>0.36</td>
</tr>
<tr>
<td>Pulmonary embolism (%)</td>
<td>0.02</td>
<td>0.06</td>
<td>0.14</td>
</tr>
<tr>
<td>Pneumonia (%)</td>
<td>0.14</td>
<td>0.62</td>
<td>1.04</td>
</tr>
<tr>
<td>Transfusion (%)</td>
<td>0.34</td>
<td>1.02</td>
<td>3.33</td>
</tr>
<tr>
<td>Thrombophlebitis (%)</td>
<td>0.02</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>Infection %</td>
<td>0.02</td>
<td>0.10</td>
<td>0.36</td>
</tr>
<tr>
<td>Length of stay (mean days)</td>
<td>1.95</td>
<td>3.42</td>
<td>4.37</td>
</tr>
<tr>
<td>Inflation-adjusted cost ($)</td>
<td>20,639</td>
<td>28,581</td>
<td>34,963</td>
</tr>
</tbody>
</table>

Source: Shamji et al., J Neurosurg: Spine, 2008;9:10-16
Myelo: Myelopathy diagnosis on inpatient record
While the results of this study suggest higher rates of complications with posterior fusion, it is often the case that contemporary posterior techniques are reserved for patients with greater disability and/or spinal instability who require multi-level interevention (Caridi, 2011). Information from administrative databases is understandably limited in controlling for differences in clinical presentation.

While not considered a “comparative cohort” study per se, an earlier study analyzing data from the NIS also showed higher rates of complications and mortality for posterior vs. anterior fusion (Wang, 2007); in multiple logistic regression analyses, posterior fusion was significantly associated with a higher risk of in-hospital complications (but not mortality).

3. Fusion Case Series

A total of 28 fusion case series stratified data according to key patient subgroups of interest; findings are presented in Appendix C, Table C23 and described in further detail below.

Single- vs. Multi-Level Surgery
Subgroup analyses of patients undergoing single- and multi-level fusion procedures were analyzed in 17 case series. In most of these studies, increases in the number of levels involved were associated with increased rates of pseudarthrosis, although the statistical significance of any observed differences was often not tested. Reoperation rates and development of adjacent segment disease were assessed in 3 studies (Matsumoto, 2009; Heidecke, 2000; Bishop, 1996); no statistically-significant differences in these rates according to number of levels involved were observed. In one series, rates of dysphagia were reported for patients undergoing 1-, 2-, and 3+ level anterior fusion; these rates increased according to the number of levels involved (11% vs. 24% vs. 43%, significance not tested) (Riley, 2005). Findings from a later systematic review by the same author were similar in nature (Riley, 2010).

Smoking Status
Six case series examined the impact of pre-operative smoking status on adverse events and clinical outcomes. While 1 study described statistically-significantly fewer cases of pseudarthrosis among non-smokers (20% vs. 50% for smokers, p=.001) (Goldberg, 2002), 2 found no correlation between smoking status and development of pseudarthrosis or adjacent segment disease (Matsumoto, 2009; Bindal, 2007). In terms of clinical outcomes, 3 series evaluated the effect of smoking status on treatment success using Odom’s criteria. In one, data from a series of 144 patients indicated that smokers had a significantly (p=.008) higher rate of fair or poor outcomes (Jensen, 2009), although actual percentages were not reported. Another study (n=190) found non-smokers to have a significantly higher rate of excellent outcomes (43.0% vs. 27.3%, p<.03) (Hilibrand, 2001). A third smaller series (n=66) found no statistically-significant differences in this measure (Samartzis, 2005).
Gender
Clinical outcomes and adverse events did not statistically differ according to patient gender in 4 case series (Chen, 2009; Matsumoto, 2009; Bindal, 2007; Javid, 2001). Data from an additional series did find statistically-significantly greater rates of dysphagia (41.2% vs. 22.9% for women vs. men, p<.05) and dysphonia (28.2% vs. 8.6%, p<.05) among women (Yue, 2005b).

Anterior vs. Posterior Fusion
Fusion approach was evaluated in a single case series of 120 patients undergoing revision procedures for pseudarthrosis (Carreon, 2006). Over 3-4 years of follow-up, the rate of subsequent reoperation was lower among patients undergoing posterior revision (2.2% vs. 44.4%), although this difference was not tested statistically.

Duration of Symptoms
Four case series analyzed the impact of symptom duration on clinical prognosis and long-term outcomes. While 1 study found no correlation (Chen, 2009), 3 case series demonstrated significantly better outcomes in patients with shorter duration of symptoms (Kadoya, 2003; Hamburger, 2001; Heidecke, 2000). In one study, patients with a symptom duration of less than 3 months had a statistically-significantly higher rate of excellent outcomes based on Odom’s criteria as compared to those with symptoms > 12 months (48.9% vs. 33.3%, p<.03) (Hamburger, 2001). In a study of anterior microdiscectomy with fusion, no significant difference in the percentage of patients with a self-reported “good” outcome was seen in patients with radiculopathy symptoms; however, patients with myeloradiculopathy symptoms for < 1 year had a significantly higher rate of good outcome (78.9% vs. 50.0%, p<.01) (Heidecke, 2000). Finally, a significant (p=.01) correlation was observed between outcome scores derived from the Neurological Cervical Spine Scale and duration of symptoms. Those with symptom duration of 6 months or less had the highest mean score (3.3); scores declined with increasing symptom duration, culminating with score of 1.2 among patients with symptom duration > 4 years (Kadoya, 2003).

Age
Seven case series provided data on patient subgroups based on age. While 3 studies described no correlation between age and clinical outcomes such as the Neck Disability Index and the Hirabayashi recovery rate (Chen, 2009; Matsumoto, 2009; Bindal, 2007), 3 other case series demonstrated statistically-significant differences based on patient age in rates of adverse events and neurologic outcomes (Cabraja, 2011; Kadoya, 2003; Heidecke, 2000). For example, the rate of neurologic improvement significantly declined with increasing age, from 71.0% among those age < 40 years to 11.1% among patients age 70 years or older (p=.014) (Kadoya, 2003). An additional case series found a greater incidence of dysphagia in younger (mean age 48 years) vs. older (mean age 55 years) patients (Yue, 2005b); differences were not tested statistically, however.

Workers’ Compensation
Goldberg et al. evaluated outcomes in 80 patients with and without worker’s compensation (Goldberg, 2002). Patients underwent anterior discectomy and fusion with a mean follow-
up of 4 years. There were no statistically-significant differences in outcome based on Odom’s criteria; incidence of donor site pain and the development of pseudarthrosis also did not differ. Seventy percent of patients with workers’ compensation returned to work without restriction vs. 80% of patients without workers’ compensation, although this difference was not tested.

In a case series involving 66 patients undergoing ACDF, a subgroup analysis of work- and non-work-related injuries demonstrated no significant differences in outcome based on Odom’s criteria at 22 months of follow-up (Samartzis, 2005).

8. Decision Analytic Model (KQ 4)

8.1 Objectives
The primary objective of this decision analytic model was to assess the relative cost and cost-effectiveness of treatment pathways involving spinal fusion versus alternative treatments for the management of cervical degenerative disc disease.

8.2 Methods
Analytic Approach
The decision model designed for this evaluation was a Markov state transition model, as shown in Figure 6 below. The clinical status of patients is represented by one of the three mutually exclusive states in the model. The model states were chosen to be reflective of available clinical data and key clinical states that patients may experience in terms of treatment success (with or without complications or other adverse events). The model was designed to shift patients between the different model states at three-month intervals over a 1-3 year time horizon.

Figure 6. Markov Disease State Diagram for Cervical Degenerative Disc Disease

SAE=serious adverse event.
Perspective
We adopted a public payer perspective for the reference case (i.e., primary analysis), thus following the majority of recommendations for health care economic evaluation (Drummond, 2005). However, an analysis taking an employer perspective was also considered as a sensitivity analysis that incorporated consideration of work productivity gains or losses from treatment.

Type of Economic Evaluation
A cost-utility analysis was conducted that reported results as an incremental cost per quality adjusted life-year (QALY) gained. This form of analysis enables comparison of cost-effectiveness estimates for spinal fusion with other treatments funded by payers. It also facilitates the identification of subpopulations where use of spinal fusion appears to be more (or less) cost-effective. In the interest of transparency, however, major model outcomes are also presented in disaggregated form (e.g., outcomes, costs). In addition, the cost per additional “treatment responder” (i.e., percentage of patients with resolution of symptoms) was calculated to provide additional context to disaggregated findings.

Target Patient Population
The target population of the decision model was the same as the population assessed in the corresponding systematic review, namely adults with cervical DDD who continue to have severe cervical pain after an initial course of conservative therapy of 6 to 12 weeks. It was further assumed that patients’ neck pain was not a consequence of systemic disease or other excluded medical diagnosis/condition (e.g., trauma, malignancy). Consistent with the findings of the systematic review, the model focused attention on patients with radiculopathic symptoms in addition to axial neck pain. The analysis therefore did not include patients having only generalized neck pain, as there was no evidence to support use of surgery in this patient population.

Intervention and Comparators
Anterior cervical discectomy and fusion (ACDF) was chosen as the primary approach of spinal fusion for the analysis. The reference case analysis compared spinal fusion with conservative management with physical therapy. Transition probabilities for spinal fusion were derived using data from an RCT which compared fusion with cervical arthroplasty (Sasso, 2011). The Sasso et al. study provided transition probabilities at three, six, twelve, twenty-four, and forty-eight months after fusion – equivalent data were not available in the Persson et al. study. Clinical effects for physical therapy versus fusion were, however, estimated from Persson, 2001, where it was assumed that fusion would initially have more pronounced effects versus conservative management but these effects would diminish over time (40% better at 6 months; 20% better at 12 months, 10% better at 24 months; equal at 48 months). Comparisons of spinal fusion with other potential treatment alternatives of interest are also reported:

- Manual therapy with spinal manipulation
- Epidural steroid injections
- Posterior laminoforaminotomy
- Anterior discectomy without fusion
Outcome Measures
This evaluation assessed key clinical outcomes related to the diagnosis of cervical DDD including the proportion of patients who resolved cervical pain symptoms, the proportion of patients who had cervical pain symptoms, and the occurrence of rare but important adverse events such as perioperative complications (e.g., nerve damage, cerebrospinal fluid leak, new-onset radiating pain, and stroke or other thrombotic events), longer-term complications such as adjacent segment disease, and death. Costs related to treatments, total costs to the payer, and the impact of different treatment pathways on quality of life (as reflected by QALYs) were also calculated.

The model was also designed to allow “break-even” analyses of cost and effect size parameters to be conducted to investigate the values that would potentially make cervical fusion cost-neutral or cost-effective across different thresholds relative to nonsurgical management.

Time Horizon
For the reference case analysis, multiple time horizons ranging from 1 to 3 years were considered (a) to match the typical duration of follow-up in relevant clinical trials; and (b) to capture the heterogeneity of treatment effect over time.

Decision Modeling & Assumptions
The decision model designed for this evaluation is shown in Figure 6 on page 83. The use of decision modelling to assess the comparative value of healthcare interventions is widespread (Briggs, 1998; Beck, 1993). Models facilitate the synthesis of data from various sources and permit the evaluation of both costs and outcomes simultaneously (Briggs, 1998; Beck, 1993). Typically, models divide the disease in question into distinct states and transition probabilities are assigned for movement between these states over a discrete time period. As noted above, the decision analysis was based on a cohort level Markov model consisting of three states. The clinical outcomes of interest, the costs over the 1- to 3-year time horizons, and the impacts of the different treatments on QALYs were incorporated into the design of the model to compare different treatment pathways for cervical DDD. The probability of transitioning between states in the model was affected by treatment. A cycle length of 3 months was selected to allow modeling of changes in the therapeutic approach to management of patients in the model and to best align model design with outcomes reported in the most relevant clinical literature. To enhance transparency and use a more parsimonious model, the impact of adverse events for each treatment within a pathway were applied by incorporating cost and QALY decrements to relevant health states rather than developing a separate health state for adverse events.

The clinical outcomes that were used in the model to evaluate each treatment pathway included the proportion of patients who resolve cervical pain symptoms, the proportion of patients who continue to have cervical pain symptoms, and the number of deaths observed. Baseline transition probabilities for clinical outcomes were based on the Sasso trial.
Resolution of cervical pain symptoms was based on the definition applied in Sasso et al.; patients had to achieve all of the following: an improvement of 15 points in the NDI, neurological improvement, no serious (WHO grade-3 or 4) adverse events related to the implant or surgical implantation procedure, and no subsequent surgery or intervention that would be classified as a treatment failure.

Listed in Table 7 below are assumptions made in designing the model for this evaluation. Our model was based to some degree on past decision models associated with the management of cervical DDD (Carreon, 2012; Van der Velde, 2008) as well as on clinical studies comparing spinal fusion with conservative treatment or cervical arthroplasty (Persson, 2001; Sasso, 2011).

### Table 7. Key Model Assumptions

<table>
<thead>
<tr>
<th>Assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior to entering the model patients have had an initial trial of conservative management lasting between 6-12 weeks which did not resolve symptoms.</td>
</tr>
<tr>
<td>The gap in clinical benefit between spinal fusion and conservative treatment narrows over time as patients with conservative treatment reach similar pain and function levels, consistent with observations from clinical studies.</td>
</tr>
<tr>
<td>All forms of conservative management (e.g., physical therapy, spinal manipulation) have approximately equal clinical effectiveness.</td>
</tr>
<tr>
<td>Patients who do not resolve symptoms of cervical pain will have a reduction in health related quality of life and will incur costs equivalent to approximately $1,983 per cycle (~ equal to cost of ongoing physical therapy).</td>
</tr>
</tbody>
</table>

**Clinical Efficacy, Effectiveness and Safety**

The probabilities associated with the different clinical outcomes included in the decision model are described in Table 8 on the following page. For the comparison treatment pathways evaluated, the probabilities of transitioning between states in the model were based on the Sasso and Persson studies. Importantly, because data were limited comparing conservative treatment with fusion, we assumed that fusion was more effective than conservative therapy, although the relative benefits diminished over time, consistent with observations from clinical studies (Persson, 2001). We assumed that fusion was 40% better at 6 months (i.e., relative risk=0.60), 20% better at 12 months, 10% better at 24 months and equal at 48 months (see Table 8). We conservatively assumed that fusion was not associated with an increased risk of mortality in the reference case. However, a sensitivity analysis was conducted where we assumed a 30% risk reduction (RR=0.70) for conservative therapy (Kadaňka, 2011).
Risk multipliers for sub-group analysis were applied to the natural history data. Table 9 below provides a summary of the important relative risk measures used to evaluate each treatment comparison and an overview of the relevant probabilities included in the economic model for both deterministic and probabilistic analyses.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Base Estimate</th>
<th>Probability Distribution</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVENT RATE ADJUSTMENTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk of neck pain resolution or improvement vs spinal fusion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical therapy</td>
<td>See Table 1</td>
<td>Fixed</td>
<td>Sasso, 2011 ; Persson, 2001</td>
</tr>
<tr>
<td>Manual therapy/Spinal manipulation</td>
<td>See Table 1</td>
<td>Fixed</td>
<td>Sasso, 2011 ; Persson, 2001</td>
</tr>
<tr>
<td>Epidural spinal injections</td>
<td>0.39</td>
<td>Lognormal (-0.94, 0.42)</td>
<td>Clinical expert input</td>
</tr>
<tr>
<td>Laminoforaminotomy</td>
<td>0.98</td>
<td>Lognormal (0.02, 0.095)</td>
<td>Systematic Review</td>
</tr>
<tr>
<td>Anterior discectomy without fusion</td>
<td>0.98</td>
<td>Lognormal (0.02, 0.095)</td>
<td>Systematic Review</td>
</tr>
</tbody>
</table>

Beta distributions parameterized by alpha and beta, lognormal distributions parameterized by log means and log standard errors

The adverse events considered in the model are reported in Table 10 on the following page. As mentioned previously, to enhance transparency and use a more parsimonious model, the impact of adverse events for each treatment within a pathway were applied by incorporating cost and QALY effects (versus developing a separate state). To enhance transparency, these are applied via a separate analysis to ensure that such inclusion does not alter findings. We applied a “disutility” (i.e., an estimate of the decrement in quality of life) for radiating pain of 0.02264 to 80% of patients and applied the costs and utility decrements associated with a very small risk of stroke to the conservative management treatment options. We also applied an incremental cost of $29,721, associated with adjacent
segment disease and $1070 for perioperative complications for spinal fusion to 4.1% and 2% of patients, respectively (Sasso, 2011), as well as a disutility estimate of 0.20 (due to the need for subsequent surgery in these patients).

<table>
<thead>
<tr>
<th>Table 10. Adverse Events Incorporated in the Decision Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
</tr>
<tr>
<td>Physical therapy</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ACDF</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Valuing Outcomes**

Utility values for the model were derived from Richardson, 2012 (see Table 11 below) using their methods for predicting SF-36 scored based on Neck Disability Index (NDI) scores with the standard gamble method. Patients with cervical neck pain were assigned a mean utility score of 0.542 (based on an NDI score of 50.2 [Sasso, 2011]). A utility gain of 0.185 (equivalent to gaining 67.5 days of perfect health) was applied for resolution of symptoms of cervical neck pain. Because this utility gain exceeds values for resolving severe complications such as a myocardial infarction or stroke (Sullivan, 2006), we also conducted a sensitivity analysis using a smaller utility gain of 0.0737 (based on ICD-9 722 [Sullivan, 2006]). There are limitations to this estimate since the ICD-9 code does not consist entirely of surgical candidates with a diagnosis of cervical radiculopathy.

<table>
<thead>
<tr>
<th>Table 11. Utility Values Incorporated in the Decision Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable Description</strong></td>
</tr>
<tr>
<td>Utility value for patients experiencing cervical pain</td>
</tr>
<tr>
<td>Utility value for resolution of cervical pain symptoms</td>
</tr>
<tr>
<td>Utility for death</td>
</tr>
</tbody>
</table>
Costs and Resource Use
The cost and resource use variables used in the decision model are provided in Table 12 below. Costs of direct medical services were estimated using the 2012 Medicare fee schedule and data from the Washington State Health Care Authority.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Cost</th>
<th># of Units per Cycle</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACDF</td>
<td>$29,722</td>
<td>1</td>
<td>Washington HCA</td>
</tr>
<tr>
<td>Laminoforaminotomy</td>
<td>$29,556</td>
<td>1</td>
<td>Washington HCA</td>
</tr>
<tr>
<td>Discectomy</td>
<td>$22,284</td>
<td>1</td>
<td>Washington HCA</td>
</tr>
<tr>
<td>Physical therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>$76</td>
<td>1</td>
<td>Washington HCA</td>
</tr>
<tr>
<td>Therapy</td>
<td>$53</td>
<td>36</td>
<td>Washington HCA</td>
</tr>
<tr>
<td>Manual therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>$72</td>
<td>1</td>
<td>CMS Medicare, 2012B</td>
</tr>
<tr>
<td>Spinal X-ray</td>
<td>$98</td>
<td>1</td>
<td>CMS Medicare, 2012B</td>
</tr>
<tr>
<td>Therapy</td>
<td>$27</td>
<td>12</td>
<td>Washington HCA</td>
</tr>
<tr>
<td>Epidural steroid injections</td>
<td>$433</td>
<td>2</td>
<td>Washington HCA</td>
</tr>
</tbody>
</table>

Each cycle has a duration of three months.

To account for lost productivity, we assumed that patients receiving spinal fusion would return to work after 61 days, and a mean wage of $23.73 per hour was applied (Bureau of Labor Statistics, 2012; Heller 2009). We conservatively assumed that patients receiving conservative therapy miss approximately 10 more days of work at the same hourly rate.

Discount Rate
Future costs (i.e., beyond 1 year) and QALYs were discounted at an annual rate of 3% in the reference case analysis. Discount rates of 1% and 5% were explored in sensitivity analyses.

Sensitivity & Variability Analyses
Deterministic Sensitivity & Variability Analyses
Several univariate sensitivity and variability analyses were also conducted to explore the impact of varying parameter values and assumptions within the model. These included the following factors of interest: severity of cervical pain (mild, moderate, and severe); treatment effect (assume fusion is consistently 10% better; 50% better); perspective (payer, employer); time horizon (1 years, 2 years); and discount rates (1% and 5%).
Probabilistic Sensitivity Analyses
Probabilistic sensitivity analyses were performed using Monte Carlo simulation and adopted standard methods for defining uncertainty around parameters. Where possible, transition probabilities were characterized by beta distributions, relative risks by log normal distributions, utilities by beta distributions, and utility decrements by normal distributions. The costs of the different treatment strategies and cost-consequences associated with clinical outcomes were assumed to be fixed unless otherwise stated. Costs and effectiveness for each treatment pathway, as derived from 1,000 Monte Carlo iterations, were plotted as cost-effectiveness acceptability curves to convey the uncertainty in the results.

Subsequent Treatments
To enhance transparency, the reference case assumes that patients in the symptoms of cervical pain health state in years 2+ all receive the same treatment at the same cost and benefits. However, because this may not be reflective of clinical practice, additional analyses were performed in which additional costs were applied assuming that that 10% of patients who initially received conservative treatment and remain symptomatic after years 2 and beyond are treated with ACDF. In patients who initially received ACDF, we assumed that 4.5% over 48 months (or 1.15% per year) had a subsequent surgery (Sasso, 2011). We also conducted another sensitivity analysis in which we assumed that 100% of patients who were initially managed using conservative treatment and did not have resolution of symptoms would receive ACDF. A cost of $29,722 was applied for ACDF for sensitivity analyses.

Analyses
All analyses were conducted using Microsoft Excel 2010 (Microsoft Corporation, Seattle, Washington).
8.3 Results

Spinal Fusion vs. Conservative Treatment
Table 13 below provides the results for adult patients with severe cervical pain who have failed an initial conservative therapy course of 6-12 weeks’ duration. A greater percentage of patients using spinal fusion improved their cervical pain symptoms in the short-intermediate term than patients treated with conservative therapy, although the benefits diminished over time; the absolute difference in the percentage of patients with complete resolution of symptoms was >15% after 1 year, but only 3.6% after 3 years. The results from Table 13 are also presented graphically in Figure 7 on the following page.

<table>
<thead>
<tr>
<th>Time Horizon</th>
<th>Symptoms of Cervical Pain</th>
<th>Resolution of Symptoms of Cervical Pain</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spinal Fusion</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>25.8%</td>
<td>74.0%</td>
<td>0.2%</td>
</tr>
<tr>
<td>2 year</td>
<td>26.3%</td>
<td>73.2%</td>
<td>0.5%</td>
</tr>
<tr>
<td>3 year</td>
<td>26.5%</td>
<td>72.9%</td>
<td>0.7%</td>
</tr>
<tr>
<td><strong>Conservative Treatment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>40.6%</td>
<td>59.2%</td>
<td>0.2%</td>
</tr>
<tr>
<td>2 year</td>
<td>33.7%</td>
<td>65.9%</td>
<td>0.5%</td>
</tr>
<tr>
<td>3 year</td>
<td>30.1%</td>
<td>69.2%</td>
<td>0.7%</td>
</tr>
<tr>
<td><strong>Absolute Difference</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>-14.8%</td>
<td>14.8%</td>
<td>0.0%</td>
</tr>
<tr>
<td>2 year</td>
<td>-7.3%</td>
<td>7.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>3 year</td>
<td>-3.6%</td>
<td>3.6%</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>Number Needed to Treat to Obtain 1 Additional Patient with Resolved Symptoms</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>7</td>
<td>7</td>
<td>NA</td>
</tr>
<tr>
<td>2 year</td>
<td>14</td>
<td>14</td>
<td>NA</td>
</tr>
<tr>
<td>3 year</td>
<td>27</td>
<td>27</td>
<td>NA</td>
</tr>
</tbody>
</table>
Figure 7: Resolution of Cervical Pain Symptoms of Spinal Fusion versus Conservative Treatment in Reference Case Analysis

Following the pattern of clinical benefit, the incremental cost to achieve 1 additional treatment response (i.e., a patient with resolution of cervical pain symptoms) increased from $174,515 in year 1 to $677,917 in year 3 (see Table 14 below).

Table 14. Results of Reference Case Cost-Effectiveness Analysis

<table>
<thead>
<tr>
<th></th>
<th>Spinal Fusion</th>
<th>Conservative Therapy</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cost</td>
<td>% With Resolution of Symptoms</td>
<td>Cost</td>
</tr>
<tr>
<td>1 year</td>
<td>$31,981</td>
<td>74.0%</td>
<td>$6,153</td>
</tr>
<tr>
<td>2 year</td>
<td>$33,957</td>
<td>73.2%</td>
<td>$8,895</td>
</tr>
<tr>
<td>3 year</td>
<td>$35,897</td>
<td>72.9%</td>
<td>$11,204</td>
</tr>
</tbody>
</table>
From a cost-utility perspective, spinal fusion produced more QALYs than conservative therapy, albeit at an increased cost (Table 15 below). The incremental cost per QALY gained for spinal fusion versus conservative therapy ranged from $347,473 to $579,428 depending on the time horizon considered.

### Table 15. Results of Reference Case Cost-Utility Analysis

<table>
<thead>
<tr>
<th></th>
<th>Spinal Fusion</th>
<th>Conservative Therapy</th>
<th>Difference</th>
<th>Cost per QALY, Spinal Fusion versus Conservative Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cost</td>
<td>QALYs</td>
<td>Cost</td>
<td>QALYs</td>
</tr>
<tr>
<td>1 year</td>
<td>$31,981</td>
<td>0.6609</td>
<td>$6,153</td>
<td>0.6163</td>
</tr>
<tr>
<td>2 year</td>
<td>$33,957</td>
<td>1.3060</td>
<td>$8,895</td>
<td>1.2435</td>
</tr>
<tr>
<td>3 year</td>
<td>$35,897</td>
<td>1.9303</td>
<td>$11,204</td>
<td>1.8592</td>
</tr>
</tbody>
</table>

The expected values of costs, effects and ICERs for probabilistic sensitivity analysis did not vary significantly from the reference case analysis. The cost-effectiveness acceptability curve for spinal fusion versus conservative therapy for the 3 year time horizon indicated that there is a 0% chance that spinal fusion is the more cost-effective treatment strategy at a willingness-to-pay threshold of $100,000 per QALY, and a 1% chance at a willingness-to-pay threshold of $150,000 per QALY.
Results of Sensitivity & Variability Analyses (3-Year Time Horizon)
Table 16 below provides the results of deterministic sensitivity analyses for the 3-year time horizon. The incremental cost-effectiveness of spinal fusion versus conservative therapy varied across sensitivity and variability analyses, ranging from $101,796 when we assumed a 50% benefit of cervical fusion is maintained over time, to $1.9 million when more conservative benefits were assumed.

Included in Table 16 below are the results when the employer perspective is taken and estimated productivity losses/gains are included for the 3-year time horizon. The incremental cost-effectiveness of spinal fusion versus conservative therapy was $322,429 when this perspective was taken, indicating that considering productivity has relatively small impact on estimates of the incremental cost-effectiveness of surgery vs. conservative management.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Incremental Cost for Spinal Fusion</th>
<th>Incremental QALYs for Spinal Fusion</th>
<th>Incremental Cost per QALY Gained (ICER), Spinal Fusion versus Conservative Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Case</td>
<td>$24,693</td>
<td>0.0711</td>
<td>$347,473</td>
</tr>
<tr>
<td>a.) Treatment effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assume 10% benefit of fusion is maintained over time horizon</td>
<td>$26,095</td>
<td>0.0384</td>
<td>$680,376</td>
</tr>
<tr>
<td>Assume 25% benefit of fusion is maintained over time horizon</td>
<td>$23,630</td>
<td>0.0959</td>
<td>$246,441</td>
</tr>
<tr>
<td>Assume 50% benefit of fusion is maintained over time horizon</td>
<td>$19,521</td>
<td>0.1918</td>
<td>$101,796</td>
</tr>
<tr>
<td>b.) Severity of cervical pain (Reference case, NDI score=50.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate (NDI score = 40; utility gain of 0.12)</td>
<td>$24,693</td>
<td>0.0459</td>
<td>$537,914</td>
</tr>
<tr>
<td>Mild (NDI score = 25; utility gain of 0.033)</td>
<td>$24,693</td>
<td>0.0128</td>
<td>$1,929,072</td>
</tr>
<tr>
<td>c.) Discount rate (Reference case, 3% for both costs and QALYs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>$24,597</td>
<td>0.0733</td>
<td>$335,482</td>
</tr>
<tr>
<td>5%</td>
<td>$24,754</td>
<td>0.0697</td>
<td>$355,369</td>
</tr>
<tr>
<td>Scenario</td>
<td>Incremental Cost for Spinal Fusion</td>
<td>Incremental QALYs for Spinal Fusion</td>
<td>Incremental Cost per QALY Gained (ICER), Spinal Fusion versus Conservative Treatment</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------</td>
<td>------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Reference Case</td>
<td>$24,693</td>
<td>0.0711</td>
<td>$347,473</td>
</tr>
<tr>
<td>d.) Perspective</td>
<td>Employer perspective with productivity included</td>
<td>$22,913</td>
<td>0.0711</td>
</tr>
<tr>
<td>e.) Inclusion of severe adverse events</td>
<td>Inclusion of side effects in conservative &amp; spinal fusion</td>
<td>$25,932</td>
<td>0.0857</td>
</tr>
<tr>
<td>f.) Subsequent treatments</td>
<td>Assume 10% of patients receiving conservative therapy who have symptoms after treatment receive ACDF and 3.4% of patients receiving surgery have repeat surgery</td>
<td>$23,027</td>
<td>0.0711</td>
</tr>
<tr>
<td></td>
<td>Assume 100% of patients receiving conservative therapy who have symptoms after treatment receive ACDF; and 3.4% of patients receiving surgery have repeat surgery</td>
<td>$14,972</td>
<td>0.0711</td>
</tr>
</tbody>
</table>
Table 16. Results of Sensitivity & Variability Analyses (3-Year Time Horizon)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Incremental Cost for Spinal Fusion</th>
<th>Incremental QALYs for Spinal Fusion</th>
<th>Incremental Cost per QALY Gained (ICER), Spinal Fusion versus Conservative Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Case</td>
<td>$24,693</td>
<td>0.0711</td>
<td>$347,473</td>
</tr>
</tbody>
</table>

  g.) Utilities from Sullivan et al., 2006

  Utility gain of 0.0737 for resolving cervical pain symptoms (baseline utility of 0.778)

  | $24,693 | 0.0283 | $872,925 |

Spinal Fusion vs. Other Treatments

Table 17 below provides a summary of comparisons of spinal fusion with the other treatment pathways of interest. Spinal fusion was slightly more effective and slightly less expensive than laminoforaminotomy (based on Washington HCA reimbursement amounts), but was more expensive than discectomy, leading to an incremental cost-effectiveness ratio of $603,558. Fusion was most cost-effective when compared to initial therapy with epidural steroid injections, as we assumed that such injections would only be one-third as effective as fusion at a constant rate over time.

Table 17. Comparisons of Spinal Fusion to Other Treatments

<table>
<thead>
<tr>
<th>Comparator</th>
<th>Incremental Cost for Fusion</th>
<th>Incremental QALYs for Fusion</th>
<th>Incremental Response (% Improvement) for Fusion versus Comparator</th>
<th>Incremental Cost per QALY Gained (ICER)</th>
<th>Incremental Cost per Responder for Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual therapy with spinal manipulation</td>
<td>$28,465</td>
<td>0.0711</td>
<td>3.6%</td>
<td>$400,544</td>
<td>$781,460</td>
</tr>
<tr>
<td>Laminoforaminotomy (relative risk (RR) of cervical pain resolution, 0.98 vs. spinal fusion, mean cost, $29,556)</td>
<td>-$328</td>
<td>0.0115</td>
<td>2.2%</td>
<td>Less expensive and more effective</td>
<td>Less expensive and more effective</td>
</tr>
</tbody>
</table>
### Table 17. Comparisons of Spinal Fusion to Other Treatments

<table>
<thead>
<tr>
<th>Comparator</th>
<th>Incremental Cost for Fusion</th>
<th>Incremental QALYs for Fusion</th>
<th>Incremental Response (% Improvement) for Fusion versus Comparator</th>
<th>Incremental Cost per QALY Gained (ICER)</th>
<th>Incremental Cost per Responder for Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discectomy (RR of cervical pain resolution, 0.98 vs. spinal fusion, mean cost, $22,284)</td>
<td>$6,945</td>
<td>0.0115</td>
<td>2.2%</td>
<td>$603,558</td>
<td>$317,757</td>
</tr>
<tr>
<td>Epidural steroid Injection (RR of cervical pain resolution, 0.39 vs. spinal fusion, mean cost, $443 and 2 per cycle)</td>
<td>$18,831</td>
<td>0.2340</td>
<td>44.4%</td>
<td>$80,488</td>
<td>$42,375</td>
</tr>
</tbody>
</table>

### 8.4 Discussion

#### Summary of Main Findings from the Decision Analytic Model
To the best of our knowledge, this is the first economic analysis that has compared the use of spinal fusion with alternative treatments in the management of patients with cervical degenerative disc disease. We found that spinal fusion is more effective than conservative therapy initially but the difference in clinical benefits diminishes over time as the proportion of patients treated conservatively whose symptoms resolve “catches up” to that of patients treated surgically. The use of spinal fusion was associated with an increased cost relative to conservative therapies. Results showed that the incremental cost per QALY gained for spinal fusion ranged from $347,473 to $579,428 in the reference case analysis over a 1-3 year time horizon.

The cost of ACDF is approximately $30,000 per surgery. This exceeds the cost of managing patients with conservative treatments such as physical therapy, manual therapy, and steroid injections where costs are typically less than $2,000 per year, although patients may incur costs for the aforementioned treatments continuously over time unless they proceed to a state where they resolve their symptoms. The cost-effectiveness of spinal fusion therefore depends on the relative efficacy of spinal fusion in comparison to other treatments.

Our analysis also suggests that the cost-effectiveness of fusion is lower when surgery is performed on patients who have milder symptoms from cervical degenerative disc disease. Patients with more severe cervical DDD will start with lower quality of life, and if treatment is successful, they will achieve relatively large QALY gains. However, when patients have
a milder form of disease, they will have a quality of life closer to that of the general population, and therefore will have lower potential to achieve substantial improvements in quality of life, which will increase the cost-effectiveness ratio of fusion relative to alternative treatments.

Our results also suggest that the cost-effectiveness of ACDF will vary depending on the comparator. There were sharp differences in cost-effectiveness estimates depending on the comparator treatment. The cost differential between spinal fusion and laminoforaminotomy or discectomy is much smaller than the cost differential between spinal fusion and physical therapy or manual therapy. As a result, the cost-effectiveness estimates will vary considerably depending on the treatment comparator chosen.

**Strengths and Limitations**

There are a number of strengths of this study. First, this is the first independent economic analysis to assess the value of spinal fusion in patients with cervical DDD. Second, clinical inputs were derived from a systematic review of available clinical evidence. Third, our analysis followed a transparent and accepted methodology and adheres to the guidelines for the economic evaluation of health technologies. Fourth, wherever possible, the model used cost data for 2012 reflective of the Washington State Health Care Authority experience. Finally, detailed sensitivity and variability analyses were performed to examine the robustness of results to variation in model parameters and assumptions.

Despite its strengths, this analysis has certain limitations that warrant discussion. First, there were considerable gaps in available clinical evidence. As a result, findings from the economic evaluation should be interpreted with caution. There was a limited body of clinical evidence comparing spinal fusion with alternative treatments other than surgery. Further, there was considerable variation in patient populations, study design, and outcome definitions across studies, which limits the comparability of evidence. As a result, we were forced to make assumptions around the clinical effects of conservative treatment versus spinal fusion, using data derived from the Persson study. Because of the underlying uncertainty around clinical effects of fusion versus conservative treatment, we conducted numerous sensitivity analyses where efficacy for surgery versus conservative treatments was varied. Nonetheless, to obtain more robust comparative clinical and cost-effectiveness estimates, more clinical studies are needed comparing fusion with alternative conservative therapies after patients have had a trial of conservative therapy.

**Comparison to Other Economic Studies of ACDF**

Carreon conducted a cost-effectiveness analysis of single level ACDF in the United States (Carreon, 2012). Carreon reported cost per QALY estimates of $106,256 at year 1, $54,622 at year 2, $38,836 at year 3, $29,454 at year 4 and $24,479 at year 5, and concluded that single-level instrumented ACDF is both effective and durable resulting in QALY gained as compared to other widely accepted healthcare interventions. The results from Carreon differ from those reported in our study largely because the Carreon study did not compare fusion's costs and effects to any alternative treatment strategy. Standard guidelines for economic evaluation recommend that a comparator be used when conducting economic evaluations when they are available. Further, Carreon applied a cost of $15, 714 which is
**almost half the cost of fusion in the Washington State Health Care Authority.** If the cost of surgery (~$30,000) in Washington State Health Care Authority was applied in the Carreon analysis, the cost per QALY estimates in the Carreon study would approximately double.

Although the Carreon study applied similar utility gains for resolving cervical pain to those applied in this report (0.18), it is important to note that assumed utility gains are equivalent to gaining 67.5 days of perfect health each year. These estimates exceed those reported for severe complications such as stroke or myocardial infarction (Sullivan, 2006). Consequently, the utility gains in the Carreon analysis (and our reference case analysis) may not be generalizable to patients with less severe forms of cervical DDD, where utility gains may be less pronounced. These aforementioned issues (no comparator, assumed cost of fusion of $15,714, and assumed utility gain of 0.18) limit the generalizability of findings from the Carreon analysis to patient populations with more severe forms of cervical DDD in which conservative treatments are not an option. Indeed, we were able to produce findings similar to those reported in the Carreon analysis (~$40,000 per QALY gained at year 3) when we: 1) did not use a comparator (i.e., assumed conservative treatment costs $0 and fusion is 100% more effective), 2) assumed fusion costs $15,714, and 3) applied utility gains of 0.185 for resolution of cervical pain symptoms.

Our analysis is more generalizable to the Washington State Health Care Authority – we consider alternative treatment strategies; use Washington State Health Care Authority specific costs; and our analysis considers multiple sources of utilities, some of which may be more generalizable to broader patient populations such as those less severe forms of disease where an alternative conservative treatment remains an option despite a previous trial of conservative treatment. Nonetheless, further research is needed exploring how gains in health related quality of life vary by disease severity.

The only other study identified compared the cost-effectiveness of various types of fusion (i.e., plating and allograft vs. autograft) (Angevine, 2005). Because this type of comparison was outside the scope of our systematic review, it is not discussed in detail here.
9. Recommendations for Future Research

As documented in this appraisal report, despite the relatively high prevalence, clinical significance, and economic impact of chronic neck pain, syntheses of all the available medical literature continue to highlight many notable areas of uncertainty that hinder comparisons of the clinical effectiveness and value of major management options. In part the uncertainty is driven by the natural history of the condition itself, in that many patients may see improvement in their symptoms over time regardless of the intervention used. However, the current body of evidence also suffers from a lack of rigor and applicability, as nearly all randomized studies have been small, conducted in single, specialized centers, and have not employed standard techniques for measuring or evaluating outcomes.

Informed by the evidence gaps highlighted by our systematic review, we present below high-level recommendations for all research on spinal fusion for cervical neck pain and related symptoms arising from degenerative disc disease.

1. To date, very few RCTs have been conducted in a population of patients who have attempted conservative management for 4-6 weeks; most studies have involved individuals suffering from neck pain for months to years. Given that case series suggest that fusion’s benefits relative to conservative management decline the longer patients have been experiencing symptoms, trials should be developed to determine if “early” fusion provides a benefit relative to treatment alternatives. These trials should include a focus on functional outcomes, particularly return to work/work productivity. Although limited data suggest possible earlier pain relief with fusion, it is unknown whether this translates into earlier return to work and/or higher work productivity.

2. All clinical trials of management options for cervical pain and radiculopathy should use a common set of outcome measures, measured at the same time points so that results across trials can be more easily pooled and evaluated. For example, pain could be measured on a VAS that uses a uniform 100 mm measurement scale across trials, or through standardized and scored instruments such as the Million VAS or McGill Pain Questionnaire. In addition, despite recommendations from clinical societies to use validated functional instruments such as the Neck Disability Index, function has only been minimally-assessed in RCTs comparing fusion to treatment alternatives. Future RCTs should always include functional outcomes.

3. Health-related quality of life has not been adequately studied in RCTs of cervical spinal fusion. All future trials should directly assess quality of life by validated instruments such as the EQ-5D or SF-36.

4. Whenever possible, studies should capture healthcare utilization and cost outcomes. Information on indirect costs should also be collected, including detailed data on work loss, disability leave, and return to normal activities.
5. More trials are needed of treatment “pathways” or “algorithms” that would characterize care for patients who need more than an initial form of treatment. In particular, trials are needed of minimally-invasive nonsurgical interventions such as radiofrequency denervation and spinal injections to assess their potential for delaying or eliminating the need for surgery.

6. More trials are needed of patient preferences for different types of treatment options and how these preferences correlate with treatment outcomes.

7. More trials are needed in more broadly representative patient populations, including among clinical providers in the community, not just the elite practitioners at top academic sites.

8. Long-term observational studies are necessary to gain a better understanding of treatment-related harms, requirements for retreatment and additional treatment, and real-world healthcare utilization. In particular, clinical registries with careful notation of clinical and psychosocial patient characteristics prior to spinal fusion could be useful in understanding those factors associated with better overall outcomes.
REFERENCES


32. Chou R, Huffman LH, American Pain S, American College of P. Nonpharmacologic therapies for acute and chronic low back pain: A review of the


76. Hirai T, Okawa A, Arai Y et al. Middle-term results of a prospective comparative study of anterior decompression with fusion and posterior decompression with


135. Persson LCG, Lilja A. Pain, coping, emotional state and physical function in patients with chronic radicular neck pain. A comparison between patients treated


APPENDIX A
Measures of Treatment Success

1. Hilibrand criteria


<table>
<thead>
<tr>
<th>Outcome</th>
<th>Pain</th>
<th>Medication</th>
<th>Activity</th>
<th>Work status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>None</td>
<td>None</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Good</td>
<td>Mild</td>
<td>Occasional use of non-steroidal anti-inflammatory drugs</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Fair</td>
<td>Moderate</td>
<td>Frequent use of non-steroidal anti-inflammatory drugs</td>
<td>Restricted</td>
<td>Limited</td>
</tr>
<tr>
<td>Poor</td>
<td>Severe</td>
<td>Oral use of narcotics</td>
<td>Incapacitated</td>
<td>Disabled</td>
</tr>
</tbody>
</table>

2. Odom’s Criteria

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Patient characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>No complaints referable to cervical disease. Patients able to carry on daily occupations without impairment.</td>
</tr>
<tr>
<td>Good</td>
<td>Intermittent discomfort related to cervical disease which did not significantly interfere with work.</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>Subjective improvement with limited physical activity.</td>
</tr>
<tr>
<td>Poor</td>
<td>No improvement/ worse compared to pre-operative condition.</td>
</tr>
</tbody>
</table>

3. **Japanese Orthopaedic Association (JOA) Cervical Myelopathy Evaluation Questionnaire**

With regard to your health condition during the last week, please circle the number of the one answer that best applies for each of the following questions. If your condition varies depending on the day or the time, circle the number of the answer that applies when your condition was at its worst.

**Q1-1 While in the sitting position, can you look up at the ceiling by tilting your head upward?**
1) Impossible 2) Possible to some degree (with some effort)
3) Possible without difficulty

**Q1-2 Can you drink a glass of water without stopping despite the neck symptoms?**
1) Impossible 2) Possible to some degree
3) Possible without difficulty

**Q1-3 While in the sitting position, can you turn your head toward the person who is seated to the side but behind you and speak to that person while looking at his/her face?**
1) Impossible 2) Possible to some degree
3) Possible without difficulty

**Q1-4 Can you look at your feet when you go down the stairs?**
1) Impossible 2) Possible to some degree
3) Possible without difficulty

**Q2-1 Can you fasten the front buttons of your blouse or shirt with both hands?**
1) Impossible 2) Possible if I spend time
3) Possible without difficulty

**Q2-2 Can you eat a meal with your dominant hand using a spoon or a fork?**
1) Impossible 2) Possible if I spend time
3) Possible without difficulty

**Q2-3 Can you raise your arm? (answer for the weaker side)**
1) Impossible
2) Possible up to shoulder level
3) Possible although the elbow and/or wrist is a little flexed
4) I can raise it straight upward
Q3-1 Can you walk on a flat surface?
1) Impossible
2) Possible but slowly even with support
3) Possible only with the support of a handrail, a cane, or a walker
4) Possible but slowly without any support
5) Possible without difficulty

Q3-2 Can you stand on either leg without the support of your hand? (Do you need to support yourself?)
1) Impossible with either leg
2) Possible on either leg for more than 10 seconds
3) Possible on both legs individually for more than 10 seconds

Q3-3 Do you have difficulty going up stairs?
1) I have great difficulty
2) I have some difficulty
3) I have no difficulty

Q3-4 Do you have difficulty with one of the following motions: bending forward, kneeling, or stooping?
1) I have great difficulty
2) I have some difficulty
3) I have no difficulty

Q3-5 Do you have difficulty walking more than 15 minutes?
1) I have great difficulty
2) I have some difficulty
3) I have no difficulty

Q4-1 Do you have urinary incontinence?
1) Always
2) Frequently
3) When retaining urine over a period of more than 2 hours
4) When sneezing or straining
5) No

Q4-2 How often do you go to the bathroom at night?
1) Three times or more
2) Once or twice
3) Rarely

Q4-3 Do you have a feeling of residual urine in your bladder after voiding?
1) Most of the time
2) Sometimes
3) Rarely

Q4-4 Can you initiate (start) your urine stream immediately when you want to void?
1) Usually not
2) Sometimes
3) Most of the time

Q5-1 How is your present health condition?
1) Poor
2) Fair
3) Good
4) Very good
5) Excellent
Q5-2 Have you been unable to do your work or ordinary activities as well as you would like?
1) I have not been able to do them at all.
2) I have been unable to do them most of the time.
3) I have sometimes been unable to do them.
4) I have been able to do them most of the time.
5) I have always been able to do them.

Q5-3 Has your work routine been hindered because of the pain?
1) Greatly 2) Moderately 3) Slightly (somewhat)
4) Little (minimally) 5) Not at all

Q5-4 Have you been discouraged and depressed?
1) Always 2) Frequently 3) Sometimes
4) Rarely 5) Never

Q5-5 Do you feel exhausted?
1) Always 2) Frequently 3) Sometimes
4) Rarely 5) Never

Q5-6 Have you felt happy?
1) Never 2) Rarely 3) Sometimes
4) Almost always 5) Always

Q5-7 Do you think you are in decent health?
1) Not at all (my health is very poor)
2) Barely (my health is poor)
3) Not very much (my health is average)
4) Fairly (my health is better than average)
5) Yes (I am healthy)

Q5-8 Do you feel your health will get worse?
1) Very much so
2) A little bit at a time
3) Sometimes yes and sometimes no
4) Not very much
5) Not at all
On a scale of 0 to 10, regarding 0 as “no pain (numbness) at all” and 10 as “the most intense pain (numbness) imaginable,” mark a point between 0 and 10 on the lines below to show the degree of your pain or numbness when your symptom was at its worst during the last week.

If you feel pain or stiffness in your neck or shoulders, mark the degree.
0___________________________________ 10

If you feel tightness in your chest, mark the degree.
0___________________________________10

If you feel pain or numbness in your arms or hands, mark the degree. (If there is pain in both limbs, judge the worse of the two.)
0___________________________________10

If you feel pain or numbness from chest to toe, mark the degree
0___________________________________10


4. Hirabayashi Recovery Rate

Hirabayashi Recovery rate= \[ \frac{\text{Post operative condition} - \text{Preoperative condition}}{\text{Normal condition} - \text{Preoperative condition}} \] * 100

Study Quality Rating Criteria

**Study Quality Rating System for Randomized Controlled Trials**

A) Was the method of randomization adequate?
   The randomization sequence should be random and unpredictable, using methods such as a computer-generated random-number sequence or use of sealed envelopes. Allocation methodology based on date of birth or hospital numbers is not adequate.

B) Was the treatment allocation concealed?
   Treatment assignment is conducted by an independent party without information about patient eligibility or any patient-specific data.

C) Were groups similar at baseline regarding the most important prognostic factors?
   For assignment of a “yes” score, patient demographics including age, sex, duration of symptoms, should be similar at baseline.

D) Was the patient blinded to the intervention?
   Sufficient information regarding patient blinding should be provided.

E) Was the care provider blinded to the intervention?
   Sufficient information regarding caregiver blinding should be provided.

F) Was the outcome assessor blinded to the intervention?
   Sufficient information regarding blinding of the outcome assessment should be provided.

G) Were co-interventions avoided or similar?
   Co-interventions should be avoided, or be similar, in the trial design.

H) Was adherence acceptable in all groups?
   Based on factors such as intensity and duration of interventions, frequency of sessions, the reviewer will assess if adherence is acceptable.

I) Was the dropout rate described and acceptable (≤15%)?
   All patient dropouts should be described or accounted for without exceeding 15% per treatment arm.

J) Was the timing of the outcome assessment in all groups similar?
   Outcome assessment should occur at similar time points for all intervention arms, with evaluation of all important outcomes.

K) Did the analysis include an intention-to-treat analysis?
   All randomized patients are evaluated according to the original allocation, regardless of adherence, and < 5% of randomized patients are excluded.

Each question is scored as Yes/No/Don’t Know with every “Yes” score assigned one point. If more than 50% of all possible points are awarded, the study is rated as “high” quality, e.g. a study is scored as 6/11 points and is a “high” quality study.

Study Quality of Comparative Cohort Studies
The quality of all included comparative cohorts was assessed using the framework employed by the U.S. Preventive Services Task Force (USPSTF Procedure Manual, 2008). Studies were rated “good,” “fair,” or “poor,” using the criteria described below:

- **Good:** Comparable group are assembled initially and maintained throughout the study (follow-up of at least 80%); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; important outcomes are considered; and appropriate attention to confounders in analysis. Intention to treat analysis is used.

- **Fair:** Generally comparable groups are assembled initially but some question remains whether moderate differences occurred in follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention to treat analysis is done.

- **Poor:** Any of the following problems exist: (1) groups assembled initially are not close to being comparable or maintained throughout the study; (2) unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); (3) key confounders are given little or no attention; and (4) intention to treat analysis is *not* conducted.
APPENDIX B
Search Strategy for OVID
Databases searched:
• Medline 1996 to Present with Daily Update
• EBM Reviews – Cochrane Central Register of Controlled Trials 4th Quarter 2012

1. Randomized controlled trial.pt
2. Controlled clinical trial.pt
3. randomized controlled trials/
4. random allocation/
5. double blind method/
6. single blind method/
7. 1 or 2 or 3 or 4 or 5 or 6
8. Animal /not human.mp
9. 7 not 8
10. Clinical trial.pt
11. Exp Clinical trial/
12. (clin$ adj25 trial$).tw
13. ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).tw
14. Placebos/
15. Placebo$.tw
16. Random$.tw
17. Research design/
18. (latin adj square).tw
19. 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18
20. 19 not 8
21. 20 not 9
22. Comparative study/
23. Exp Evaluation Studies/
24. Follow-up studies/
25. Prospective studies/
26. (control$ or prospective$ or volunteer$).tw
27. Cross-over studies/
28. 22 or 23 or 24 or 25 or 26 or 27
29. 28 not 8
30. 29 not (9 or 21)
31. 9 or 21 or 30
32. Exp cervical vertebrae/
33. Cervical.mp
34. Degenerative.mp
35. 32 or 33 or 34
36. Fusion.mp
37. Exp spinal fusion/
38. Interbody.mp
39. Spondylodes*.mp
40. 36 or 37 or 38 or 39
41. 31 and 35 and 40

Search limited to studies published from 2000 to present, English language only and filtered by randomized controlled trials.

Cohort search limited to Medline with focused elimination of animal and cadaver studies.

**Search Strategy for EMBASE**

1. 'clinical article'/exp AND [embase]/lim AND [1996-2013]/py
2. 'clinical study'/exp AND [embase]/lim AND [1996-2013]/py
3. 'clinical trial'/exp AND [embase]/lim AND [1996-2013]/py
4. 'controlled study'/exp AND [embase]/lim AND [1996-2013]/py
5. 'multicenter study'/exp AND [embase]/lim AND [1996-2013]/py
6. 'randomized controlled trial'/exp AND [embase]/lim AND [1996-2013]/py
7. 'major clinical study'/exp AND [embase]/lim AND [1996-2013]/py
8. 'phase 3 clinical trial'/exp AND [embase]/lim AND [1996-2013]/py
9. 'phase 4 clinical trial'/exp AND [embase]/lim AND [1996-2013]/py
10. 'crossover procedure'/exp AND [embase]/lim AND [1996-2013]/py
11. 'double blind procedure'/exp AND [embase]/lim AND [1996-2013]/py
12. 'single blind procedure'/exp AND [embase]/lim AND [1996-2013]/py
13. 'placebo'/exp AND [embase]/lim AND [1996-2013]/py
14. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13
15. 'human'/exp AND [embase]/lim AND [1996-2013]/py
16. 'nonhuman'/exp AND [embase]/lim AND [1996-2013]/py
17. 'animal'/exp AND [embase]/lim AND [1996-2013]/py
18. 'animal experiment'/exp AND [embase]/lim AND [1996-2013]/py
19. #16 OR #17 OR #18
20. #15 AND #19
21. #14 NOT #19
22. #14 AND #20
23. #21 OR #22
24. 'cervical spine'/exp AND [embase]/lim AND [1996-2013]/py
25. cervical:de,ab,tii AND [embase]/lim AND [1996-2013]/py
26. degenerative:de,ab,tii AND [embase]/lim AND [1996-2013]/py
27. #24 OR #25 OR #26
28. fusion:de,ab,tii AND [embase]/lim AND [1996-2013]/py
29. 'spine fusion'/exp AND [embase]/lim AND [1996-2013]/py
30. interbody:de,ab,tii AND [embase]/lim AND [1996-2013]/py
31. spondylodes*:de,ab,tii AND [embase]/lim AND [1996-2013]/py
32. #28 OR #29 OR #30 OR #31
33. #23 AND #27 AND #32
34. #27 AND #32
35. #34 AND 'randomized controlled trial'/de

Additional databases searched
• OT Seeker: Occupational Therapy Systematic Evaluation of Evidence
• PEDro: Physiotherapy Evidence Database