

Draft Comparative Effectiveness Review

Number XX

Spinal Fusion for Treating Painful Lumbar Degenerated Discs or Joints

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No.

Prepared by:

[to be added in final draft]

Investigators:

[to be completed in final draft]

This information is distributed solely for the purposes of predissemination peer review. It has not been formally disseminated by the Agency for Healthcare Research and Quality. It does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services (AHRQ) determination or policy.

AHRQ Publication No. xx-EHCxxx

<Month Year>

Statement of Funding and Purpose

This report is based on research conducted by the Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. XXX-XXX-XXXX). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

Disclaimer Regarding 508-Compliance

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectivehealthCare@ahrq.hhs.gov.

Financial Disclosure Statement

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Standard Suggested Citation

Suggested citation: To be added to final report.

Preface

The Agency for Healthcare Research and Quality (AHRQ) conducts the Effective Health Care Program as part of its mission to organize knowledge and make it available to inform decisions about health care. As part of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Congress directed AHRQ to conduct and support research on the comparative outcomes, clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services to meet the needs of Medicare, Medicaid, and the Children's Health Insurance Program (CHIP).

AHRQ has an established network of Evidence-based Practice Centers (EPCs) that produce Evidence Reports/Technology Assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care. The EPCs now lend their expertise to the Effective Health Care Program by conducting comparative effectiveness reviews (CERs) of medications, devices, and other relevant interventions, including strategies for how these items and services can best be organized, managed, and delivered.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about systematic reviews, see <http://www.effectivehealthcare.ahrq.gov/reference/purpose.cfm>

AHRQ expects that CERs will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that consumers who make decisions about their own and their family's health can benefit from the evidence.

Transparency and stakeholder input from are essential to the Effective Health Care Program. Please visit the Web site (<http://www.effectivehealthcare.ahrq.gov>) to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input. Comparative Effectiveness Reviews will be updated regularly.

We welcome comments on this CER. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Elise Berliner
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

[to be inserted for final version]

Key Informants

[to be inserted for final version]

Technical Expert Panel

[to be inserted for final version]

Peer Reviewers

[to be inserted for final version]

Spinal Fusion for Treating Painful Lumbar Degenerated Discs or Joints

Structured Abstract

Objectives: To assess whether previous research is sufficient to support evidence-based conclusions about the benefits and harms of lumbar fusion relative to nonsurgical treatments or other invasive treatments or to support conclusions about outcomes following the use of different fusion strategies.

Data Sources: Database searches included MEDLINE, PreMEDLINE, EMBASE, CINAHL, and The Cochrane Library. Hand-search methods included review of studies recommended by public comment on key questions and review of reference lists.

Review Methods: Methodologists developed key questions and the protocol in cooperation with the Agency for Healthcare Research and Quality, Key Informants, and Technical Expert Panel. Original research was systematically searched, selected, and reviewed and patient-oriented outcomes of interest such as pain, function and adverse events were assessed. Extracted data included the following: study, patient, and treatment characteristics; study methodology (risk-of-bias assessment); study outcomes data; and prognostic factors.

Results: The overall evidence base for this report consisted of 25 studies (2 studies addressed more than 1 question): 5 studies compared fusion surgery with continued noninvasive treatment, 3 compared fusion surgery with other invasive procedures (e.g., decompressive laminectomy), 10 compared different spinal fusion approaches and techniques, and 7 studies considered patient or treatment factors associated with patient outcomes following spinal fusion. For all but one of the comparative studies that met the inclusion criteria for this report, the overall risk-of-bias rating was moderate. The moderate rating was largely because of lack of concealment of allocation and/or blinding of patients or outcome assessors to treatment received, or not reporting if concealment or blinding took place in the study. One study earned a high risk-of-bias rating due to high treatment crossover and other limitations that compromised the randomization of the study. Limited evidence (of low quantity and consistency) suggest the following findings: 1) fusion improves back pain relief and function compared to physical therapy at 2-year followup; 2) arthroplasty is associated with shorter surgical time, less blood loss, and shorter inpatient stays in adults with degenerated discs; and 3) rhBMP-2 is associated with less blood loss than autogenous bone graft, while surgery time and length of hospital stay do not differ substantially for these products for adults undergoing fusion for low back pain due to degenerated disc(s).

Conclusions: Overall, limited evidence suggests that spinal fusion compared with physical therapy improves pain and function for adults undergoing fusion for low back pain due to disc degeneration. Because of insufficient reporting and variation in surgical methods used in the different studies, the incidence of adverse events (serious and minor) associated with fusion could not be determined conclusively. The evidence was insufficient to draw evidence-based conclusions for the benefits and harms of spinal fusion for patients with degenerative stenosis or degenerative spondylolisthesis of the lumbar spine. The evidence was also largely insufficient to draw conclusions about the benefits and harms of fusion compared with other invasive treatments or different fusion approaches or techniques. Thus, future research is needed in these areas. However, many of the studies reviewed in full for this review were ultimately excluded for lack of relevance to modern treatment practices in the United States. Because implantable devices are frequently replaced by new products and generations of products, either by product line updates or withdrawal of previous implants and instrumentation from the market because of adverse events, ongoing clinical studies of new devices and materials are needed.

Contents

Executive Summary	ES-1
Background.....	ES-1
Objectives	ES-2
Methods.....	ES-4
Results	ES-6
Summary of Conclusions and Strength of Evidence.....	ES-12
Applicability	ES-13
Research Gaps	ES-13
Conclusions.....	ES-14
Introduction	1
Degenerative Lumbar Discs or Joints	1
Noninvasive Treatments for Degenerative Lumbar Conditions.....	2
Invasive Treatments for Degenerative Lumbar Conditions.....	5
Lumbar Fusion.....	6
Assessment Methods	7
Key Questions and Scope	7
Criteria for Study Inclusion/Exclusion.....	12
Publication Criteria.....	12
Study Design Criteria	12
Patient Criteria	13
Treatment Criteria	13
Data Criteria.....	13
Searching for the Evidence: Literature Search Strategies for Identifying Relevant Studies To Answer the Key Questions.....	13
Abstracting and Managing Data	14
Assessing Individual Studies' Methodological Quality (i.e., Risk of Bias)	15
Synthesizing Data	17
Strength-of-Evidence Grading.....	19
Assessing Applicability.....	20
Results	21
Introduction.....	21
Results of Literature Searches	21
Description of Included Studies	26
Key Question 1: Spinal Fusion Compared to Continued Noninvasive Treatment for Painful Lumbar Degenerated Discs.....	27

Key Points: Fusion Versus Physical Therapy	27
Detailed Synthesis	27
Strength-of-Evidence Ratings: Fusion Versus Physical Therapy	31
Applicability: Fusion Versus Physical Therapy	32
Summary: Fusion Versus Physical Therapy	32
Key Question 2: Spinal Fusion Compared to Continued Noninvasive Treatment for Painful Degenerative Lumbar Spinal Stenosis	32
Key Question 3: Spinal Fusion Compared to Continued Noninvasive Treatment for Painful Lumbar Degenerative Spondylolisthesis	32
Key Point: Fusion Versus Noninvasive Treatment	32
Detailed Synthesis	32
Strength-of-Evidence Ratings: Fusion Versus Noninvasive Treatment	36
Applicability	37
Summary: Fusion Versus Noninvasive Treatment.....	37
Key Question 4: Spinal Fusion Compared to Other Invasive Procedures for Painful Lumbar Degenerated Disc(s)	38
Key Points: Fusion Versus Arthroplasty.....	38
Detailed Synthesis	38
Strength-of-Evidence Ratings: Fusion Versus Arthroplasty	42
Applicability	43
Summary: Fusion Versus Arthroplasty.....	43
Key Question 5: Spinal Fusion Compared to Other Invasive Procedures for Painful Degenerative Lumbar Spinal Stenosis	43
Key Point: Fusion Versus Decompression	43
Detailed Synthesis	43
Strength-of-Evidence Ratings: Fusion Versus Decompression.....	46
Applicability: Fusion Versus Decompression	47
Summary: Fusion Versus Decompression	47
Key Question 6: Spinal Fusion Compared to Other Invasive Procedures for Painful Degenerative Lumbar Spondylolisthesis	47
Key Question 7: Spinal Fusion Approaches (e.g., Anterior, Posterior, Combined) and Techniques (e.g., Instrumentation, Graft Material) Compared to One Another for Painful Degenerated Lumbar Disc(s)	47
Key Points: Fusion With rhBMP-2 Versus Autogenous Bone Graft.....	47
Detailed Synthesis	47
Applicability: Fusion With rhBMP-2 Versus Autogenous Bone Graft	50
Summary: Fusion With rhBMP-2 Versus Autogenous Bone Graft.....	51
Key Points: Open Mini ALIF Versus Laparoscopic ALIF.....	51
Detailed Synthesis	51
Applicability: Open Mini ALIF Versus Laparoscopic ALIF	52
Summary: Open Mini ALIF Versus Laparoscopic ALIF	52

Key Points: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach	53
Detailed Synthesis: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach...	53
Applicability: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach.....	54
Summary: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach.....	54
Key Points: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion.....	54
Detailed Synthesis: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion	54
Applicability: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion.....	56
Summary: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion.....	56
Strength-of-Evidence Ratings: Spinal Fusion Approaches Compared to One Another for Painful Degenerated Lumbar Disc(s)	57
Key Question 8: Spinal Fusion Approaches (e.g., Anterior, Posterior, Combined) and Techniques (e.g., Instrumentation, Graft Material) Compared to One Another for Painful Degenerative Lumbar Spinal Stenosis	58
Key Points: Posterolateral Fusion Versus Posterolateral Fusion Plus Transforaminal Interbody Fusion.....	58
Detailed Synthesis: Posterolateral Fusion Versus Posterolateral Fusion Plus Transforaminal Interbody Fusion	58
Key Points: Autograft Versus Coralline Hydroxyapatite Versus Both.....	59
Detailed Synthesis: Autograft Versus Coralline Hydroxyapatite Versus Both.....	59
Strength-of-Evidence Ratings: Spinal fusion approaches and techniques compared to one another for painful degenerative lumbar spinal stenosis.....	61
Applicability: Spinal fusion approaches and techniques compared to one another for painful degenerative lumbar spinal stenosis	62
Summary: Spinal fusion approaches and techniques compared to one another for painful degenerative lumbar spinal stenosis.....	62
Key Question 9: Spinal Fusion Approaches and Techniques Compared to One Another for Painful Degenerative Lumbar Spondylolisthesis	62
Key Points: Instrumentation Versus No Instrumentation	62
Detailed Synthesis: Instrumentation Versus No Instrumentation	62
Key Points: Bilateral Instrumentation Versus Unilateral Instrumentation	64
Detailed Synthesis: Bilateral Instrumentation Versus Unilateral Instrumentation	65
Strength-of-Evidence Ratings: Spinal Fusion Approaches and Techniques Compared to One Another for Painful Degenerative Lumbar Spondylolisthesis.....	67
Applicability: Spinal Fusion Approaches and Techniques Compared to One Another for Painful Degenerative Lumbar Spondylolisthesis	68
Summary: Spinal Fusion Approaches and Techniques Compared to One Another for Painful Degenerative Lumbar Spondylolisthesis	68

Key Question 10: Patient Characteristics Predictive of Outcomes After Spinal Fusion	68
Key Points	68
Detailed Synthesis	69
Summary: Patient Characteristics Predictive of Outcomes After Spinal Fusion	72
Discussion	73
Key Findings and Strength of Evidence	73
Findings in Relationship to What is Already Known	76
Applicability	77
Implications for Clinical and Policy Decisionmaking	79
Limitations of the Comparative Effectiveness Review Process	79
Limitations of the Evidence Base	80
Research Gaps	80
References	83
Tables	
Table 1. Scope of key questions: comparators, outcomes, and primary indications	9
Table 2. Risk-of-bias assessment items.....	16
Table 3. Minimally clinically important differences of primary outcomes	18
Table 4. Included studies for Key Questions 1–9 (comparative benefit and harm)	23
Table 5. Outcome measurement scales in included studies (Key Questions 1–9)	25
Table 6. Key Question 1: strength-of-evidence ratings.....	31
Table 7. Key Question 3: strength-of-evidence ratings.....	36
Table 8. Key Question 4: strength-of-evidence ratings.....	42
Table 9. Key Question 5: strength-of-evidence ratings.....	46
Table 10. Key Question 7: transperitoneal versus retroperitoneal anterior surgical approach: strength-of-evidence ratings.....	57
Table 11. Key Question 8: strength-of-evidence ratings.....	61
Table 12. Key Question 9: strength-of-evidence ratings.....	67
Table 13. Patient and treatment variables	69
Table 14. Summary of conclusion and strength-of-evidence ratings	75
Figures	
Figure 1. Analytic framework for spinal fusion for treating painful lumbar degenerated disc(s) or degenerative joint disease	11
Figure 2. Study selection for Key Questions 1–9 (comparative benefit and harm).....	22
Figure 3. Study selection for Key Question 10 (qualitative review of prognostic factors).....	26
Appendixes	
Appendix A. Search Strategy	
Appendix B. Excluded Studies	
Appendix C. Data Tables	

Executive Summary

Background

Most adults in the United States will experience low back pain during their lives.¹ A U.S. Centers for Disease Control and Prevention 2009 National Health Interview Survey reported that 28 percent of respondents had low back pain in the 3 months before the interview.² Among people with severe low back pain, about half reported disability.³ Fortunately, an estimated 80 to 90 percent of people with acute low back pain experience complete resolution within 6 weeks, and only 5 to 10 percent develop chronic pain, although recurrences may occur.¹ In about 85 percent of patients, the cause of low back pain is never identified.⁴

Degeneration of discs and bones in the low back (i.e., lumbar spine) can cause chronic low back pain. Although degenerative lumbar conditions do not always cause symptoms, they can cause severe chronic low back pain due to vertebral instability and abnormal biomechanics and/or compression of other anatomical sites including nerves. Such conditions may occur in isolation or in combination and include the following: disc degeneration (degeneration of intervertebral discs), stenosis (narrowing of the spinal canal to less than 10 mm in diameter), and spondylolisthesis (change in vertebrae position relative to other vertebrae). This report addresses only degenerative causes of stenosis and spondylolisthesis (e.g., degenerative spondylolisthesis versus isthmic spondylolisthesis).

Spinal fusion (also known as spinal arthrodesis) is an inpatient surgical procedure intended to permanently immobilize the functional spinal unit (2 adjacent vertebrae and the disc between them), limiting painful movement.⁴ Lumbar fusion is the fusion of two vertebrae in the lumbar spine (low back). Most fusions performed today use instrumentation such as pedicle and facet screws, rods, and cages⁵ and incorporate a combination of graft materials made from a patient's own bone (autograft), donor bone (allograft), or a synthetic substance such as recombinant human bone morphogenetic protein to promote fusion. Surgeons can initiate the procedure through the peritoneum (membrane lining the abdomen) or retroperitoneum (anterior approach), the back (posterior approach), or a combination of sites (anteroposterior approach). Surgical techniques include the following:⁶

- Posterolateral fusion (PLF): Dorsal surgery (from the back of the body) that joins vertebrae by their transverse processes
- Posterior lumbar interbody fusion (PLIF): Dorsal surgery that joins vertebrae by their bodies
- Transforaminal lumbar interbody fusion: A form of PLIF that joins vertebrae on one side only
- Anterior lumbar interbody fusion (ALIF): Anterior surgery (from the front of the body) that can be performed by open transperitoneal (through the peritoneum) or, more commonly, retroperitoneal (from behind the abdominal cavity), mini-open, or laparoscopic techniques
- Circumferential fusion: 360° fusion that joins vertebrae by their entire bodies and transverse processes, typically performed by combining PLF and ALIF

Lumbar fusion's main potential advantage is to provide pain relief and restore quality of life and function when less extensive and invasive treatments cannot. However, it poses potential harms ranging from anesthesia risks and surgical complications to a need for subsequent reoperation for later complications.⁷

Objectives

To assess whether previous research is sufficient to support evidence-based conclusions about the benefit and harm of lumbar fusion relative to nonsurgical treatments or other invasive treatments or to support conclusions about outcomes following the use of different fusion strategies. We addressed these objectives by thoroughly summarizing the evidence pertaining to 10 key questions (listed below and presented graphically in Figure A). The questions and figure follow the PICOTS framework by incorporating the patient population (P), intervention (I), comparator (C), outcomes (O), timeframe (T), and setting (S). In this report, we considered all followup timepoints and settings.

Key Question 1

For adults with pain associated with degenerated disc(s) of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from continued noninvasive treatment in:

- a. Patient-centered outcomes such as function, quality of life, or pain?
- b. Adverse events?

Key Question 2

For adults with pain associated with degenerative (not congenital) stenosis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from continued noninvasive treatment in:

- a. Patient-centered outcomes such as function, quality of life, or pain?
- b. Adverse events?

Key Question 3

For adults with pain associated with degenerative spondylolisthesis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from continued noninvasive treatment in:

- a. Patient-centered outcomes such as function, quality of life, or pain?
- b. Adverse events?

Key Question 4

For adults with pain associated with degenerated disc(s) of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from other invasive procedures (e.g., total disc replacement, disc decompression) in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 5

For adults with pain associated with degenerative stenosis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from other invasive procedures (e.g., decompressive laminectomy and minimally invasive procedures, including those using devices) in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 6

For adults with pain associated with degenerative spondylolisthesis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from other invasive procedures (e.g., repair, vertebrectomy) in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 7

For adults with pain associated with degenerated disc(s) of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation, graft material) performed alone or in conjunction with additional surgery differ in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 8

For adults with pain associated with degenerative stenosis of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation, graft material) performed alone or in conjunction with additional surgery differ in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 9

For adults with pain associated with degenerative spondylolisthesis of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation, graft material) performed alone or in conjunction with additional surgery differ in:

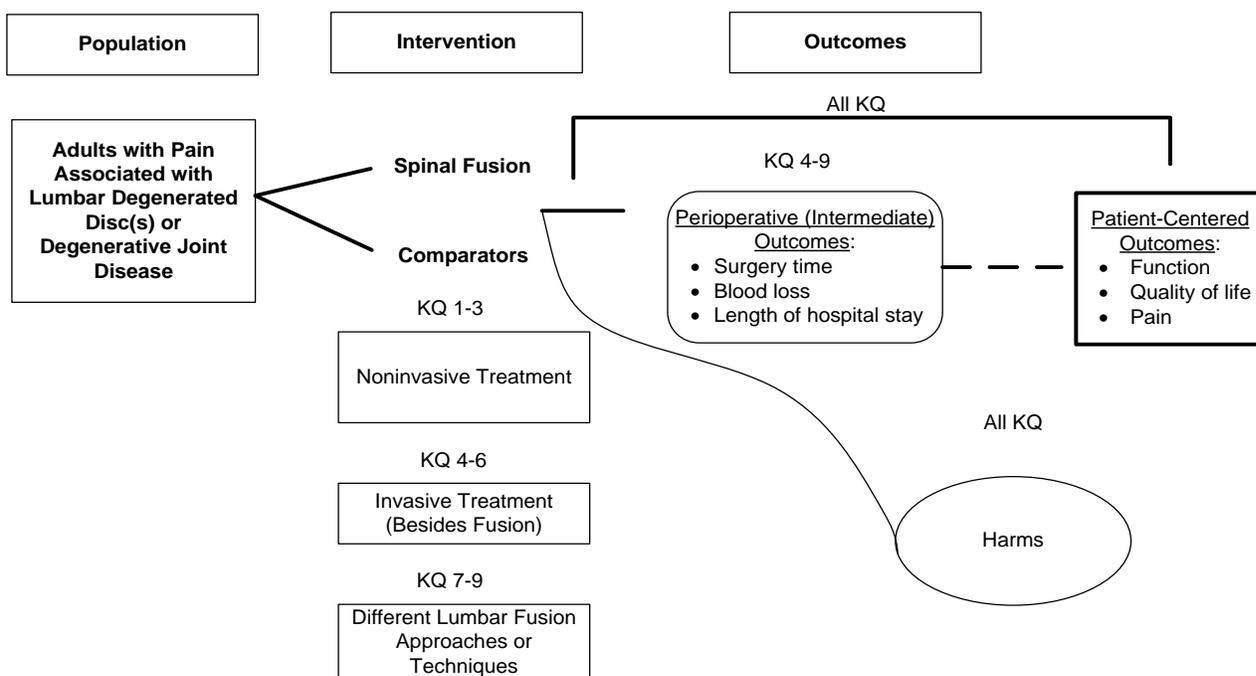
- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 10

Which, if any, patient characteristics (e.g., pain severity, type or duration of prior treatment) does the evidence show are associated with better or worse outcomes after spinal fusion in adults with degenerative disc changes, degenerative stenosis, or degenerative spondylolisthesis?

- a. Patient-centered outcomes such as function, quality of life, or pain
- b. Adverse events

Figure A. Analytic framework



KQ = Key question

Methods

We developed and refined the topic in late 2011 in collaboration with the Agency for Healthcare Research and Quality (AHRQ) and eight key informants: one neurosurgeon, one individual from a payer organization, two industry representatives, one hospital purchasing representative, one president of a patient advocacy foundation (who is also an orthopedic surgeon), and two patients who had previously undergone the surgical procedure. We finalized the review protocol in the winter of 2012 based on input from nine technical experts: two physicians (one internist and one family practitioner), two surgeons (one orthopedic surgeon and one neurosurgeon), one biostatistician, one chiropractor, one physical therapist, and two individuals from payer organizations.

Information professionals in the Evidence-based Practice Center Information Center performed literature searches and followed established guidelines and procedures as identified by the Director of Health Technology Assessment/Evidence-based Practice Center Information Center. We searched MEDLINE and PreMEDLINE; EMBASE; the Cochrane Library, including the Central Register of Controlled Trials, the Cochrane Database of Methodology Reviews, the Cochrane Database of Systematic Reviews, the Database of Abstracts of Reviews of Effects, and the Health Technology Assessment Database; and the United Kingdom National Health Service Economic Evaluation Database. The search dates were January 1, 1995 to February 7, 2012. Other mechanisms used to retrieve additional relevant information included a review of bibliographies/reference lists from peer-reviewed and gray literature. Gray literature includes reports, studies, articles, and monographs produced by federal and local government agencies, private organizations, educational facilities, consulting firms, and corporations that typically do not appear in the peer-reviewed journal literature.

For inclusion in the review, we selected only full articles published in English. For questions comparing interventions (i.e., all key questions except Key Question 10 on identifying prognostic factors), the study must have either randomly assigned patients to treatments or used an analytic method to address selection bias, such as intentional baseline matching on multiple characteristics, propensity scoring, or other analytic approaches. The study and its data collection must have been prospective. Further, the treatments being compared must have been administered during the same time period, so any observed difference between treatment outcomes were not attributable to differential time frames. For Key Question 10, cohort studies (prospective or retrospective) were also included, provided that the primary objective of the study was to evaluate potential patient-level prognostic factors.

For a study to be included for a given key question, at least 85 percent of its patients must have had the condition specified in the key question unless data is reported separately for the subgroup population. The risk of doing a mixed analysis that may include fewer than 85 percent of one condition (e.g. disc degeneration) and greater than 15 percent or more of patients with a different condition (e.g. stenosis) is that these conditions may not respond equally to fusion surgery, and therefore could skew any observed difference in effectiveness between treatments. Additionally, at least 85 percent of patients must have been undergoing primary, not revision, fusion. Studies that evaluated fusion methods no longer in use in the United States or not commercially available in the United States were not reviewed. Finally, the study must have reported data on at least one of the included outcomes for at least one of the key questions; outcome data must not have relied on retrospective recall; and data must have reported on at least 10 patients with the condition of interest who represented at least 50 percent of enrolled patients.

For each key question, we provided a description of the included studies, including the following: patient indications; method(s) of diagnosis; inclusion and exclusion criteria, including diagnostic criteria for degenerated discs or degenerative joint disease; prior, concurrent, and subsequent treatments; ancillary treatments; surgical techniques and devices used; and all reported baseline data. We also extracted the numerical data needed to compute an effect size (such as an odds ratio or standardized mean difference) and measure of variance for all included outcomes for each study. To limit clinical heterogeneity, different treatment comparisons were addressed separately within the key questions.

We assessed the risk of bias (i.e., internal validity) separately for each outcome and each time point of each study using 13 risk-of-bias items (e.g., randomization, concealment of allocation, blinding of outcome assessors, use of methods to enhance group comparability in nonrandomized studies, whether the comparisons of interest were prospectively planned, whether the outcomes of interest were measured objectively). Based on these items, each data point from each study was assigned a risk-of-bias category of “Low,” “Moderate,” or “High.” This was performed in duplicate, with disagreements resolved by consensus.

Within each treatment comparison, we examined all included outcomes from all relevant studies. The outcomes were divided into two categories: perioperative (Key Questions 4–9) and patient-centered. Perioperative outcomes included surgery time, blood loss, and length of hospital stay and patient-centered outcomes included function, quality of life, and pain. We also examined adverse events, which might include reoperation, neurological injury, blood clots, and infection.

We planned to perform meta-analysis whenever the evidence base for a key question met the following minimum criteria: it consisted of at least two studies addressing the same outcome at the same duration of followup and the studies were clinically similar in terms of patient

characteristics, surgical approach and strategy, and comparability of control groups. However, none of the evidence bases met the minimum criteria, so we did not attempt to use meta-analysis to determine summary effect sizes. We therefore performed qualitative analyses in which the studies comprising the evidence base for each key question were described, compared, and contrasted. Similarly, we planned to statistically investigate patient and treatment factors (e.g., presence or absence of radicular pain, prior treatments, and preoperative pain severity) for association with patient outcomes in Key Question 10. However, there were insufficient data to permit such analyses, so we conducted a qualitative review of primary literature reporting patient-level data.

To aid interpretation, for each outcome in the review we set the smallest difference between groups that could still be considered clinically significant (minimum clinically significant difference, MCID). This MCID aids interpretation by determining whether a statistically significant difference is important or whether a statistically nonsignificant difference is small enough to exclude the possibility of an important difference. For instance, a change of <8.2 points on the Oswestry Disability Index (scale 0 to 100 points) is unlikely to be considered a clinically significant change, whereas a change of 20 or more points is very likely to be considered clinically significant. For instruments measuring pain or function for which no literature-based MCID could be identified, we considered a 30 percent difference to be clinically significant. This number is based on a study by Raymond et al. in which the objective was to determine meaningful changes for back pain and function using different methodologies.⁸ These authors looked at several commonly used instruments to measure pain and function and determined that a 30 percent change from baseline may be considered a clinically meaningful improvement when comparing pre to post-treatment scores.

For major comparisons and outcomes, we rated the strength of evidence using the Evidence-based Practice Center system described by Owens and colleagues.⁹ This system includes four core domains (risk of bias, consistency, precision, and directness) as well as four optional domains (large magnitude of effect, all plausible confounders that would reduce the effect, publication bias, and dose-response association). The directness domain does not encompass applicability, which is considered outside the evidence rating system. The various domains were considered together using transparent rules to rate the evidence for the outcome as “High,” “Moderate,” “Low,” or “Insufficient.” We performed strength-of-evidence rating for all key questions except Key Question 10, which did not involve comparing treatments, but rather an assessment of potential patient-level prognostic factors.

Applicability was assessed by considering important patient characteristics (e.g., diagnosis, presurgical pain level, presurgical functional status, workers compensation or other occupational factors, prior surgery, other patient characteristics) and treatment characteristics (e.g., surgical approach, device or other materials used in surgery, adjunctive surgery). Based on a review of the data abstracted, we narratively summarized any patterns reflected from these factors that might affect the applicability of the evidence to the general population and to the Medicare-beneficiary population. We made no attempt to generate any rating or score for the applicability of the evidence. Our narrative summaries are intended to draw stakeholders’ attention to potential applicability issues embedded in the evidence.

Results

Extensive literature searches identified 4,378 citations potentially addressing the comparative benefits and harms of lumbar fusion (Key Questions 1–9). Of those, 4,230 were

excluded upon title and abstract review for clearly not meeting inclusion criteria (e.g., not pertinent to the topic, uncontrolled or historically controlled, not a clinical study). Of the 148 articles retrieved and reviewed in full for the review of benefits and harms, 128 were excluded. Reasons for exclusion are summarized in Figure B below. Eighteen clinical studies (reported in 20 publications) remained for inclusion, of which four addressed Key Question 1, one addressed Key Question 3, two addressed Key Question 4, one addressed Key Question 5, six addressed Key Question 7, two addressed Key Question 8, and three addressed Key Question 9. One study addressed more than one key question and no studies addressed Key Questions 2 or 6.

We conducted additional searches for Key Question 10 to identify studies in which the primary objective was to examine patient and/or treatment factors associated with patient outcomes following fusion surgery. Our searches identified 1,452 potentially relevant studies. Of those, 1,383 were excluded upon title and abstract review for clearly not meeting inclusion criteria (e.g., not relevant to the condition or treatment, not assessing prognostic factors). Of the 69 articles retrieved and reviewed in full, 61 were excluded and 7 clinical studies (8 publications) remained for inclusion. Reasons for exclusion are summarized below in Figure C.

Figure B. Study selection for Key Questions 1–9 (comparative benefits and harms)

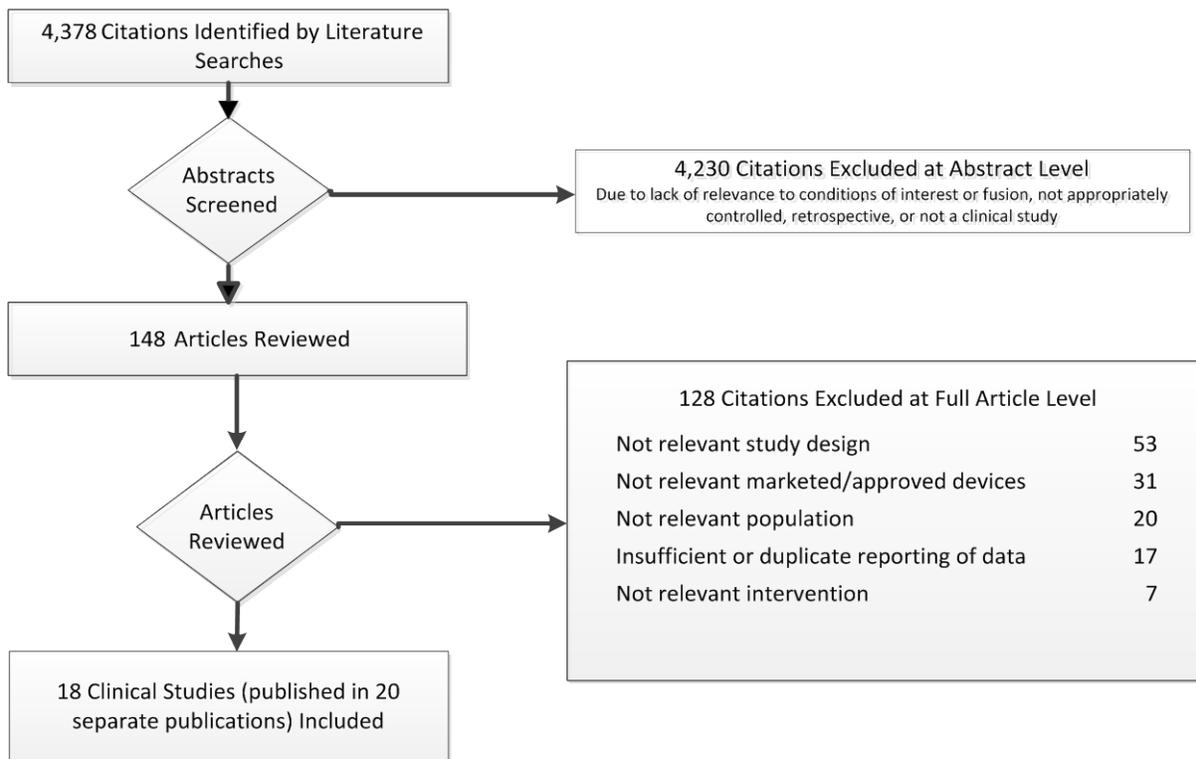
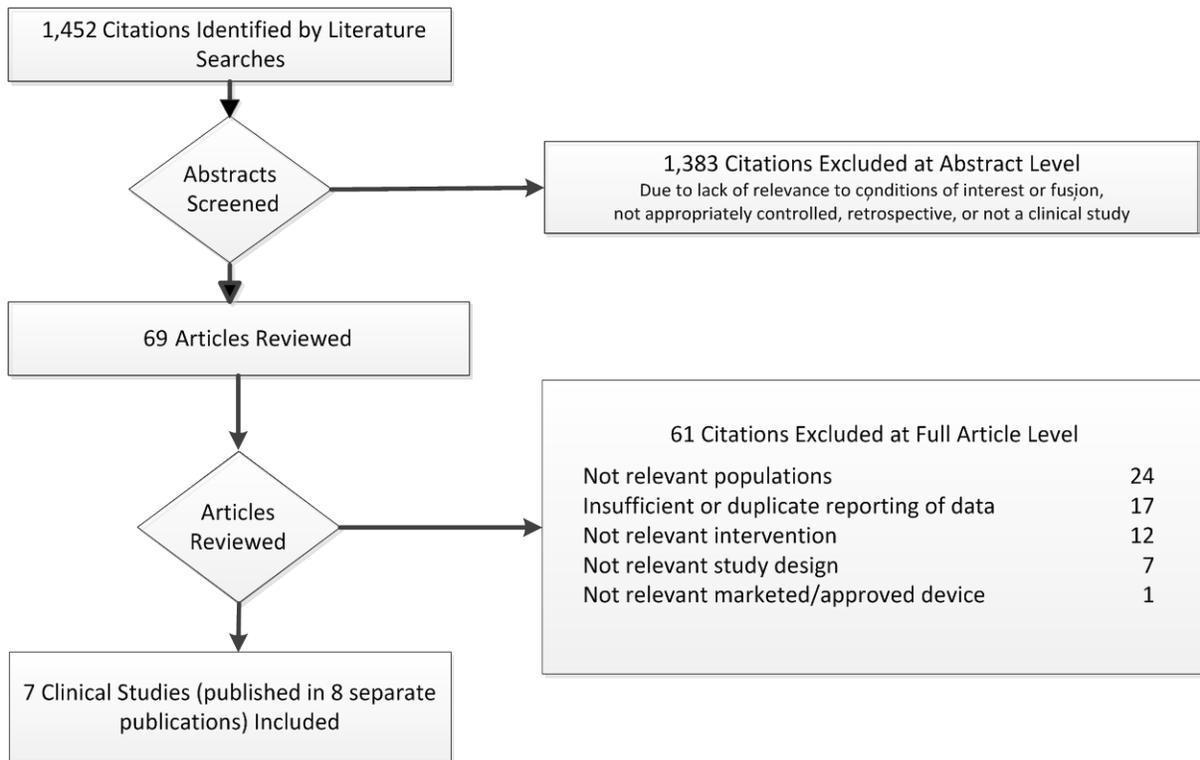


Figure C. Study selection for Key Question 10 (qualitative review of prognostic factors)



Most studies addressing Key Questions 1-9 were RCTs and generally well-designed. All but one study earned a moderate risk-of-bias rating. The moderate rating was largely because of lack of concealment of allocation and/or blinding of patients or outcome assessors to treatment received, or not reporting if concealment or blinding took place in the study. While blinding patients and providers would have been impossible in most of the studies, knowledge of the treatment received and related expectations present a potential source of bias nevertheless. One study earned a high risk-of-bias rating due to high treatment crossover and other limitations that compromised the randomization of the study and introduced selection and other biases.

Below, we summarize the key findings for each question. Details about the patients enrolled in each of the studies and the treatments they received are presented more fully in the full comparative effectiveness report.

Key Question 1: For adults with pain associated with degenerated disc(s) of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from continued noninvasive treatment?

The evidence base for this question consisted of four randomized controlled studies (RCTs), all of which compared fusion to physical therapy.¹⁰⁻¹³ Because of differences in the treatments (the conditions of physical therapy varied considerably across the studies) and patient populations across the studies, no meta-analysis was performed. The following conclusions are based on a qualitative assessment of the evidence.

- Limited data suggests that fusion leads to greater improvement in back pain relief and function than physical therapy at 2-year followup (strength of evidence: low); however, whether the difference is clinically significant is unclear (the confidence intervals overlap

with what is considered a clinically significant difference), and findings at 1 year are insufficient to allow conclusions.

- No other conclusions are possible, because of insufficient evidence or uninformative statistical findings. Serious and minor adverse events occurred in the fusion group that could not occur in a noninvasive-intervention group; however, because of insufficient reporting and variation in surgical methods used in the different studies, the incidence of these events cannot be determined conclusively.

Key Question 2: For adults with pain associated with degenerative (not congenital) stenosis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from continued noninvasive treatment?

- We identified no studies that met inclusion criteria and addressed this key question.

Key Question 3: For adults with pain associated with degenerative spondylolisthesis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from continued noninvasive treatment?

The evidence base for this question consisted of one study reported in two separate publications (one reporting followup data for up to 2 years¹⁴ and the other reporting 4-year followup data¹⁵).

- Because only one study compared fusion versus noninvasive treatment, the evidence is insufficient to support evidence-based conclusions regarding the comparative effectiveness or safety of fusion in adults with low back pain due to degenerative spondylolisthesis.

Key Question 4: For adults with pain associated with degenerated disc(s) of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from other invasive procedures (e.g., total disc replacement, disc decompression)?

The evidence base for this question consisted of two multicenter RCTs.^{16,17} In both studies the surgical approach was anterior for arthroplasty, and anterior and posterior for fusion.

- **Fusion versus arthroplasty:** Limited evidence suggests that shorter surgical time, less blood loss, and shorter inpatient stays are associated with arthroplasty and that disc recipients have better Oswestry Disability Index (ODI) functions scores at 6 weeks postsurgery (strength of evidence: low). The difference in ODI functions were not observed at later followup times.
- For all other outcomes, the data were insufficient to support any conclusions, typically because the mean difference was uninformative (i.e., captured the possibility of being either superior or equivalent) or because only one study addressed that outcome or duration of followup.

Key Question 5: For adults with pain associated with degenerative stenosis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from other invasive procedures (e.g., decompressive laminectomy and minimally invasive procedures, including those using devices)?

One study comparing single or bilateral foraminotomy with nerve root decompression to posterolateral instrumented pedicular fusion made up the evidence base for this key question.¹⁸

- **Fusion versus decompression:** Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the comparative

effectiveness or safety of fusion and decompression in adults with low back pain due to stenosis with degenerative disc.

Key Question 6: For adults with pain associated with degenerative spondylolisthesis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from other invasive procedures (e.g., repair, vertebrectomy)?

- We identified no studies that met inclusion criteria and address this key question.

Key Question 7: For adults with pain associated with degenerated disc(s) of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation, graft material) performed alone or in conjunction with additional surgery differ?

The evidence base for this question consisted of six RCTs: three compared fusion with rhBMP-2 versus autogenous bone graft,¹⁹⁻²¹ one compared mini ALIF to laparoscopic ALIF,²² one compared posterolateral fusion with or without variable screw placement with circumferential fusion,²³ and one compared transperitoneal to retroperitoneal anterior surgical approach.²⁴

- **Fusion with rhBMP-2 versus autogenous bone graft:** rhBMP-2 is associated with less blood loss than autogenous bone graft, while surgery time and length of hospital stay do not differ substantially for these products in adults undergoing fusion for low back pain due to degenerated disc (strength of evidence: low). For all other outcomes the data were insufficient to support any conclusions because of inconsistencies in the study's findings, insufficient reporting of data, or because only one study addressed that outcome or duration of followup.
- **Open mini ALIF versus laparoscopic ALIF:** Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the relative perioperative outcomes or safety of open mini or laparoscopic surgery in patients undergoing fusion for disc degeneration.
- **Posterolateral fusion, with or without variable screw placement, or circumferential fusion:** Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the efficacy or safety of posterolateral fusion, with or without variable screw placement, or circumferential fusion.
- **Transperitoneal versus retroperitoneal anterior surgical approach:** Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the comparative efficacy or safety of transperitoneal versus retroperitoneal anterior surgical approach.

Key Question 8: For adults with pain associated with degenerative stenosis of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation, graft material) performed alone or in conjunction with additional surgery differ?

The evidence base for this question included two studies: one comparing posterolateral fusion with posterolateral fusion plus transforaminal interbody fusion¹⁸ and one comparing fusion with autograft with fusion using coralline hydroxyapatite.²⁵

- **Posterolateral fusion versus posterolateral fusion plus transforaminal interbody fusion:** Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the comparative effectiveness of posterolateral fusion versus posterolateral fusion plus transforaminal interbody fusion.

- **Autograft versus coralline hydroxyapatite versus both:** Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the comparative effectiveness of autograft versus coralline hydroxyapatite versus both.

Key Question 9: For adults with pain associated with degenerative spondylolisthesis of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation, graft material) performed alone or in conjunction with additional surgery differ?

The evidence base for this question included three studies: two comparing fusion with instrumentation with fusion with no instrumentation^{26,27} and one comparing fusion with bilateral instrumentation with fusion using unilateral instrumentation.²⁸

- **Instrumentation versus no instrumentation:** Two studies addressed this comparison. The evidence is insufficient to support evidence-based conclusions regarding the comparative effectiveness and safety of instrumentation versus no instrumentation mainly due to dissimilarities in the outcomes reported in the studies.
- **Bilateral instrumentation versus unilateral instrumentation:** Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the comparative effectiveness and safety of bilateral instrumentation versus unilateral instrumentation.

Key Question 10: Which, if any, patient characteristics (e.g., pain severity, type or duration of prior treatment) does the evidence show are associated with better or worse outcomes after spinal fusion in adults with degenerative disc changes, degenerative stenosis, or degenerative spondylolisthesis?

The evidence base for this question consisted of seven studies reported in eight separate publications (two studies were companion studies with one reporting data for up to 2 years followup¹⁴ and one reporting 4-year followup data.¹⁵ Of the studies, one was an RCT²⁹, one was a nonrandomized comparative trial,^{14,15} and the other five were cohort studies (3 retrospective studies and 2 prospective studies).³⁰⁻³⁴ The studies examined a number of factors that could potentially affect patient outcomes following spinal fusion surgery. The most commonly assessed factors include age, gender, and workers compensation status.

- Older age (65 years or older) appears to be associated with worse patient outcomes following spinal fusion. Three of the six studies considered for this question showed a statistically significant association between older age (≥ 65 years) and poor patient outcomes.

Because of insufficient reporting and variation in surgical methods used in the different studies, the incidence of adverse events (serious and minor) associated with fusion could not be determined conclusively.

Summary of Conclusions and Strength of Evidence

The Table A below summarizes our conclusions and strength-of-evidence ratings for key questions for which there were sufficient number of studies (more than one) to potentially draw evidence-based conclusions.

Table A. Summary of conclusions and strength of evidence

Comparison	Outcome	Time	Number of Studies (Total N)	Risk of Bias	Consistency	Directness	Precision ¹	Evidence Favors	Strength of Evidence Rating
Fusion vs. physical and exercise therapies (Key Question 1)	Pain, back, VAS	1 YR	3 (N=153)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		2 YR	2 (N=299)	Moderate	Consistent	Direct	Precise	Fusion	Low
	Pain, leg, VAS	1 YR	2 (N=118)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Pain, drugs	1 YR	2 (N=118)	Moderate	Insufficient data	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	1 YR	3 (N=153)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		2 YR	2 (N=299)	Moderate	Consistent	Direct	Precise	Fusion	Low
	Function, GFS	1 YR	2 (N=118)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
Function, return to work	1 YR	2 (N=118)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient	
Fusion vs. artificial intervertebral disc (Key Question 4)	Surgical time	Perioperative	2 (N=472)	Moderate	Consistent	Direct	Precise	Disc	Low
	Blood loss	Perioperative	2 (N=470)	Moderate	Consistent	Direct	Precise	Disc	Low
	Inpatient stay	Perioperative	2 (N=473)	Moderate	Consistent	Direct	Precise	Disc	Low
	Pain, VAS	6 WK–2 YR	2 (N=465)	Moderate	Consistent	Direct	Imprecise	Inconclusive	Insufficient
	Pain, drugs	2 YR	2 (N=469)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	6 WK	2 (N=467)	Moderate	Consistent	Direct	Precise	Disc	Low
3 MO–2 YR		2 (N=467)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient	
rhBMP-2 vs. autogenous bone graft (Key Question 7)	Surgical time	Perioperative	3 (N=371)	Moderate	Consistent	Direct	Unknown	No substantial difference	Low
	Blood loss	Perioperative	3 (N=371)	Moderate	Consistent	Direct	Unknown	rhBMP-2	Low
	Inpatient stay	Perioperative	3 (N=371)	Moderate	Consistent	Direct	Unknown	No substantial difference	Low
	Back pain, analog	2 YR	2 (N=271)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Leg pain, analog	2 YR	2 (N=271)	Moderate	Consistent	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	1 YR	2 (N=298)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	2 YR	2 (N=271)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient

Comparison	Outcome	Time	Number of Studies (Total N)	Risk of Bias	Consistency	Directness	Precision ¹	Evidence Favors	Strength of Evidence Rating
	Function, return to work	2 YR	3 (N=316)	Moderate	Consistent	Direct	Imprecise	Inconclusive	Insufficient

¹ We considered a study to have an imprecise outcome when the intervention that was favored could not be determined. GFS = General Function Scale; ODI = Oswestry Disability Index; VAS = visual analog scale

Applicability

For our assessment of applicability, we narratively summarized any patterns reflected in patient or treatment characteristics (including comparators, outcomes, and setting) that might affect the applicability of the evidence to the general population and to the Medicare-beneficiary population. Our review of the comparative studies that made up the evidence base for Key Questions 1 to 9, indicated that overall the studies enrolled mostly middle aged patients between 40 to 45 years old. In only four studies was the average age of patients greater than 50 years. Women made up less than 50 percent of the patient population in half of the included studies, and race and ethnicity were poorly reported across the studies. Duration of symptoms varied widely across studies, with one study reporting that the enrolled patients were symptomatic for less than six months and another study reporting that patients were symptomatic for more than seven years. Overall, comorbidities and other health related factors were poorly reported.

Although devices in this report have been evaluated for inclusion by members of the technical expert panel, many of the included studies were conducted over a decade ago, which may limit their applicability to currently used fusion instrumentation (e.g., rods, pedicle and facet screws) and techniques. Further, in the included studies that compared fusion to nonsurgical alternatives, the nonsurgical comparator treatments varied across studies in terms of the duration and intensity of the physical therapy component and in the supplemental treatments (e.g., acupuncture, injections, advice, and/or cognitive therapy) received by many patients. In one study, patients in the nonsurgical comparator group received a range of therapies that at minimum included active physical therapy, education/counseling with home exercise instruction, and a non-steroidal anti-inflammatory drug.

The above study, however, was designed to be a pragmatic trial. As such, the heterogeneity of treatments in the non-surgical group is probably at least somewhat representative of the range of treatments that patients would likely receive in clinical practice. Thus, the findings of the study may be more representative of what is likely to occur in clinical practice. However, such heterogeneity makes replication of the findings of pragmatic studies difficult because the exact mix of the alternative or comparator treatment typically varies among patients within the studies and is likely to vary from one clinical practice to another.

The majority of studies captured important patient outcomes, such as pain, function, and quality of life, using recognized, validated instruments. However, in a couple of studies outcome reporting was restricted to peri-operative outcomes and adverse events. Finally, nine (56%) studies were conducted outside the U.S., and many of the studies were conducted in spine specialty centers and orthopaedic centers. Such factors are likely to limit the applicability of the findings of these studies for both U.S. Medicare beneficiaries and primary care populations.

Research Gaps

Through our review of the evidence, we identified a number of gaps in the literature that need further research. In particular, research is needed on the benefits and harms of fusion for individuals over 60 years of age. The number of fusion surgeries in this population is growing despite a lack of evidence that surgery is more beneficial than other noninvasive treatments for individuals over 60. In only a few of the studies included in this review was the mean age of the patients over 50 years. Further, our qualitative review of studies evaluating prognostic factors associated with patient outcomes following surgery suggest that older age (>65 years) is associated with poorer patient outcomes.

In general, more studies are needed that focus on identifying patients who are more or less likely to benefit from fusion. Our searches identified only a handful of prognostic studies, and the patient and treatment characteristics evaluated in those studies varied. Further, patient characteristics, particularly patient comorbidities and other health related factors were poorly reported in many of the comparative trials included in the review.

Poor reporting along with variation in the surgical methods used among the comparative trials that addressed key questions 1 to 9 limited our ability to conclusively determine the incidence of adverse events associated with fusion. Thus, more complete reporting of all adverse events (serious and minor) associated with fusion and its comparators is needed in future research. Further, sufficient followup to capture late adverse events is also needed in studies comparing fusion to other invasive procedures.

One overarching problem with the evidence base in this report is the variation in the therapies used in the noninvasive comparator group among the studies that compared fusion to noninvasive alternatives. For instance, in the studies that compared fusion to physical therapy, the physical therapy component varied considerably in terms of intensity, duration, and use of supplemental therapies such as acupuncture, cognitive behavioral therapy, or injections. Efforts to develop a standardized approach to defining and delivering physical therapy for these conditions would make replication of comparisons between fusion and physical therapy possible. Similarly, clearly describing what patients received in all treatment groups is important to replication of comparisons of fusion to other treatments. In at least one study that made up the evidence base for this report, the specific surgical tools and techniques used in the fusion group were not clearly described and the therapies provided in the nonsurgical group varied and were not fully described.

Overall, more studies are needed that compare fusion to other noninvasive therapies, such as exercise therapy or cognitively-oriented therapies. More studies that compare spinal fusion and noninvasive treatment(s) for patients with degenerative stenosis are also needed. Our searches did not identify any studies that met the inclusion criteria for this report that compared spinal fusion and continued noninvasive treatment for patients with degenerative stenosis. Ideally, future studies would compare certain types of fusion for certain indications to non-operative care.

Our searches did not identify any studies that met inclusion criteria that compared spinal fusion to other invasive procedures for patients with spondylolisthesis. Thus, more studies are needed in this area. Finally, more studies that compare different fusion methods and techniques are needed to clarify optimal surgical procedures. Because implantable devices are frequently replaced by new products and product generations, either by product line updates or withdrawal of previous implants and instrumentation from the market because of adverse events, ongoing clinical studies of new devices and materials are needed. Many of the studies retrieved for our review were ultimately excluded for lack of relevance to modern treatment practices in the United States.

Lastly, for most comparisons considered in this review only one or two studies were identified. However, to support an evidence-based conclusion, replication of findings is generally needed. Replication of comparisons in clinically comparable populations across multiple studies also enables meta-analysis, increases the power of the evidence base to detect a difference between treatments overall, and decreases the likelihood that the overall findings will be imprecise. Future studies ideally would perform randomized assignment of patients to treatment

arms and, while blinding of patients and practitioners is not always practical, outcome assessors would be blinded to treatment assignment if possible.

Conclusions

Limited evidence suggests that spinal fusion compared with physical therapy improves pain and function for adults undergoing fusion for low back pain due to disc degeneration. Because of insufficient reporting and variation in surgical methods used in the different studies, the incidence of adverse events (serious and minor) associated with fusion could not be determined conclusively. More complete reporting of all adverse events associated with fusion is needed in future research. The evidence was insufficient to draw evidence-based conclusions for the benefits and harms of spinal fusion for patients with degenerative stenosis or degenerative spondylolisthesis of the lumbar spine. The evidence was also largely insufficient to draw conclusions about the benefits and harms of fusion compared with other invasive treatments or different fusion approaches or techniques. Thus, future research is needed in these areas.

Executive Summary Reference List

1. Manchikanti L. Epidemiology of low back pain. *Pain Physician* 2000;3(2):167-92.
2. Pleis JR, Ward BW, Lucas JW. Summary health statistics for U.S. adults: National Health Interview Survey, 2009. *Vital Health Stat* 10 2010 Dec;(249):1-207. PMID: 21905346
3. National Center for Health Statistics. Health, United States, 2010: with special feature on death and dying. Hyattsville (MD): National Center for Health Statistics; 2011. 547 p.
4. Hu SS, Tribus CB, Tay BK, et al. Chapter 5: Disorders, diseases, and injuries of the spine. Osteomyelitis of the spine. In: Skinner HB, editor. *CURRENT diagnosis and treatment in orthopedics*. 4th ed. New York: McGraw-Hill; 2006. Also available: <http://www.accessmedicine.com/content.aspx?aID=2319334>.
5. Maghout Juratli S, Franklin GM, Mirza SK, et al. Lumbar fusion outcomes in Washington State workers' compensation. *Spine (Phila Pa 1976)* 2006 Nov 1;31(23):2715-23. PMID: 17077741
6. Resnick DK. Spinal fusion for discogenic back pain: patient selection, operative techniques, and outcomes. *Tech Neurosurg* 2003;8(3):176-190.
7. McCrory DC, Turner DA, Patwardhan MB, et al. Spinal fusion for treatment of degenerative disease affecting the lumbar spine [draft]. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2006 Nov 1. 93 p.
8. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 2008 Jan 1;33(1):90-4. PMID: 18165753
9. Owens D, Lohr K, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions. In: Agency for Healthcare Research and Quality (AHRQ). *Methods guide for comparative effectiveness reviews*. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2009 Jul. Also available: http://effectivehealthcare.ahrq.gov/ehc/products/122/328/2009_0805_grading.pdf.
10. Ivar Brox J, Sorensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine (Phila Pa 1976)* 2003 Sep 1;28(17):1913-21. PMID: 12973134
11. Brox JI, Reikeras O, Nygaard O, et al. Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: a prospective randomized controlled study. *Pain* 2006 May;122(1-2):145-55. PMID: 16545523
12. Fritzell P, Hagg O, Wessberg P, et al. 2001 Volvo Award Winner in Clinical Studies: lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine (Phila Pa 1976)* 2001 Dec 1;26(23):2521-34. PMID: 11725230
13. Ohtori S, Koshi T, Yamashita M, et al. Surgical versus nonsurgical treatment of selected patients with discogenic low back pain: a small-sized randomized trial. *Spine (Phila Pa 1976)* 2011 Mar 1;36(5):347-54.
14. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med* 2007 May 31;356(22):2257-70. PMID: 17538085

15. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *J Bone Joint Surg Am* 2009 Jun;91(6):1295-304. PMID: 19487505
16. Zigler J, Delamarter R, Spivak JM, et al. Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine (Phila Pa 1976)* 2007 May 15;32(11):1155-62; discussion 1163. PMID: 17495770
17. Delamarter R, Zigler JE, Balderston RA, et al. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement compared with circumferential arthrodesis for the treatment of two-level lumbar degenerative disc disease: results at twenty-four months. *J Bone Joint Surg Am* 2011 Apr 20;93(8):705-15. PMID: 21398574
18. Hallett A, Huntley JS, Gibson JN. Foraminal stenosis and single-level degenerative disc disease: a randomized controlled trial comparing decompression with decompression and instrumented fusion. *Spine (Phila Pa 1976)* 2007 Jun 1;32(13):1375-80. PMID: 17545903
19. Dawson E, Bae HW, Burkus JK, et al. Recombinant human bone morphogenetic protein-2 on an absorbable collagen sponge with an osteoconductive bulking agent in posterolateral arthrodesis with instrumentation: a prospective randomized trial. *J Bone Joint Surg Am* 2009 Jul 1;91(7):1604-13.
20. Burkus JK, Transfeldt EE, Kitchel SH, et al. Clinical and radiographic outcomes of anterior lumbar interbody fusion using recombinant human bone morphogenetic protein-2. *Spine (Phila Pa 1976)* 2002 Nov 1;27(21):2396-408. PMID: 12438990
21. Burkus JK, Gornet MF, Dickman CA, et al. Anterior lumbar interbody fusion using rhBMP-2 with tapered interbody cages. *J Spinal Disord Tech* 2002 Oct;15(5):337-49. PMID: 12394656
22. Chung SK, Lee SH, Lim SR, et al. Comparative study of laparoscopic L5-S1 fusion versus open mini-ALIF, with a minimum 2-year follow-up. *Eur Spine J* 2003 Dec;12(6):613-7. PMID: 14564558
23. Fritzell P, Hagg O, Wessberg P, et al. Chronic low back pain and fusion: a comparison of three surgical techniques: a prospective multicenter randomized study from the Swedish lumbar spine study group. *Spine (Phila Pa 1976)* 2002 Jun 1;27(11):1131-41.
24. Sasso RC, Burkus JK, LeHuec JC. Retrograde ejaculation after anterior lumbar interbody fusion: transperitoneal versus retroperitoneal exposure. *Spine (Phila Pa 1976)* 2003 May 15;28(10):1023-6. PMID: 12768143
25. Korovessis P, Koureas G, Zacharatos S, et al. Correlative radiological, self-assessment and clinical analysis of evolution in instrumented dorsal and lateral fusion for degenerative lumbar spine disease. Autograft versus coralline hydroxyapatite. *Eur Spine J* 2005 Sep;14(7):630-8. PMID: 15789231
26. Fischgrund JS, Mackay M, Herkowitz HN, et al. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal instrumentation. *Spine (Phila Pa 1976)* 1997;22(24):2807-2812.
27. Abdu WA, Lurie JD, Spratt KF, et al. Degenerative spondylolisthesis: does fusion method influence outcome? Four-year results of the spine patient outcomes research trial. *Spine (Phila Pa 1976)* 2009 Oct 1;34(21):2351-60. PMID: 19755935

28. Fernandez-Fairen M, Sala P, Ramirez H, et al. A prospective randomized study of unilateral versus bilateral instrumented posterolateral lumbar fusion in degenerative spondylolisthesis. *Spine (Phila Pa 1976)* 2007 Feb 15;32(4):395-401. PMID: 17304127
29. Hagg O, Fritzell P, Ekselius L, et al. Predictors of outcome in fusion surgery for chronic low back pain. A report from the Swedish Lumbar Spine Study. *Eur Spine J* 2003 Feb;12(1):22-33.
30. Fukuta S, Miyamoto K, Hosoe H, et al. Kidney-type intervertebral spacers should be located anteriorly in cantilever transforaminal lumbar interbody fusion: analyses of risk factors for spacer subsidence for a minimum of 2 years. *J Spinal Disord Tech* 2011 May;24(3):189-95. PMID: 20634726
31. Keorochana G, Laohacharoensombat W, Wajanavisit W, et al. Functional outcome after decompression and instrumented arthrodesis in degenerative lumbar spinal stenosis: factors influencing unsuccessful outcome change. *J Med Assoc Thai* 2011 Dec;94(12):1487-94. PMID: 22295737
32. Carreon LY, Glassman SD, Djurasovic M, et al. Are preoperative health-related quality of life scores predictive of clinical outcomes after lumbar fusion? *Spine (Phila Pa 1976)* 2009 Apr 1;34(7):725-30. PMID: 19333106
33. Okuda S, Oda T, Miyauchi A, et al. Surgical outcomes of posterior lumbar interbody fusion in elderly patients. *J Bone Joint Surg Am* 2006 Dec;88(12):2714-20.
34. Schuler TC, Burkus JK, Gornet MF, et al. The correlation between preoperative disc space height and clinical outcomes after anterior lumbar interbody fusion. *J Spinal Disord Tech* 2005 Oct;18(5):396-401.

Introduction

Degenerative Lumbar Discs or Joints

Most adults in the United States will experience low back pain during their lives.¹ A U.S. Centers for Disease Control and Prevention 2009 National Health Interview Survey reported that 28 percent of respondents had low back pain in the 3 months before the survey interview.² Among people with severe low back pain, about half reported disability.³ Fortunately, an estimated 80 to 90 percent of people with acute low back pain experience complete resolution within 6 weeks, and only 5 to 10 percent develop chronic pain, although recurrences may occur.¹ In about 85 percent of patients, the cause of low back pain is never identified.⁴

Degeneration of discs and bones in the low back (i.e., lumbar spine) can cause chronic low back pain although a high prevalence of individuals with degenerative discs are asymptomatic.³⁵ Painful degenerative lumbar conditions have been attributed to vertebral instability and abnormal biomechanics and/or compression of other anatomical sites including nerves.³⁵ However attributing pain to an abnormal disc on imaging studies is controversial.

Magnetic resonance imaging (MRI), computed tomography (CT) and lumbar discography (also known as provocative discography) have all been used to determine disk pathology. While degenerative changes of the disk (e.g., radial tears of the disk annulus or loss of disk height) can be detected by MRI or CT scanning, both modalities have been associated with significant false-positive rates³⁶ Lumbar discography, an invasive procedure, entails injecting contrast material into the nucleus pulposus of a lumbar disc to determine whether the disc itself is the source of pain (i.e., a diagnosis of discogenic pain). Discography yields two types of results: pain provocation (whether the patient's typical pain was reproduced by the injection) and morphology (whether the dye images an abnormal pattern in the disc, often based on CT scan). Controversy exists about the relative importance of these two test results.

Degenerative lumbar conditions may occur in isolation or in combination and include the following:

- **Disc degeneration:** Degeneration of intervertebral discs, thought to be genetically influenced and due to mechanical loading associated with aging or trauma.^{35,37} Disc degeneration may include “real or apparent desiccation, fibrosis, narrowing of the [disc] space, diffuse bulging of the annulus beyond the disc space, extensive fissuring (i.e., numerous annular tears) and mucinous degeneration of the annulus, defects and sclerosis of the endplates, and osteophytes at the vertebral apophyses.”³⁵
- **Stenosis:** Narrowing of the spinal canal to less than 10 mm in diameter, sometimes associated with intervertebral disc herniation or degeneration; or narrowing of nerve root canals or intervertebral foramina, which can cause nerve root compression.^{4,35,38} Degeneration can be due to infection, trauma, or surgery. This report addresses only degenerative causes of stenosis.
- **Spondylolisthesis:** Change in vertebrae position relative to other vertebrae (sometimes called a “slip”), which can be due to apophyseal joint degeneration or facet arthropathy.^{35,39} Underlying degenerative causes include arthritis of facet joints, ligament malfunction, and insufficient muscle stabilization.⁴⁰ This report addresses only degenerative causes of spondylolisthesis.

The 2009 National Health Interview Survey found the prevalence of back pain in general is higher among women, people who are poor or near poor, and people at least 44 years of age.²

Sixty-nine percent of low back pain medical visits were for adults aged 25–64 years.⁴¹ Prevalence of disc degeneration is greatest among people at least 40 years old,⁴² and prevalence of herniated discs is highest among people 35–45 years old.⁴ Spondylolisthesis most commonly occurs in people aged at least 40–50 years.^{39,40} Most people older than 60 years of age have radiographic evidence of spinal stenosis,³⁸ and most people with signs and symptoms are older than 70 years of age.⁴ General risk factors for degenerative spinal conditions include age at least 50 years, female gender, pregnancies, African heritage, joint laxity, and anatomic predisposition.⁴⁰

Patients with low back pain and a “red flag” symptom or sign require immediate evaluation and may require prompt surgery. Red flags include pain due to trauma, sudden or unexplained loss or change in bowel or bladder control or urinary retention, sudden or unexplained saddle anesthesia or bilateral leg weakness, and signs and symptoms suggestive of cancer or spinal infection.⁴³ Treatments for these patients are outside the scope of this review.

Noninvasive Treatments for Degenerative Lumbar Conditions

Although degenerative causes of low back and leg pain vary, nonsurgical treatments are generally the same. Treatments include the following (from most to least conservative): bed rest (fewer than 2 days), nonsteroidal anti-inflammatory drugs (NSAIDs) or acetaminophen, antispasmodics, opioids, cardiovascular exercise, physical therapy, complementary and alternative medicine, chiropractic care, and cognitive behavior therapy.^{4,6,38-40,44-46} External braces may also be prescribed for spondylolisthesis or stenosis.^{38,39} Below, we further describe and present evidence for several noninvasive modalities used to treat patients who have degenerative lumbar conditions.

Exercise therapy has been described as the most widely used type of conservative treatment worldwide.⁴⁷ Various therapeutic exercises exist and include aerobic, aquatic (e.g., pool rehabilitation), directional preference (e.g., McKenzie), flexibility (e.g., yoga), and strengthening.⁴⁸ Exercise programs vary in intensity, frequency, and duration.⁴⁷ They also vary in terms of supervision or whether they are home-based programs.⁴⁹ Some exercises may target specific muscle groups (e.g., back) while others may target specific muscles (e.g., transversus abdominus [TrA]).

Evidence from several systematic reviews has shown that exercise therapy is effective for chronic low back pain.⁴⁷ In 2010, vanMiddelkoop et al. reviewed evidence from 37 randomized controlled trials (RCTs) evaluating some type of exercise therapy for adults with nonspecific chronic low back pain. The authors concluded that exercise therapy is effective at reducing pain and function in treating chronic low back pain. However, they were unable to identify one specific type of exercise therapy as being more effective than others. Exercise interventions included in this review were aerobic exercise training programs, lumbar flexion exercise programs, general exercise (strengthening and stretching), motor control exercise programs (designed to improve function of specific trunk muscles, e.g., TrA), yoga (designed to improve flexibility), and conventional exercise.⁴⁷

In 2008, the North American Spine Society sponsored a series of articles in *The Spine Journal* focusing on evidence-informed management of nonsurgical approaches to chronic low back pain. One article in this series reviewed the evidence for lumbar stabilization exercises (also referred to as core stabilization, or segmental stabilization).⁵⁰ The aim of these exercises is to maintain stability in the lumbar spine by improving the neuromuscular control, strength, and endurance of muscles central to maintaining dynamic spinal and trunk stability (e.g., TrA,

lumbar multifidi). Lumbar stabilization exercises are generally administered by physical therapists, exercise physiologists, personal trainers, and Pilates instructors. In the United States, stabilization training is typically administered over 6–12 sessions in a physical rehabilitation setting.⁵⁰

Standaert et al. 2008⁵⁰ included three studies (2 high quality, 1 low quality) in a review. Studies included in this review collectively excluded patients with prior spine surgery, grade III or IV spondylolisthesis, spinal stenosis, and inflammatory spinal disease. Exercise interventions included stabilization exercises (includes structured endurance training of the deep abdominal and back extensors); spinal stabilization exercises (designed to selectively retrain TrA, multifidus, pelvic floor, and diaphragmatic muscles); and motor control exercises. Comparative treatments included conventional physical therapy (active exercise and minimal use of passive modalities), manual therapy (administered by a physical therapist), and spinal manipulation. Evidence suggests that lumbar stabilization exercises are effective at improving pain and function in a heterogeneous group of patients. The authors caution that until more data are available, lumbar stabilization exercise should be considered a “useful tool” for treating patients with chronic low back pain.

Another article in this series focused on the management of chronic low back pain with lumbar extensor strengthening exercises.⁴⁸ These exercises, based on standard resistance training principles, target the lumbar erector spinae and multifidus muscles. Exercises are performed using equipment such as isotonic machines, benches and Roman chairs, free weights, and floor and stability balls. Resistance training programs are typically administered in one to three sessions per week with cost ranging from \$50 to \$200 per month, depending on the level of supervision.⁴⁸ Programs are prescribed and monitored by licensed clinicians (e.g., physicians, physical therapists), but can be supervised by a personal/athletic trainer or exercise physiologist at an outpatient setting such as a health and fitness facility. Mayer et al. 2008⁴⁸ state that home exercises are commonly prescribed during the final stages of rehabilitation. VanMiddelkoop et al. 2010 note that supervision is recommended when home exercises are prescribed, because of poor adherence to exercise prescription.⁴⁷ Harms from lumbar strengthening exercises include delayed onset lumbar muscle soreness.⁵¹

In 2008, Mayer et al. reviewed 11 RCTs evaluating lumbar extensor strengthening exercise for chronic low back pain.⁴⁸ Patients randomly assigned to exercise interventions performed isolated lumbar extensor progressive resistance exercises using equipment such as isokinetic and isotonic machines, and variable resistance dynamometer machines. Comparative treatments included spinal manipulation, aerobic exercise, or no intervention. When compared with no treatment and most passive modalities, results indicated that lumbar extensor strengthening exercises alone or with co-interventions provided short-term benefits in pain, disability, and other patient-oriented outcomes for patients with chronic low back pain. In the long term, however, some of the relative benefits in pain and disability of lumbar extensor strengthening exercise versus other interventions were not maintained.

Multidisciplinary and behavioral treatments have also been reported as effective in treating chronic low back pain.⁵² Multidisciplinary back training, partly based on physical training and cognitive behavioral training (CBT), focuses on long-term daily functioning of individuals.⁵³ VanGeen and colleagues evaluated the long-term effectiveness of multidisciplinary back training in individuals with chronic low back pain.⁵³ In this systematic review, multidisciplinary training programs contained a range of two (physical and educational or psychological) to four elements (physical, educational, psychological, and social). The duration and intensity of the intervention

and control treatments (no treatment or low-intensity multidisciplinary back training) ranged from 2 hours to 35 hours a week. The authors identified a positive effect of multidisciplinary back training on work participation and quality of life, but no long-term effect on pain and function.

The Spine Journal series also reviewed evidence on behavioral treatments for chronic low back pain.⁵⁴ This article discusses a comprehensive guide by Gatchel and Robinson⁵⁵ of a typical CBT intervention for a patient with chronic low back pain. This “short-term, skills-oriented therapy” consists of new skills being taught in successive sessions. Skills included the following: (1) correct negative (distorted) thinking about chronic pain, (2) control emotional reactions to chronic pain, and (3) cope more effectively with chronic pain and other stressors. CBT programs are typically administered in a private practitioner’s office or a specialized outpatient pain or spine clinic by a therapist or other licensed mental health professional.⁵⁴ The authors report that evidence from previous systematic reviews and “numerous well-conducted studies” have demonstrated that “CBT and behavioral treatments for chronic pain reduced patients’ pain, distress, and pain behavior, and improved daily functioning.” A 2005 article on cognitive behavioral treatments for chronic pain called CBT effective with “potential economic benefits.”⁵⁶

A separate review by vanMiddelkoop et al. 2011⁵² evaluating effectiveness of physical and rehabilitation interventions for nonspecific chronic low back pain included 83 RCTs (6 studies evaluating multidisciplinary treatment, 21 studies evaluating behavioral treatment). Conclusions of interest include the following: (1) all types of behavioral therapy were more effective in reducing pain intensity, compared with waiting list controls; (2) behavioral components were shown to potentially reduce sick leave and costs due to sick leave; and 3) multidisciplinary treatment was more effective in reducing pain intensity compared with no treatment or waiting list controls and active treatments (e.g., exercise therapy, physiotherapy, usual care).

Other popular treatments for chronic low back pain are chiropractic and massage therapy. According to the American Chiropractic Association, the chiropractic approach is to pinpoint the cause of back pain and treat it directly. A chiropractic doctor may use one or a combination of the following: realign the spine or extremities by chiropractic adjustments, use physiotherapy for the muscles and ligaments, or involve rehabilitative exercises.⁵⁷ A 2005 study in the *Journal of Manipulative and Physiological Therapeutics* compared 2,780 patients with mechanical low back pain self-referred to chiropractic care (e.g., spinal manipulation, physical therapies) or medical care (e.g., prescription drugs, exercise). Chiropractic patients with chronic low back pain demonstrated clinically important differences in pain and disability improvement, compared with patients with chronic low back pain undergoing medical care. Evidence indicated that chiropractic care is relatively cost effective for treating patients with chronic low back pain.⁵⁸

Massage is commonly defined as soft-tissue manipulation using hands or a mechanical device on any body part.^{59,60} Massage may be applied to a localized region (e.g., lumbar region) or to the entire body by extensively trained licensed massage therapists, physical therapists, or chiropractors. Common types of massage therapy are acupuncture (Shiatsu), Rolfing, Swedish Massage, reflexology, and craniosacral therapy.⁶⁰ Massage therapy is considered safe with minimal side effects (e.g., soreness). Furlan et al. 2008⁵⁹ reviewed the evidence from 13 RCTs assessing various types of massage therapy for low back pain. The authors concluded that massage might be beneficial for patients with subacute and chronic nonspecific low back pain and more beneficial when combined with exercises (such as stretching) and education. Evidence indicated that acupuncture massage (massage applied to acupuncture points) is more effective than classic massage.

Invasive Treatments for Degenerative Lumbar Conditions

Several surgical and nonsurgical invasive treatments are available to treat degenerative lumbar conditions. Artificial total disc replacement (TDR, also referred to as artificial intervertebral disc replacement) is increasingly used as an alternative to treat lumbar degenerative disc disease.⁶¹ TDR involves surgically removing a diseased vertebral disc (discectomy) and implanting a synthetic one. It is hypothesized that by undergoing lumbar TDR, a patient's normal intervertebral segment motion is restored and maintained while the adjacent level is prevented from nonphysiologic loading, thus relieving pain.⁶² TDR is performed under general anesthesia by an orthopedic spine surgeon or neurosurgeon. Although commercially available since the late 1980s,⁶³ only two lumbar total disc prostheses have received approval from the U.S. Food and Drug Administration (FDA): the Charité (DePuy Spine, Inc., Raynham, MA), and ProDisc-L (Synthes Spine, West Chester, PA);⁶¹ these manufacturers are both owned by Johnson & Johnson (New Brunswick, NJ). Device migration out of the implanted location was the most frequently reported adverse event reported in the Manufacturer and User Facility Device Experience database from August 2003 through November 2005.⁶⁴ In September 2007, the U.S. Centers for Medicare & Medicare Services (CMS) announced the decision of noncoverage of lumbar artificial disc replacement for Medicare beneficiaries older than 60 years of age.⁶⁴

Discectomy (open or microsurgical) involves removal of a diseased disc; laminectomy (also known as lumbar decompression) involves the removal of bone or soft tissue compressing the contents of the spinal canal.⁶⁵ Laminectomy is commonly performed to treat spinal stenosis and may be combined with foraminotomy (surgical enlargement of the opening through which the spinal nerves leave the spinal column) or spinal fusion. Risks from surgery include damage to a spinal nerve, partial or no postoperative relief, and infection in vertebral bones.⁶⁵

Nonsurgical invasive measures to treat degenerative lumbar conditions include injection therapy or denervation procedures. Facet or epidural steroid injections may be administered to treat degenerative spondylolistheses.⁶⁶ With fluoroscopy guidance, facet injections are delivered into the facet joint (also known as zygapophysial or z-joint; intraarticular) or into a nerve (medial branch blocks) using a local anesthetic.⁶⁷ Pain relief from the injection (containing an anesthetic and steroid) may take from 48 hours to 2 weeks. Although uncommon, transient leg weakness or numbness and tingling have been reported.⁶⁷

Delivery of epidural steroid injections includes placement of steroids and local anesthetic into the epidural space. Three commonly used methods for delivering epidural injections include the interlaminar (commonly known as an epidural injection), caudal, and transforaminal approaches (also referred to as a nerve block). Interlaminar and caudal approaches allow steroid delivery to reach several spinal segments and both sides of the spinal canal.⁶⁷ The transforaminal delivery is more concentrated, usually reaching one segment and one side of the spinal canal. Epidural injections, performed on an outpatient basis, may take from one to several days to relieve pain. Serious complications such as allergic reaction, bleeding, infection, nerve damage, or paralysis have been rarely reported.⁶⁷ Both facet and epidural steroid injections may be safely repeated and are typically combined with other therapeutic methods (e.g., physical therapy).

Denervation procedures may be used to treat lumbar discogenic pain, a major problem in lumbar degenerative disc disease. Percutaneous intradiscal radiofrequency thermocoagulation (PIRFT) involves placing an electrode or catheter into the intervertebral disc. A similar procedure, intradiscal electrothermal therapy (IDET), involves placing an electrode around the annulus or nucleus of the intervertebral disc.^{68,69} During PIRFT and IDET, alternating

radiofrequency current and electrothermal energy, respectively, are used to alter pain receptors.⁶⁸ CMS currently considers PIRFT and IDET to be investigational in nature and thus medically unnecessary for relief of discogenic pain.⁶⁹

Lumbar Fusion

Spinal fusion (also known as spinal arthrodesis) is an inpatient surgical procedure intended to permanently immobilize the functional spinal unit (2 adjacent vertebrae and the disc between them), limiting painful movement.⁴ Lumbar fusion is the fusion of two or more vertebrae in the lumbar spine (low back). Most fusions performed today use instrumentation such as pedicle and facet screws, rods, and cages⁵ and incorporate a combination of graft materials made of a patient's own bone (autograft), donor bone (allograft), or a synthetic substance such as recombinant human bone morphogenetic protein (rhBMP) to promote fusion. Surgeons can initiate the procedure through the peritoneum (membrane lining the abdomen) or retroperitoneum (anterior approach), the back (posterior approach), or a combination of sites (anteroposterior approach). Surgical techniques include the following:⁶

- Posterolateral fusion (PLF): Dorsal surgery (from the back of the body) that joins vertebrae by their transverse processes
- Posterior lumbar interbody fusion (PLIF): Dorsal surgery that joins vertebrae by their bodies
- Transforaminal lumbar interbody fusion: A form of PLIF that joins vertebrae on one side only
- Anterior lumbar interbody fusion (ALIF): Anterior surgery (from the front of the body) that can be performed by open transperitoneal (through the peritoneum) or, more commonly, retroperitoneal (from behind the abdominal cavity), mini open, or laparoscopic techniques
- Circumferential fusion: 360° fusion that joins vertebrae by their entire bodies and transverse processes, typically performed by combining PLF and ALIF

Lumbar fusion's main potential advantage is to provide pain relief and restore quality of life and function when less extensive and invasive treatments cannot. However, it poses potential harms ranging from anesthesia risks and surgical complications to a need for subsequent reoperation for later complications.⁷

Information on FDA-approved devices (fusion and nonfusion) used in this evidence base is included in Table 15 of Appendix C.

Assessment Methods

Key Questions and Scope

In part, this report updates a previous report on spinal fusion conducted for the Agency for Healthcare Research and Quality (AHRQ) in November 2006 by the Duke University Evidence-Based Practice Center (McCroory et al.).⁷ The Duke review (currently available only in draft form) was primarily focused on outcomes of lumbar fusion in patients age 65 or older with degenerative disc disease compared to nonsurgical management or other surgical strategies. The patient population for the current report has been expanded at the request of AHRQ to include adults 18 years of age and older. Key questions have also been added for comparing different spinal fusion strategies. For surgical comparisons, perioperative outcomes have been added. Finally, patients addressed in this report must have pain attributed to the condition for which they undergo treatment.

This report addresses 10 key questions (KQs), which are listed below. The subsequent table (Table 1) and following text clarify the scope of each key question using a standard format that addresses populations, interventions, comparisons, outcomes, timing, and settings (PICOTS). Initially a panel of eight key informants, which included one neurosurgeon, one individual from a payer organization, two industry representatives, one hospital purchasing representative, one president of a patient advocacy foundation (who was also an orthopedic surgeon), and two patients who had previously undergone the surgical procedure, gave input on the key questions to be examined. These KQs were then posted on AHRQ's website for public comment in January 2012 for 4 weeks and revised as needed.

We then drafted a protocol for the report and recruited a panel of technical experts to provide high-level content and methodological expertise throughout the development of the review. We finalized the review protocol in the winter of 2012 based on input from nine technical experts: two physicians (one internist and one family practitioner), two surgeons (one orthopedic and one neurosurgeon), one biostatistician, one chiropractor, one physical therapist, and two individuals from payer organizations.

Key Question 1

For adults with pain associated with degenerated disc(s) of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from continued noninvasive treatment in:

- a. Patient-centered outcomes such as function, quality of life, or pain?
- b. Adverse events?

Key Question 2

For adults with pain associated with degenerative (not congenital) stenosis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from continued noninvasive treatment in:

- a. Patient-centered outcomes such as function, quality of life, or pain?
- b. Adverse events?

Key Question 3

For adults with pain associated with degenerative spondylolisthesis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from continued noninvasive treatment in:

- a. Patient-centered outcomes such as function, quality of life, or pain?
- b. Adverse events?

Key Question 4

For adults with pain associated with degenerated disc(s) of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from other invasive procedures (e.g., total disc replacement, disc decompression) in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 5

For adults with pain associated with degenerative stenosis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from other invasive procedures (e.g., decompressive laminectomy and minimally invasive procedures, including those using devices) in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 6

For adults with pain associated with degenerative spondylolisthesis of the lumbar spine, does spinal fusion performed alone or in conjunction with additional surgery differ from other invasive procedures (e.g., repair, vertebrectomy) in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 7

For adults with pain associated with degenerated disc(s) of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation, graft material) performed alone or in conjunction with additional surgery differ in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 8

For adults with pain associated with degenerative stenosis of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation, graft material) performed alone or in conjunction with additional surgery differ in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 9

For adults with pain associated with degenerative spondylolisthesis of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation, graft material) performed alone or in conjunction with additional surgery differ in:

- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?
- b. Patient-centered outcomes such as function, quality of life, or pain?
- c. Adverse events?

Key Question 10

Which, if any, patient characteristics (e.g., pain severity, type or duration of prior treatment) does the evidence show are associated with better or worse outcomes after spinal fusion in adults with degenerative disc changes, degenerative stenosis, or degenerative spondylolisthesis?

- a. Patient-centered outcomes such as function, quality of life, or pain
- b. Adverse events

The intent of these key questions is to assess the performance of different treatments in groups of reasonably similar patients (as defined in the inclusion criteria). Comparing outcomes of incomparable patients (e.g., different indication, different severity, different serious comorbidity, such as osteoporosis) who undergo different treatments does not inform this objective and is outside the scope of this report.

For each key question, we provide a description of the included studies, including the following: patient indications; method(s) of diagnosis; inclusion and exclusion criteria, including diagnostic criteria for degenerated discs or degenerative joint disease; prior, concurrent, and subsequent treatments; ancillary treatments; and surgical techniques and devices used. To limit clinical heterogeneity, different treatment comparisons were addressed separately within the key questions. We planned to statistically investigate patient and treatment factors (e.g., presence or absence of radicular pain, prior treatments, preoperative pain severity) for association with outcome in Key Question 10; however, there were insufficient data to permit such analyses, so we conducted a qualitative review of primary literature reporting patient-level data.

Table 1 depicts the comparators, outcomes, and primary indications for each key question. Populations are addressed in the text below. For all of the key questions, the intervention is lumbar spinal fusion, and any timing of followup and setting will be considered.

Table 1. Scope of key questions: comparators, outcomes, and primary indications

Comparator	Outcomes	Primary Indication	Key Question Number
Noninvasive treatment	Patient-centered* Adverse events	Degenerated disc(s)	1
		Degenerative stenosis	2
		Degenerative spondylolisthesis	3
Invasive treatment other than fusion	Perioperative† Patient-centered Adverse events	Degenerated disc(s)	4
		Degenerative stenosis	5
		Degenerative spondylolisthesis	6
Different fusion surgeries	Perioperative Patient-centered Adverse events	Degenerated disc(s)	7
		Degenerative stenosis	8
		Degenerative spondylolisthesis	9
All	All	All	10

* Patient-centered outcomes are function, quality of life, and pain

† Perioperative outcomes are surgery time, blood loss, and length of hospital stay

Populations

The patient population comprises adults with low back pain and/or leg pain associated with lumbar degenerated disc(s) or degenerative joint disease (e.g., stenosis, spondylolisthesis) in the absence of adult deformity. Patients must have both pain and have received a diagnosis of degenerated disc(s), degenerative stenosis, or degenerative spondylolisthesis thought to be causing the low back pain. Using statistical methods such as subgroup analysis or meta-regression, we intended to explore if diagnostic criteria or patient characteristics (e.g., race and ethnicity, duration of symptoms, presence or absence of radicular pain, occupational status, workers compensation or litigation, and presence of comorbid conditions) were associated with differences in outcome. However, the data from the studies that made up the evidence base for the key questions did not permit such exploration. We, therefore, conducted a qualitative review of primary studies that used patient-level data to assess which factors may indicate which patients were more or less likely to experience fusion success or failure.

Intervention

The intervention is spinal fusion. Any clinically relevant method of performing spinal fusion—both the surgical approach and the instrumentation or other implanted material (e.g., bone morphogenetic protein) used—is included. Abandoned methods (those no longer being used) were not reviewed. The Technical Expert Panel (TEP) was consulted to identify which methods have been abandoned or are no longer relevant to practice in the United States. Differences in treatments (e.g., hardware, surgical approach) are investigated for association with outcome where data allow.

Comparators

For comparators, only direct comparisons are considered. Indirect comparisons and historically controlled trials are not reviewed because of high risk of bias.

- Key Questions 1–3: Noninvasive treatments. Any noninvasive treatment or combination of noninvasive treatments were considered (e.g., antispasmodics, bed rest, cardiovascular exercise, chiropractic care, cognitive behavior therapy, complementary and alternative medicine, facet or epidural steroid injections, NSAIDs or acetaminophen, opioids, physical therapy).
- Key Questions 4–6: Other invasive procedures (e.g., total disc replacement, discectomy, surgical decompression, injections, percutaneous procedures).
- Key Questions 7–9: Direct comparison of different spinal fusion surgeries (e.g., different approaches, techniques).

Outcomes

- Perioperative outcomes (Key Questions 4–9)
 1. Surgery time
 2. Blood loss
 3. Length of hospital stay
- Patient-centered outcomes (short- and long-term effectiveness)
 1. Function (e.g., Oswestry Disability Index; return to work or other daily activities; activities of daily living)
 2. Quality of life (e.g., Short Form [36] Health Survey [SF-36])
 3. Pain (e.g., visual analog scales (VASs) or numerical rating scales; pain medication needed)

- Adverse effects of intervention(s)
Any harms reported in the literature including, but not limited to the following: reoperation (e.g., for revision or device removal with or without replacement), total reoperation rate (which might include reoperation due to adverse events or additional reasons for further surgery such as adjacent segment degeneration), neurologic injury, blood clots, and infection

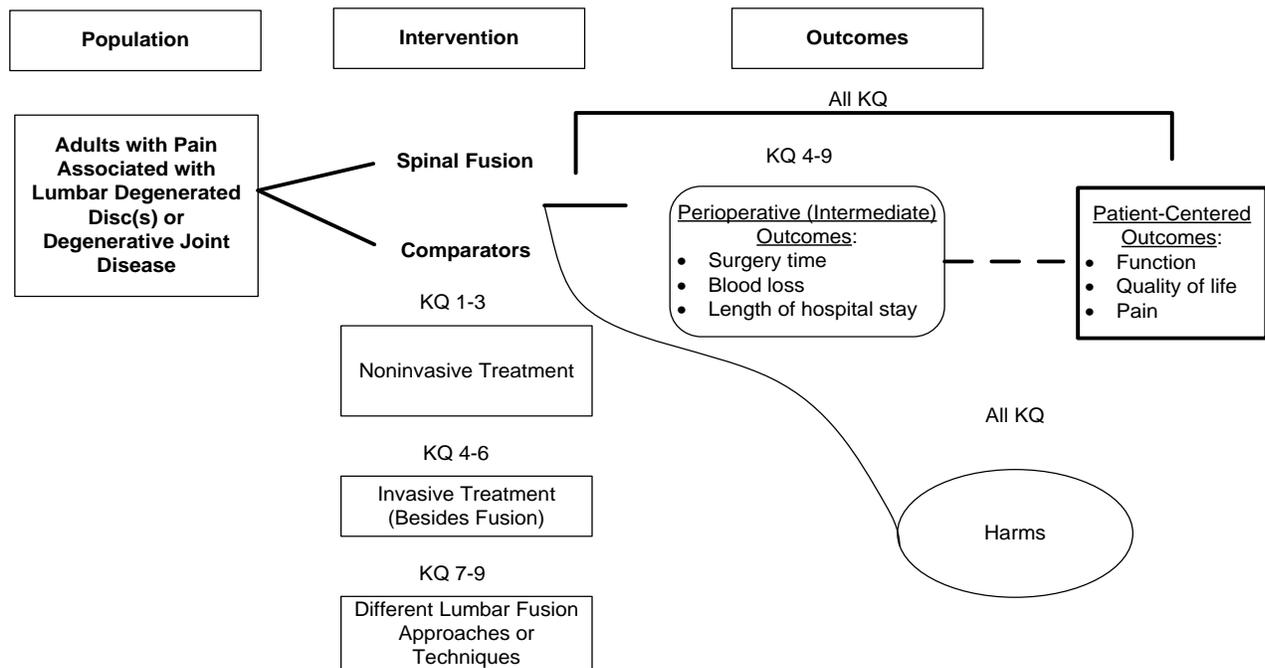
Timing

- Perioperative: up to 2 weeks postsurgery (Key Questions 4–9)
- Short term: 2 weeks to 6 months
- Intermediate term: 6 months to 4 years
- Long term: longer than 4 years

Setting

- We included all settings.

Figure 1. Analytic framework for spinal fusion for treating painful lumbar degenerated disc(s) or degenerative joint disease



KQ = Key question

Criteria for Study Inclusion/Exclusion

The inclusion criteria are listed below in separate categories pertaining to the following: publication, study design, patient characteristics, treatment characteristics, and data.

Publication Criteria

1. Publication must be a full clinical study; abstracts alone were not included. Abstracts do not include sufficient details about experimental methods to permit an evaluation of study design and conduct, and they may not report all outcomes.^{70,71} Similarly, letters, editorials, and other publications that are not full-length, clinical studies were not accepted as evidence. Abstracts of randomized studies that did not subsequently appear as full articles would have been flagged as possible evidence of publication bias.
2. To capture the most relevant data, we included studies published on or after January 1, 1995.
3. To avoid double-counting of patients, in cases in which we found more than one report in which patient populations overlap, only outcome data from the most recent report, or if the duration of followup is the same, the report with the largest number of patients will be included. We checked smaller reports for data on outcomes that were not provided in the largest report. Multiple publications of the same study (e.g., publications reporting subgroups, other outcomes, or longer followup) were identified by examining author affiliations, study designs, enrollment criteria, number of patients enrolled, and enrollment dates.
4. Studies must be published in English. Non-English-language studies are unlikely to reflect care settings applicable to care in the United States.

Study Design Criteria

1. The study and its data collection must be prospective.
2. The study must either randomly assign patients to treatments or use an analytic method to address and protect against selection bias, such as baseline matching on multiple characteristics or propensity scoring. Studies with large differences at baseline between groups (regardless of whether they were randomized) or confounding by indication (e.g., 2 groups are treated for different indications and then compared), are excluded. If the importance of baseline differences were unclear, we solicited TEP expertise to determine inclusion. Studies in which patients were allocated to different treatments based on their characteristics were excluded because of excessive risk of selection bias and lack of comparability between groups.
3. Fusion and the comparator treatment must have been administered during the same time period to eliminate potential bias due to differential time frames and related factors.
4. Studies addressing Key Question 10 (regarding identification of prognostic factors in primary literature) must have either met items 2a, 2b, and 2c above, or be primary studies with the primary objective of evaluating prognostic factors (as determined by title and abstract screening) for outcomes following lumbar spinal fusion using patient-level data from patients with degenerated discs, degenerative stenosis, and/or degenerative spondylolisthesis using statistical methods such as subgroup comparisons or regression analysis.

Patient Criteria

1. The study must provide data for which at least 85 percent of the patients had the condition specified in the key question unless data is reported separately for the subgroup population. This criterion is necessary since if >15% of patients had a different condition, the estimate of treatment effect may be inaccurate.
2. At least 85 percent of patients must have the cause of the disc changes, stenosis, or spondylolisthesis associated with degenerative changes (e.g., not isthmic or traumatic).
3. The mean number of levels treated, or proportion of patients with one, two, and three levels treated, must be similar in the treatment and comparison groups.
4. At least 85 percent of patients will be undergoing primary, not revision, fusion.
5. We define “adults” as people at least 18 years of age.
6. No specific diagnostic criteria are required so long as the primary investigators state that the conditions being treated are degenerated disc(s), degenerative stenosis, or degenerative spondylolisthesis; however, all reported details regarding how disease was diagnosed or selected for inclusion in the study were extracted and considered as co-variants potentially associated with the outcomes.
7. Assessment of treatment for other indications, including but not limited to deformity, is outside the scope of this report.

Treatment Criteria

1. Lumbar fusion outcomes must be directly compared with the outcomes of a comparator treatment. Literature and historical controls were not considered because of excessive risk of bias.
2. Fusion methods no longer used in the United States or not commercially available in the United States were not reviewed. We solicited TEP expertise to determine which methods and devices are not relevant to current U.S. practices.

Data Criteria

1. The study must report data on at least one of the included outcomes for at least one of the key questions.
2. Outcome data must not rely on retrospective recall (e.g., preoperative data collected long after the procedure was performed) because such outcomes may not accurately reflect patients’ experiences.
3. We included data points capturing at least 10 patients with the condition of interest who represented at least 50 percent of eligible enrolled patients.

Searching for the Evidence: Literature Search Strategies for Identifying Relevant Studies To Answer the Key Questions

Medical librarians in the Evidence-based Practice Center (EPC) Information Center performed literature searches following established systematic search protocols. The searches were developed by a Senior Information Specialist and approved by the Director of Health Technology Assessment/EPC Information Center.

We searched the following databases on the OVID SP platform using the one-search and deduplication features: MEDLINE, PreMEDLINE, and EMBASE. The CINAHL database

(Cumulative Index to Nursing and Allied Health Literature) was searched on the EBSCOhost platform. The Cochrane Library (including the Central Register of Controlled Trials, the Cochrane Database of Methodology Reviews, and the Cochrane Database of Systematic Reviews), the Database of Abstracts of Reviews of Effects, the Health Technology Assessment Database, and the United Kingdom National Health Service Economic Evaluation Database were also searched for unique reviews, trials, economic analyses, and technology assessments. Our searches covered the time period of January 1, 1995 through February 7, 2012.

Other mechanisms used to retrieve additional relevant information included a review of bibliographies/reference lists from peer-reviewed and gray literature. Gray literature includes reports, studies, articles, and monographs produced by federal and local government agencies, private organizations, educational facilities, consulting firms, and corporations that typically do not appear in the peer-reviewed journal literature.

Search terms included a combination of subject headings and key words identified by the following: (1) reviewing relevant systematic reviews on similar topics that are identified by the research staff; (2) reviewing how other relevant studies are indexed, their subject heading terms, and their keywords; and (3) reviewing MeSH and Emtree indexes for relevant and appropriate terms. Each strategy was reviewed by the investigators and the Director of the Health Technology Assessment/EPC Information Center. A study-design filter was applied to retrieve systematic reviews and clinical trials. Feedback from the Senior Investigator and the Director of the Health Technology Assessment/EPC Information Center—including details regarding gaps in the search strategy and articles (identified by the Senior Investigator) not retrieved by the searches—was integrated into the search strategy using key terms and subject headings. Updated strategies are rerun in all identified databases. Additional results were scanned, and their relevancy assessed by the Medical Librarians. New results were downloaded and forwarded to the Senior Investigator for review. Hand searches of reference lists in identified articles were also reviewed for possible inclusion. All the resources and detailed search strategies used in this report are presented in the *Literature Search Methods* section of Appendix A.

All abstracts were screened in duplicate using DistillerSR (EvidencePartners, Ottawa, Ontario, Canada). In cases of unresolved conflict, the article was ordered. Two reviewers reviewed the full-length articles for inclusion. In cases of unresolved conflict, a third senior methodologist mediated. Full articles meeting the inclusion criteria were retained for extraction of general study characteristics, patient characteristics, treatment characteristics, risk-of-bias items, and outcome data (see next section).

Abstracting and Managing Data

We extracted study information and outcomes data into Word (Microsoft Corp., Redmond, WA). The forms were pilot tested by three team members, and a single reviewer extracted data. A second reviewer randomly selected and audited 10 percent of the data to verify accuracy. Discrepancies would have prompted a greater audit rate. Risk-of-bias items were judged in duplicate, and discrepancies were resolved with discussion. Information extracted from each study included:

- **General study characteristics:** Author, publication year, country, study design, dates of patient enrollment, length of followup, funding source, and key question(s) addressed.
- **Patient characteristics:** Number of enrolled patients, age, sex, type of degeneration, how diagnosed, presurgical pain level, presurgical quality-of-life scores, presurgical functional activity scores, level(s) affected, number of levels affected, primary/recurrent, prior

treatment, pending litigation or other legal factors, workers compensation or other occupational factors, and depression or other psychological factors.

- **Treatment characteristics:** Surgical approach, device(s) or other materials used, adjunctive surgery, surgeon experience and caseload, and facility type (e.g., tertiary care, community hospital).
- **Risk-of-bias items:** See the next section on methodological quality.
- **Outcome data:** Study methods of followup for data collection were extracted, as well as the time point(s) of evaluation. For each included outcome, we extracted the number of patients contributing data to each extracted time point. We extracted the numerical data necessary to compute an effect size and measure of variance for all included outcomes for each study. These may have included means, standard deviations, counts, proportions, results of authors' statistical tests, or other statistical details, depending on what was reported.

Assessing Individual Studies' Methodological Quality (i.e., Risk of Bias)

As stated above, because of the possibility of subjective interpretation, methodological-quality assessments were performed by two extractors for each study, and discrepancies were resolved by consensus. We assessed the risk of bias (i.e., internal validity) separately for each outcome and for each time point. Some subjective outcomes are more susceptible to bias than others, and longer durations of followup often result in attrition or right censoring. This may lead to outcomes assessment only in patients who are somewhat different from the full set of enrolled patients and thereby introduce a systematic difference between the groups being compared.

Table 2 lists the risk-of-bias assessment items. Achieving some of these items may be logistically prohibitive in surgical studies. The purpose of this instrument is to gauge the potential risk of bias in a study, and such risk can exist regardless of whether achieving the item is reasonably possible. It is possible that study organizers may do an exemplary job of designing and coordinating a clinical trial, but the study will not be rated as low risk of bias (or high quality) because of risks of biases that were unavoidable. Taking risks of biases into account when determining risk-of-bias ratings is important because factors such as lack of blinding and differences in group allocation can lead to overestimation of treatment effects, especially when outcomes are subjective (e.g., pain, function, quality of life).⁷²

For all studies with control groups (regardless of whether patients were randomly assigned to groups), we assessed the risk of bias using the items in the table below. Each of these items were answered as "Yes," "No," or "Not reported." Some items were always answered "Yes" (see comments) because of the inclusion criteria. While these factors will not differentiate risk of bias among included studies, they will reflect the decrease in potential risk of bias that the inclusion criteria confer to the entire evidence base.

Table 2. Risk-of-bias assessment items

Item	Comment
1. Were patients randomly assigned to treatment?	
2. Was group allocation concealed?	
3. For nonrandomized trials, did the study employ any other methods to enhance group comparability?	All included studies will have a “Yes” for this question, because nonrandomized studies are required to use an analytic control to address selection bias.
4. Was patient assignment to treatment made independently from physician and patient preference?	
5. Did patients in different study groups have similar levels of performance on the outcome of interest when assigned to groups?	This will be used only for outcomes that have a baseline measure (e.g., quality of life).
6. Were the study groups comparable for all other important factors at the time they were assigned to groups?	
7. Was the comparison of interest prospectively planned?	
8. If patients received ancillary treatment(s), did groups have a $\leq 5\%$ difference between them in the proportion of patients receiving each ancillary treatment?	If ancillary treatments differed substantially between groups, the study will be excluded.
9. Were the 2 groups treated concurrently?	All included studies will have a “Yes” for this question, because it is a requirement for inclusion.
10. Were patients blinded to their treatment assignment?	For most comparisons this will not be possible, but it remains a potential source of bias.
11. Were outcome assessors blinded to which treatment patients had?	
12. Was the outcome measure of interest objective, and was it objectively measured?	Objective outcomes include adverse events, length of hospital stay, and other perioperative outcomes. Subjective outcomes include pain, quality of life, and functional status, including return to work and return to daily activities.
13. Was there $\leq 15\%$ difference in the length of followup for the 2 groups?	This item is not relevant to perioperative outcomes.

We categorized each study as “Low,” “Moderate,” or “High” risk of bias using the following method:

- To be considered low risk of bias, the study must meet ALL of the following conditions:
 - Randomized (item 1).
 - Allocation concealment (item 2) OR blinded patients and outcome assessors (items 10 and 11) OR both.
 - Good baseline comparability for both outcome (item 5) and other patient characteristics (item 6).
 - If NOT blinded outcome assessors (item 11) (or not reported blinded outcome assessors), then it was an objective outcome (item 12).
 - ≤ 15 percent difference in length of followup between groups (item 13).

- To be considered high risk of bias, the study must meet AT LEAST TWO of the following criteria:
 - Process of assigning patients to groups NOT made independently from physician and patient preference (item 4).
 - Poor baseline comparability for either the outcome (item 5) or other patient characteristics (item 6).
 - Retrospective (post-hoc) analysis (item 7).
 - Difference in ancillary treatments ≥ 5 percent (item 8).
 - Not a blinded outcome assessor (item 11) AND a subjective outcome (item 12).
- To be considered moderate risk of bias, the study neither meets the criteria for low risk of bias nor the criteria for high risk of bias.

Synthesizing Data

We planned to perform meta-analysis wherever appropriate and possible. Our minimum criteria were:

- At least two studies addressing the same outcome at the same duration of followup
- Clinical similarity in those studies, including patient characteristics, surgical approach and strategy, and comparability of control groups

We also planned to use subgroup analyses and meta-regression to explore whether variations among studies, such as those that follow, led to differences in outcome.

- Patient characteristics (e.g., comorbid diagnoses, previous treatments)
- Surgical approach
- Device(s) or other implanted materials (e.g., bone morphogenetic proteins) used

None of the evidence bases met the minimum criteria for performing meta-analysis, so we did not attempt to use meta-analysis to either determine quantitative effect sizes or to explore which variations in studies may have led to differences in patient outcomes.

We therefore performed qualitative analyses in which the studies comprising the evidence base for each key question were described, compared, and contrasted. Results are reported and interpreted, with supporting data tabled in Appendix C. For each study at each outcome and time point, p-values are reported (author-reported p-values were sometimes used when reporting was insufficient for the EPC to perform the calculation) to determine whether the difference between groups was statistically significant, and mean differences were calculated to determine whether the differences between groups were clinically significantly different (unless reported data were insufficient to enable this calculation). The mean difference shows how much difference there was between groups in terms of the scale the study used to measure the outcome. For instance, if one group had a mean pain score of 50 and the other had a mean pain score of 20, the mean difference is 30. We also calculated 95 percent confidence intervals (CIs) for the mean difference because the difference between groups can only be considered clinically significant if the lower interval is greater than the minimally clinically important difference (MCID).

The MCID is the minimum amount of change in an outcome measure, usually after treatment, which is considered to represent a perceptible benefit (or detriment). MCID is an objective way to determine whether a treatment outcome resulted in a meaningful change. This is in contrast to statistically significant change, which shows numerical differences that may not correspond to genuine improvement or detriment as measured by an outcome instrument. For instance, if a study finds patients have an average change in pain score from 8.5 to 8.0 on a scale

of 0–10, this difference may be statistically significant (particularly if the study is large) but is unlikely to represent a treatment benefit that is clinically important.

Determining thresholds for MCID is challenging, and MCID must be determined with respect to the patient population being assessed and the treatment in question.⁷³ Therefore, MCID is never a “fixed” value.⁷⁴ MCID can depend on baseline score, especially when it is a percentage change from baseline because the same percentage of two numbers yields a larger change in points for the larger number, and clinical experiences at opposite ends of the same scale can be very dissimilar.^{75,76} Expecting a larger MCID for a riskier treatment is justifiable because a greater tradeoff in potential benefit can be expected when risk of harm is higher.⁷³ The selection of the mathematical method of determining MCID can also be challenging, because several calculation methods exist (standard error of the mean, half of a standard deviation, an effect size, or a receiver operating characteristic curve) but no consensus has been reached regarding which method is best.⁷⁷ A general recommendation is that any MCID must be greater than the measurement error and correspond to patient perception of important change.⁷⁷

We divided MCID values of primary outcomes of interests (selected for being by far the most commonly used) identified in the literature (shown in Table 16 of Appendix C) into three tiers as follows: very unlikely to be clinically significant (lower than the lowest MCID identified), possibly clinically significant (greater than the lowest MCID identified but less than the highest MCID identified), and very likely to be clinically significant (equal to or greater than the highest value identified). These values, summarized in Table 3, represent the change from baseline score to followup score, not the difference between two groups at followup.

For instrumentes measuring pain or function for which no literature-based MCID could be identified, we considered a 30 percent difference to be clinically significant. This number is based on a study by Raymond et al. in which the objective was to determine meaningful changes for back pain and function using different methodologies.⁸ These authors looked at several commonly used instruments to measure pain and function and determined that a 30 percent change from baseline may be considered a clinically meaningful improvement when comparing pre to post-treatment scores. Ideally, we would be able to compare the proportions of patients who achieve at least MCID value changes; however, such data was not usually reported. No literature on perioperative MCID values (i.e., surgery time, blood loss, and length of inpatient stay) or adverse events for this patient population were identified.

Table 3. Minimally clinically important differences of primary outcomes

Domain	Instrument	Very Unlikely To Be Clinically Significant	Potentially Clinically Significant	Very Likely To Be Clinically Significant
Function	Oswestry Disability Index	<8.2 points	8.2–20 points	≥20 points
Pain, back	11-point Visual Analogue Scale	<1.2 points	1.2–2.0 points	≥2.1 points OR reduction to <3/10
Pain, back	101-point Visual Analogous Scale	<10.0 points	10.0-18.0 points	≥18 points
Pain, leg	11-point Visual Analogue Scale	<1.6 points	1.6–2.7 points	≥2.8 points OR reduction to <3/10
Quality of life	SF-36 physical function	<4.9 points	≥4.9 points OR 30% reduction from baseline when baseline score ≥50 points	Not applicable

We computed effect sizes and standard deviations using standard methods. Because meta-analysis was not possible, it was not possible to use statistical methods such as meta-regression

and subgroup analyses to identify prognostic factors for fusion success or failure. We therefore conducted a qualitative review of primary studies that used patient-level data to assess which factors may indicate which patients are more or less likely to experience fusion success or failure.

This report presents findings organized by key questions. Answers to key questions are stated first, in boldface, along with the evidence rating supporting that conclusion. Subdivisions of the evidence base will be made as needed. Alternately, a statement specifies that the evidence is insufficient to answer the question.

Strength-of-Evidence Grading

To grade the evidence for each key question, we implemented the rating system described by Owens et al.⁹ (also part of the “Methods Guide”⁷⁸) and used many of the principles described by Treadwell et al.⁷⁹ We rated confidence in each conclusion “High,” “Moderate,” “Low,” or “Insufficient.” A rating of “Insufficient” applies to evidence bases that involve no studies or where the evidence does not support a conclusion (e.g., when the uncertainty is so high that the evidence cannot discriminate among conflicting conclusions). We were fully transparent about the ratings process, so readers can replicate the process or use the individual domains to arrive at their own overall ratings.

Owens et al.⁹ describe the following eight domains to be considered: risk of bias, consistency, directness, precision, magnitude of effect, publication bias, dose-response association, and all plausible confounders that would reduce the effect. Judgments concerning applicability are not included in the rating of the strength of evidence, but see the next section for information on how we assessed applicability in this report.

Owens and colleagues defined the four ratings as:

- High—“High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect.”
- Moderate—“Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate.”
- Low—“Low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate of effect and is likely to change the estimate.”
- Insufficient—“Evidence either is unavailable or does not permit a conclusion.”

We used the following approach to grade each of the following domains for each outcome and duration of followup:

- Risk of bias: We used the methods described in the previous section, *Assessing Methodological Quality (i.e., Risk of Bias) of Individual Studies*. Because all studies were rated as moderate, the overall risk-of-bias rating of any evidence base was moderate.
- Consistency: For evidence bases with two or more studies addressing the same outcome at the same duration of followup, we evaluated the studies for qualitative consistency by determining whether they both had findings favoring the same intervention. If they did not, the evidence base was downgraded. For single study evidence bases, this item was considered “not applicable” and the strength of evidence was downgraded (generally to Insufficient because of lack of replication).
- Directness: Because we included only directly controlled studies that assessed patient-oriented outcomes, all studies were considered direct.

- Precision: We considered the precision of the individual studies, since we did not perform meta-analysis. We considered a study to have an imprecise outcome when the intervention that was favored could not be determined. Finding the effect size to be imprecise depends on the size of the CI around the mean difference of a single study or, if multiple studies are combined in meta-analysis, the random-effects summary statistic. If this interval is so wide that it includes a clinically significant (or substantial) effect in one direction and also a clinically significant effect in the opposite direction, then the evidence is inconclusive, and therefore uninformative.^{80,81} We considered studies with statistically significant findings to be sufficiently precise to support a qualitative conclusion, even if clinical importance could not be established.

If there were no studies for a given treatment comparison or key question, we rated the evidence as Insufficient. As noted earlier, a single study with moderate or high risk of bias was generally considered insufficient evidence because of lack of replication of findings, even if it reported statistically significant findings.

We provided evidence ratings for perioperative outcomes, and pain, function, and quality of life for Key Questions 1–9 (addressing comparative efficacy) when there was at least one study addressing at least one of these outcomes per question. We did not rate adverse events because event reporting was generally sparse, suggesting underreporting.

Assessing Applicability

Applicability was assessed by considering important patient characteristics (e.g., diagnosis, presurgical pain level, presurgical functional status, workers compensation or other occupational factors, prior surgery, other patient characteristics) and treatment characteristics (e.g., surgical approach, device or other materials used in surgery, adjunctive surgery). Based on a review of the data abstracted, we narratively summarized any patterns reflected in these factors that might affect the applicability of the evidence to the general population and to the Medicare-beneficiary population. We made no attempt to generate any rating or score for the applicability of the evidence. Our narrative summaries are intended to draw stakeholders' attention to potential applicability issues embedded in the evidence.

Results

Introduction

In this chapter, the reader will find the results of our literature search, including information about how many articles were retrieved and reviewed in full for the review of benefits and harms. Reasons for exclusion of articles reviewed in full are presented in this chapter in Figure 2 and Figure 3 and in Appendix B. This is followed by the key findings for Key Question 1 to 10. The presentation of the findings includes a detailed description of the included studies, which includes basic study design information, inclusion/exclusion criteria, description of the enrolled patients and treatments, outcomes reported, and a description of the instruments used to measure each outcome. This is followed by a more in-depth description of the study findings, a description of individual study risk-of-bias assessments, and strength of evidence ratings for the body of evidence.

Results of Literature Searches

Extensive literature searches identified 4,378 citations potentially addressing the comparative benefits and harms of lumbar fusion (Key Questions 1–9). Of those, 4,230 were excluded upon title and abstract review for clearly not meeting inclusion criteria (e.g., not pertinent to the topic, uncontrolled or historically controlled, not a clinical study). Of the 148 articles retrieved and reviewed in full for the review of benefits and harms, 128 were excluded. Reasons for exclusion are summarized in Figure 2 below. This figure collapses reasons for exclusion into broad categories. For instance, the category “not relevant population” includes studies, such as Fairbank et al.⁸² in which fewer than 85 percent of patients had degenerative disc disease and Weinstein et al.⁸³ in which fewer than 85 percent of patients received primary fusion (89 percent of patients in this study received decompressive laminectomy without fusion). Reasons for exclusion are also itemized by study in Appendix B.

Eighteen clinical studies (reported in 20 publications) remained for inclusion, of which four addressed Key Question 1, one addressed Key Question 3, two addressed Key Question 4, one addressed Key Question 5, six addressed Key Question 7, two addressed Key Question 8, and three addressed Key Question 9. One study addressed more than one key question and no studies addressed Key Questions 2 or 6. The included studies are listed in Table 4, which follows, and described in the finding section by Key Question addressed. The outcomes and description of the instruments used to measure the outcomes are presented in Table 5.

Figure 2. Study selection for Key Questions 1–9 (comparative benefit and harm)

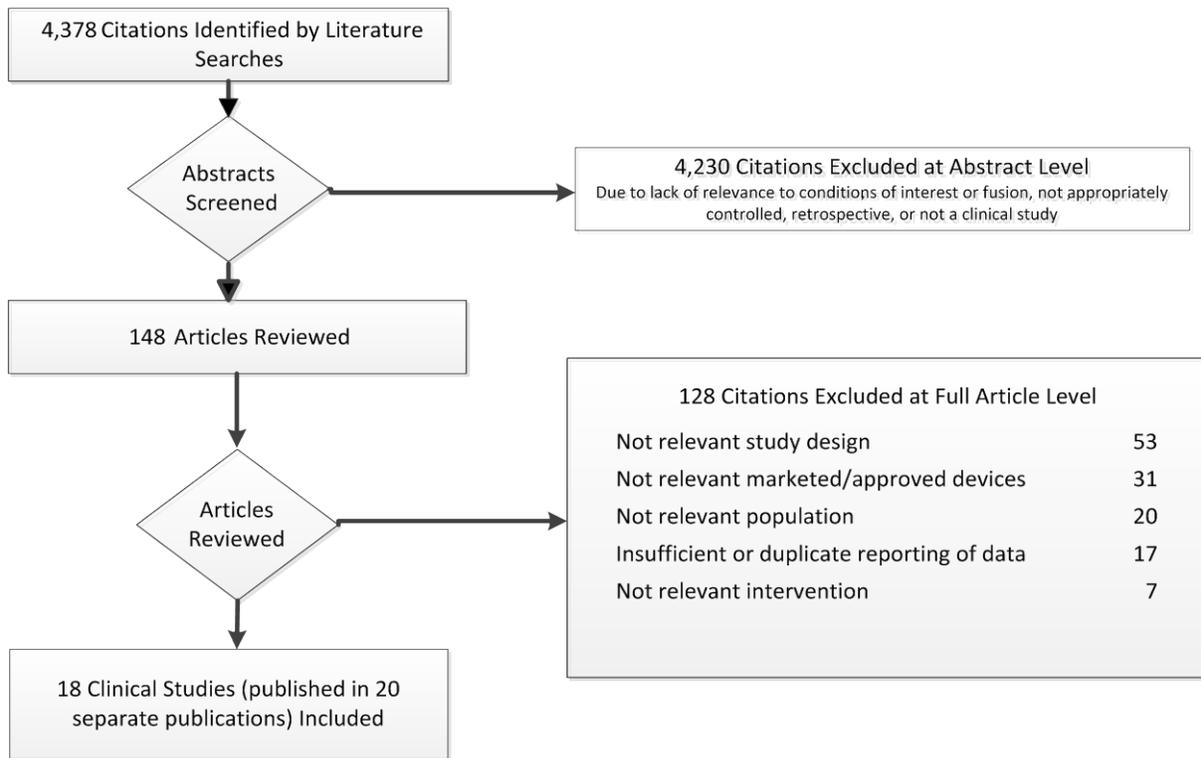


Table 4. Included studies for Key Questions 1–9 (comparative benefit and harm)

Study	KQ(s) Addressed	Fusion Strategy	Comparison
Brox et al. 2009 ¹⁰	1	Posterolateral fusion with transpedicular screws and autologous bone graft	Exercise and cognitive interventions
Brox et al. 2006 ¹¹	1	Posterolateral fusion with transpedicular screws and autologous bone graft	Exercise and cognitive interventions
Fritzell et al. 2001 ¹²	1	Posterolateral fusion, with or without variable screw placement (VSP) with pedicle screws and plate (DePuy Acromed, Raynham MA), or circumferential fusion; all had autologous bone graft	Physical therapy with 1 or more adjunctive treatment
Ohtori et al. 2011 ¹³	1	Anterior discectomy and anterior interbody fusion using iliac bone	Exercise
Weinstein et al. 2007 & 2009 ^{14,15}	3	As per protocol, ⁸⁴ standard posterior decompressive laminectomy with or without bilateral single-level fusion (iliac crest bone grafting with or without posterior pedicle-screw instrumentation) (devices not specified)	As per protocol, ⁸⁴ usual care, recommended to include at least active physical therapy, education or counseling (including instructions for home exercise), and nonsteroidal anti-inflammatory agents if tolerated (devices not specified)
Zigler et al. 2007 ¹⁶	4	Anterior lumbar interbody fusion with femoral ring allograft and posterolateral fusion with autologous iliac crest bone graft with pedicle screws	Artificial intervertebral disc replacement with ProDisc-L (Synthes Spine, West Chester, PA)
Delamarter et al. 2011 ¹⁷	4	Anterior lumbar interbody fusion with femoral ring allograft and posterolateral fusion with autologous iliac crest bone graft with pedicle screws	Artificial intervertebral disc replacement with ProDisc-L (Synthes Spine)
Hallett et al. 2007 ¹⁸	5, 8	Posterolateral instrumented pedicular fusion OR transforaminal lumbar interbody fusion with anterior graft and 2 circular titanium interbody cages and additional graft posterolaterally; both with autologous iliac crest bone graft	Key Question 5: Single or bilateral foraminotomy with nerve root decompression Key Question 8: The 2 fusion groups were compared
Sasso et al. 2003 ²⁴ subgroup of Burkus et al. 2002 ²¹	7	Transperitoneal exposure, anterior lumbar interbody fusion with or without rhBMP	Retroperitoneal exposure, anterior lumbar interbody fusion with or without rhBMP
Burkus et al. 2002 ²¹	7	Posterolateral arthrodesis with rhBMP-2 on collagen sponge with osteoconductive bulking agent (LT-CAGE Lumbar Tapered Fusion Device [Medtronic Sofamor Danek, Memphis, TN] and Infuse Bone Graft [Medtronic Sofamor Danek])	Posterolateral arthrodesis with iliac crest bone graft and instrumentation (Cotrel-Dubousset Horizon Spinal System [Medtronic Sofamor Danek])
Burkus et al. 2002 ²⁰	7	Anterior lumbar interbody fusion with InFUSE Bone Graft (rhBMP-2 on absorbable collagen sponge carrier, Medtronic Sofamor Danek) plus MD-II threaded cortical bone dowel (Regeneration Technologies, Inc. [Alachua, FL])	Anterior lumbar interbody fusion, open, with iliac crest bone graft and MD-II threaded cortical bone dowel, Regeneration Technologies, Inc.
Chung et al. 2003 ²²	7	Anterior laparoscopic approach, using Brantigan carbon cage (DePuy, Raynham, MA) and autologous bone graft	Anterior mini open approach, using Brantigan carbon cage (DePuy) and autologous bone graft

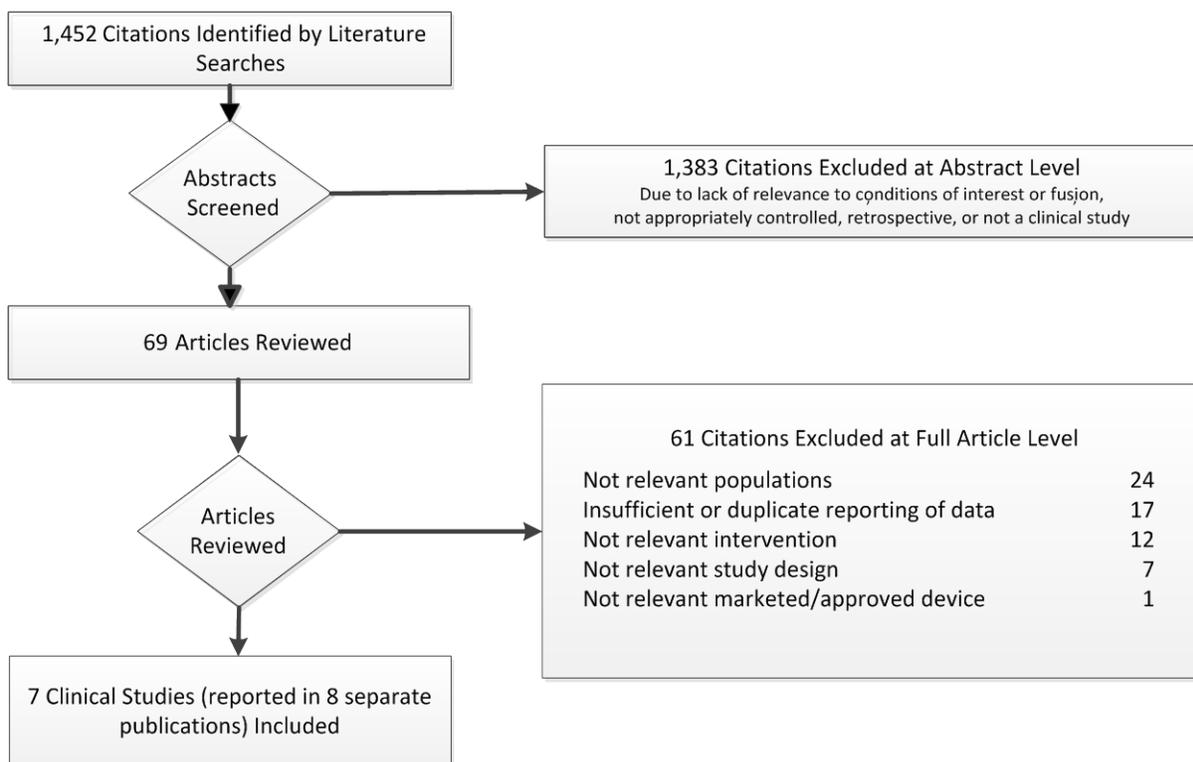
Study	KQ(s) Addressed	Fusion Strategy	Comparison
Dawson et al. 2009 ¹⁹	7	Posterolateral arthrodesis with rhBMP-2 on collagen sponge with osteoconductive bulking agent (INFUSE/MASTERGRAFT Medtronic Sofamor Danek) and Cotrel-Dubousset Horizon Spinal System (Medtronic Sofamor Danek)	Posterolateral arthrodesis with iliac crest bone graft and instrumentation (Cotrel-Dubousset Horizon Spinal System [Medtronic Sofamor Danek])
Fritzell et al. 2002 ²³	7	Posterolateral fusion, with or without variable screw placement (VSP) with pedicle screws and plate (DePuy Acromed, Raynham MA), or circumferential fusion; all had autologous bone graft	Fusion groups compared
Korovessis et al. 2005 ²⁵	8	Fusion with Pro-Osteon 500-R (Interpore Cross International, Irvine, CA) graft material and only local autograft	Fusion using both Pro-Osteon-500-R and iliac crest autograft Fusion using iliac crest autograft only
Abdu et al. 2009 ²⁷	9	Decompression with instrumented posterolateral fusion with pedicle screws (devices not specified)	Decompression with posterolateral fusion (devices not specified)
Fischgrund et al. 1997 ²⁶	9	Decompressive laminectomy and single level bilateral lateral autogenous intertransverse process arthrodesis with transpedicular instrumentation (Pedicle Screws: VSP, Acromed [Cleveland, OH]; Plates: Acromed	Decompressive laminectomy and single level autogenous bilateral intertransverse process arthrodesis without instrumentation
Fernandez-Fairen et al. 2007 ²⁸	9	“Classic bilateral” Iliac crest bone placed bilaterally onto and between transverse processes and facet joints, implemented with granulated bicalcium phosphate and instrumentation (Xia pedicular screw system [Stryker Spine, Allendale, NJ])	“Unilateral pedicle instrumentation” Iliac crest bone placed unilaterally onto and between transverse processes and facet joints, implemented with granulated bicalcium phosphate and instrumentation (Xia pedicular screw system [Stryker Spine])

Table 5. Outcome measurement scales in included studies (Key Questions 1–9)

Outcome	Instrument	Description
Pain	Visual analog scale (VAS)	Patients select their pain level on a line representing a point between 0 (no pain) and 10 or 100 (maximum pain). Higher scores indicate more severe pain.
Function	Oswestry Disability Index (ODI)	The ODI assesses function activities affected by low back pain on a converted scale of 0–100, with higher scores indicating greater disability (<20 minimal disability, 20–40 moderate disability, 40–60 severe disability, 60–80 “crippled,” 80–100 bed bound or “severe magnifier.” ⁸⁵
	Japanese Orthopaedic Association (JOA) back pain questionnaire	This scale measures function on a scale of 0–100, with higher scores indicating better function. ⁸⁶ Posttreatment increases of more than 20 points are clinically important, and scores over 90 are considered satisfactory. ⁸⁶
	General Function Scale (GFS)	This scale measures physical disability in low back pain patients on an overall scale of 0–100, with 100 indicating maximal physical disability. ⁸⁵
	Million Visual Analog Scale	A disability and pain questionnaire for low back pain on which greater scores indicate more impaired function. ⁸⁵
	Low Back Outcome Scale	This questionnaire includes daily activities and yields a calculated score of 0–75, with lower scores indicating greater disability. ⁸⁵
	Roland Morris Disability Scale (RMDS)	This scale was designed to gauge low back pain– related physical disability in daily activities on a scale of 0–24, with 24 indicating severe disability. ⁸⁵
Quality of Life	Short Form-36	The SF-36 measures health-related quality of life on a scale of 0–100, with lower scores indicating lower quality of life. A score of 50 indicates normal health-related quality of life; lower scores indicate impairment. ⁸⁵

We conducted additional searches for Key Question 10 to identify studies in which the primary objective was to examine patient and/or treatment factors associated with patient outcomes following primary fusion surgery. Our searches identified 1,452 potentially relevant studies. Of those, 1,383 were excluded upon title and abstract review for clearly not meeting inclusion criteria (e.g., not relevant to the condition or treatment, not assessing prognostic factors). Of the 69 articles retrieved and reviewed in full, 61 were excluded. Reasons for exclusion are summarized in Figure 3 below and itemized by study in Appendix B. Seven clinical studies (reported in 8 separate publications) remained for inclusion and are listed Table 13 under the findings for Key Question 10.

Figure 3. Study selection for Key Question 10 (qualitative review of prognostic factors)



Description of Included Studies

Information about the patients who participated in the studies is detailed (to the level permitted by original study reporting) in each key question, since different studies address each key question. In general, patients in the studies addressing Key Questions 1–9 (regarding comparative benefit and harm) had moderate-to-severe or severe pain at baseline and functional impairment. They had pain lasting at least 6 months prior to study enrollment, and had tried generally unspecified nonsurgical interventions with sufficient success for at least 6 months. Where reported, a substantial minority had previous decompression. The mean patient age was usually less than 65 years. Most studies were conducted outside the United States.

Most were RCTs and generally well-designed. All but one study earned a moderate risk-of-bias rating. The moderate rating was largely because of lack of concealment of allocation and/or blinding of patients or outcome assessors to treatment received, or not reporting if concealment or blinding took place in the study. While doing so would have been impossible in most of the studies, knowledge of the treatment received and related expectation present a potential source of bias nevertheless. One study earned a high risk-of-bias rating due to high treatment crossover and other limitations that compromised the randomization of the study and introduced selection and other biases.

Key Question 1: Spinal Fusion Compared to Continued Noninvasive Treatment for Painful Lumbar Degenerated Discs

Key Points: Fusion Versus Physical Therapy

- Limited data suggests that fusion leads to greater improvement in back pain relief and function than physical therapy at 2-year followup (strength of evidence: low); however, whether the difference is clinically significant is unclear (the confidence intervals overlap with what is considered a clinically significant difference), and findings at 1 year are insufficient to allow conclusions.
- No other conclusions are possible, because of insufficient evidence or uninformative statistical findings. Serious and minor adverse events occurred in the fusion group that could not occur in a noninvasive intervention group; however, because of insufficient reporting and variation in surgical methods used in the different studies, the incidence of these events cannot be determined conclusively.

Detailed Synthesis

We identified four RCTs that met inclusion criteria for this key question: Ohtori et al. 2011,¹³ Brox et al. 2006,¹¹ Brox et al. 2003,¹⁰ and Fritzell et al. 2001.¹² All the studies compared fusion with physical therapy. Physical therapies varied among studies and included walking and stretching for a total of 1.5 hours twice a day,¹³ education about physical activity despite back pain and various exercises for an average of 3.5 hours per day,^{10,11} and a nonstandardized program including education about physical activity with physical therapies not described and adjunctive nonsurgical treatments such as acupuncture or injection.¹² Differences among studies in the intensity of physical therapy have the potential to influence the degree of relative patient improvement in the comparison of nonsurgical therapy to fusion.

All studies performed fusion at vertebral levels L4/L5 or L5/S1^{11,13} or both.^{11,12} Ohtori et al. performed fusion using an anterior surgical approach,¹³ Fritzell et al. performed posterolateral fusion or circumferential fusion¹², and Brox et al. 2006¹¹ and Brox et al. 2003¹⁰ performed posterolateral fusion. All used iliac crest autograft, and Ohtori et al. used no additional instrumentation, Brox et al. 2003 and Brox et al. 2006 used transpedicular screws, and Fritzell et al. 2001 used variable screw placement with pedicle screws and plates manufactured by DePuy. Ohtori et al. performed adjunctive discectomy while the others did not. Treatment details are reported in Table 21 and Table 22 of Appendix C.

The study by Ohtori and colleagues was conducted in Japan, the studies by Brox and colleagues were conducted in Norway, and the study by Fritzell and colleagues was conducted in Sweden. The authors of all studies had university affiliations; Brox et al. 2003 also reported some authors had National Hospital affiliations.¹⁰ Fritzell et al. reported commercial sponsorship.¹² These general study characteristics are summarized in Table 17 of Appendix C.

All four studies enrolled patients with chronic low back pain and degenerative discs. Most were fairly small, enrolling 41,¹³ 60,¹¹ or 64¹⁰ patients, but Fritzell et al. was larger, enrolling 294 patients.¹² Brox et al. 2006¹¹ enrolled only patients who had a previous surgery for disc herniation at least 1 year prior. Studies enrolled patients with pain lasting at least a year^{10,11} or two prior to surgery.^{12,13} All required imaging confirmation of disc degeneration, from plain

radiographs^{10,11} to plain radiograph or computed tomography (CT) or magnetic resonance imaging (MRI),¹² or MRI, positive discography, and relief from discoblock.¹³ The two studies by Brox and colleagues required functional impairment indicated by a score of at least 30/100 on the Oswestry Disability Index (ODI), and the study by Fritzell and colleagues required that back pain predominate and be severe enough to require sick leave for at least a year, and functional impairment indicated by a score of at least 7/10 on the Function-Working Disability Score. Study inclusion criteria are shown in Table 18 of Appendix C.

Three studies reported the number of potential participants screened and the number enrolled; in each only about half of screened patients met criteria and enrolled.^{10,11,13} The adults were fairly young in all studies, with mean age of 34 years in one study¹³ and 43 years in the rest. Roughly half of each population was women. Race or ethnicity was not reported. In the three studies that reported occupational status, only a minority of participants were working (7 percent,¹¹ 23 percent,¹⁰ 21 percent¹²). The percentage of participants receiving workers compensation was 0 percent in Ohtori et al. (recipients were excluded), 11 percent in Brox et al. 2003, and 21 percent in Fritzell et al. No patients in Ohtori et al. had pending litigation, and 61 percent in Fritzell et al. did (for the studies by Brox and colleagues, this was not reported). These patient characteristics are summarized in Table 19 of Appendix C.

Few details regarding patients' health history were reported. Patients were affected at one vertebral level in two studies,^{11,13} and one or two levels in two studies.^{10,12} In Brox et al. 2006,¹¹ all patients had previous disc surgery (but not fusion), and in Fritzell et al. 2001¹² 19 percent had previous successful disc removal. In the other two studies, no patients had previous surgery. In the studies that reported it, the duration of pain prior to study enrollment lasted a mean of 7 years,¹³ 7.8 years¹² or 11.2 years.¹¹ Duration and nature of nonsurgical treatments prior to study enrollment were not reported. No patients in Ohtori et al. had radicular pain,¹³ but 82 percent in Brox et al. 2006 did.¹¹ Few comorbid diagnoses were reported. Brox et al. 2006¹¹ reported 40 percent had any comorbidity and 67 percent smoked, and Fritzell et al. 2001¹² reported 35 percent had any comorbidity and 43 percent smoked. Patient health characteristics are summarized in Table 20 of Appendix C.

All of these studies merited a moderate risk-of-bias category. None blinded patients to the treatment they received, and outcomes measures were subjective. Most did not conceal allocation or blind outcome assessors to the treatments patients received. More information and itemization by studies is provided in Table 28 of Appendix C.

Because of the differences in the treatments among studies and the differences in the treated patient populations and insufficient numbers of studies to perform statistical investigation of differences among the studies (such as meta-regression or subgroup analyses), we did not perform meta-analysis and instead performed a qualitative review.

1a Patient-Centered Outcomes: Fusion Versus Physical Therapy

Pain: Three of the studies reported back pain levels before surgery and 1 year after in terms of a VAS. All reported baseline and followup data for the patients who remained to report outcomes at 2 years. At 1 year, Ohtori et al.¹³ reported, fusion patients had statistically significant lower mean back pain levels (2.5 ± 0.5) than nonsurgical patients (5.6 ± 1.4) ($p < 0.01$), while neither study from Brox and colleagues found a significant difference between groups and the mean difference was uninformative (please refer to *Methods* for more information about determination of uninformative findings). The study by Ohtori and colleagues differed from those of Brox and colleagues in many ways, including a less-intensive physical therapy program;

anterior noninstrumented fusion instead of posterolateral or circumferential fusion with screws; adjunctive discectomy; and exclusion of patients with radiculopathy, on workers compensation, or with pending litigation. It is impossible to determine which, if any, of these differences led to the different outcome, or if the difference was due to an unidentified factor. Therefore, no conclusions for this outcome are possible.

Two-year back pain data were reported by Ohtori et al. and Fritzell et al. (Fritzell et al. did not report 1-year data). Ohtori indicated that both groups reported lower back pain scores than at 1 year, scores that were low in the fusion group and moderate in the nonsurgical group and remained statistically significantly different. Fritzell reported that mean back pain scores fell from moderate-to-high levels at baseline to moderate levels at 2 years, and that the improvement in the fusion group was statistically significantly greater. The difference in means in Ohtori et al. (3.40, 95% CI: 2.62 to 4.18) was large enough to be clinically important, but in Fritzell et al. it was not (15.10, 95% CI: 8.35 to 21.85). Ohtori et al. used an 11-point scale. A >2.1 point differential would indicate a benefit that is very likely to be clinically significant. Fritzell et al. used a 101-point scale which required an 18 point differential to indicate a benefit that is very likely to be clinically significant (See Table 3).

The findings from these studies suggest that fusion leads to lower pain scores than physical therapies, though perhaps not clinically significantly lower scores, at 2-year followup. However, the fact that two other studies reported no significant difference at 1 year calls into question whether this is a genuine treatment benefit at 2 years, or if it has more to do with the differences in the studies included in the 2-year followup evidence base. For this reason, the strength of evidence was judged to be low. One notable difference is that both Ohtori and Fritzell used a less-intensive physical therapy program than the studies by Brox et al., which has the potential to increase the relative benefit of fusion over nonsurgical therapy.

The two studies by Brox and colleagues measured leg pain in terms of a 101-point VAS at 1-year followup. In Brox et al. 2006,¹¹ median leg pain levels were moderate at both baseline and 1-year followup in both groups and were similar and not statistically different and also uninformative at 1 year. In Brox et al. 2003,¹⁰ baseline scores were moderate, and 1-year scores were lower only in the fusion group, and the difference between groups was statistically significant when adjusted for the difference in change in baseline scores (although the unadjusted mean difference at followup was not, but because of the differences at baseline this is a less useful measurement). Because the two studies had inconsistent results, it is not possible to determine whether fusion is associated with better leg pain control than physical therapy at 1-year followup. At 2-year followup, Fritzell et al. reported, an improvement was noted in mean leg pain in the fusion group and an increase in pain in the physical therapy group. Leg pain was statistically significantly improved in the fusion group compared with the physical therapy group, but a single study provides insufficient evidence to support a conclusion.

Both studies by Brox and colleagues also measured pain medication use at 1 year in terms of daily defined doses. Mean doses changed little, and neither trial found a significant difference between groups at 1-year followup. For Brox et al. 2006,¹¹ the effect size was uninformative; for Brox et al. 2003¹⁰ insufficient data were reported to calculate it. Therefore, no conclusions are possible.

One-year pain data are summarized in Table 23, and 2-year pain data are summarized in Table 24; both tables can be found in Appendix C.

Function: Ohtori et al. and the two studies by Brox et al. measured function at 1 year in terms of the ODI, and at least one other instrument. The general pattern for the ODI was the same as

for back pain at 1 year. Ohtori et al. found improvements in fusion that were statistically and clinically significantly superior to nonsurgical treatment, while both studies from Brox reported improvements in each group, but did not find the difference in improvements to be statistically significant; their results were uninformative. Because these findings are inconsistent, no conclusions can be drawn. Ohtori et al. also found significantly greater improvements for the fusion group in terms of the Japanese Orthopaedic Association Score. Both studies by Brox et al. assessed function in terms of the General Function Score, but only one found a statistically significant difference.¹¹ Findings in terms of these additional functional scales do not elucidate whether there are important differences in outcomes between fusion and nonsurgical treatments because they do not provide consistency or replication.

As was the case for back pain, Ohtori et al. and Fritzell et al. were the only studies that reported 2-year data for function in terms of the ODI. Ohtori reported continued improvements in both groups, but with statistically and clinically significantly greater improvement in the fusion group. In Fritzell et al., the physical therapy group did not improve in average, and the difference in 2-year scores was clinically significant. Ohtori et al. also administered the Japanese Orthopaedic Association Score and Fritzell et al. also administered the General Function Score and Million Visual Analog Score; both reported statistically significantly greater improvement in the fusion group for these instruments as well. The findings from these studies suggest that fusion leads to clinically significantly greater improvements in function than physical therapies. However, the fact that findings were inconclusive at 1 year calls into question whether this is a genuine treatment benefit at 2 years, or if it has more to do with the differences in the studies included in the 2-year followup evidence base. For this reason, the strength of evidence was judged to be low.

Brox et al. 2003 and Brox et al. 2006 reported function in terms of return to work at 1 year. Both found significantly more people in the nonsurgical group were working at followup; however, Brox et al. 2003 reported this finding was not significant once adjusted for baseline function score and gender. Although this evidence suggests that return to work is greater in the nonsurgical group at 1-year followup, further assessment by Brox et al. 2003 render findings for this outcome insufficient to support an evidence-based conclusion. Fritzell et al. 2001 found that significantly more participants returned to work following fusion. This information conflicts with the 1-year findings trend. These return-to-work data do not support any conclusions.

One-year function data are summarized in Table 25 of Appendix C, and 2-year data are summarized in Table 26.

1b Adverse Events: Fusion Versus Physical Therapy

No study reported any adverse events associated with nonsurgical intervention.

Reporting on surgical and postsurgical adverse events was sparse, with the exception of Fritzell et al. Inconsistent reporting and variation in surgical methods used make it difficult to generate a meaningful harms summary. Most fusion-related adverse events afflicted only a few patients in each study; however, some were very serious (e.g., major intraoperative bleeding, deep wound infection). Many of these harms could not occur in a nonsurgically treated population. Clearly, more potential for harm accompanies surgical intervention, but determining the risk is not possible from this evidence base because of insufficient reporting. All reported adverse events are shown in Table 27 of Appendix C.

Strength-of-Evidence Ratings: Fusion Versus Physical Therapy

Strength-of-evidence ratings for Key Question 1 are summarized in Table 6 below.

Table 6. Key Question 1: strength-of-evidence ratings

Comparison	Outcome	Time	Number Studies	Risk of Bias	Consistency	Directness	Precision ¹	Evidence Favors	Rating
Fusion vs. physical and exercise therapies	Pain, back, VAS	1 YR	3	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		2 YR	2	Moderate	Consistent	Direct	Precise	Fusion	Low
	Pain, leg, VAS	1 YR	2	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		2 YR	1	Moderate	Not Applicable	Direct	Precise	Fusion	Insufficient
	Pain, drugs	1 YR	2	Moderate	Unknown	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	1 YR	3	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		2 YR	2	Moderate	Consistent	Direct	Precise	Fusion	Low
	Function, GFS	1 YR	2	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		2 YR	1	Moderate	Not Applicable	Direct	Precise	Fusion	Insufficient
	Function, JOA	1 YR	1	Moderate	Not Applicable	Direct	Precise	Fusion	Insufficient
		2 YR	1	Moderate	Not Applicable	Direct	Precise	Fusion	Insufficient
	Function, Million	2 YR	1	Moderate	Not Applicable	Direct	Precise	Fusion	Insufficient
	Function, return to work	1 YR	2	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		2 YR	1	Moderate	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient

¹ We considered a study to have an imprecise outcome when the intervention that was favored could not be determined.

GFS = General Function Scale; JOA = Japanese Orthopaedic Association back pain questionnaire; Million = Million Visual Analog Scale; ODI = Oswestry Disability Index; VAS = visual analog scale

Applicability: Fusion Versus Physical Therapy

None of the studies in this evidence base were conducted in the United States. Two of the Scandinavian studies administered complex and time-consuming exercise and cognitive programs, which are dissimilar from typical outpatient physical therapy in the United States. In the study conducted in Japan, the nonsurgical treatment was administered for over 2 years, which also may not be representative of U.S. practices. Economic incentives to return to work may also differ from those in the United States.

In three of these studies, patients were highly selected: Only about half evaluated were enrolled. Patients in these studies were young, with mean ages in mid-thirties to mid-forties. Reporting on health history was limited, but the two studies reporting on whether patients had any comorbid diagnoses reported that only a minority did. These findings may be weakly applicable to older patients or those with comorbidities.

Summary: Fusion Versus Physical Therapy

Although four RCTs addressed this key question, the comparative efficacy of fusion and exercise or physical therapies remains somewhat unclear, primarily because of differences in treatments administered across the studies and insufficient evidence for some outcomes and time points. However, the evidence was judged to be minimally sufficient to conclude that fusion was associated with improved back pain and function at 2 years compared with physical therapy, although the clinical significance of these findings is uncertain.

Key Question 2: Spinal Fusion Compared to Continued Noninvasive Treatment for Painful Degenerative Lumbar Spinal Stenosis

We identified no studies that met inclusion criteria and addressed this key question.

Key Question 3: Spinal Fusion Compared to Continued Noninvasive Treatment for Painful Lumbar Degenerative Spondylolisthesis

Key Point: Fusion Versus Noninvasive Treatment

- Because only one study with significant limitations compared fusion versus noninvasive treatment, the evidence is insufficient to support evidence-based conclusions regarding the comparative effectiveness or safety of fusion in adults with low back pain due to degenerative spondylolisthesis.

Detailed Synthesis

We identified one multicenter study reported in two separate publications (one reporting followup data for up to 2 years¹⁴ and the other reporting 4-year followup data¹⁵) that met inclusion criteria and addressed this key question. In this study, Weinstein et al compared fusion to nonsurgical alternatives for the management of degenerative spondylolisthesis with spinal stenosis. The study took place in 13 medical centers located in 11 states across the United States.

The patients in the nonsurgical group (233 patients) received a mix of alternative therapies. Overall, 42 percent of patients in the study received physical therapy, 45 percent epidural steroid injections, 51 percent nonsteroidal antiinflammatory drugs, and 34 percent received opioids. Patients in this group could also have received additional nonsurgical therapies deemed appropriate by their physician.

Patients in the surgical group (368 patients) received standard posterior decompressive laminectomy with or without an additional bilateral single level fusion (autogenous iliac crest bone grafting at the level of the listhesis with or without posterior instrumentation using pedicle screws). Overall, 73 percent of patients received fusion with instrumentation, 21 percent received fusion without instrumentation, and 6 percent of patients received decompression only. Further information about the devices or specific surgical techniques used in the study was not reported. However, according to the study protocol, participating physicians agreed not to use any experimental devices or biologics as part of the surgical procedure. Physicians did have the discretion to differ from the standard surgical protocol if they thought that the patient required a different procedure. Further information about the treatments received in this study is presented in Table 33 and Table 34 of Appendix C.

All patients in this study had a diagnosis of degenerative spondylolisthesis and spinal stenosis with neurologic claudication or radicular leg pain with associated neurological signs. Overall, 85 percent of patients had neurogenic claudication and 77 percent had associated dermatomal pain radiation. Patients had to be at least 18 years of age and they must have experienced symptoms of pain and dysfunction for at least 12 weeks. Patients with spondylolysis and isthmic spondylolisthesis were excluded from the study. The average age of patients in the fusion group was 64.7 years and 68.2 years in the nonsurgical group. Sixty nine percent of patients were women and 84 percent were white. Overall, 68 percent of patients received physical therapy prior to enrollment, 55 percent received epidural injections, 25 percent chiropractic treatment, 63 percent anti-inflammatory agents, and 30 percent opioid agents. A total of 41 patients (34 in the fusion group and 7 in the nonsurgical group) reported being on workers compensation. Comorbid conditions were not reported in the study. Further information about general study information, patient inclusion and exclusion criteria, and patient characteristics are presented in Table 29 through Table 32 of Appendix C.

This study was designed as a randomized controlled trial in which patients meeting inclusion criteria were assigned to receive fusion or nonsurgical care. Patients who chose not to be randomized could still participate in the study as part of an observational cohort. Patients in this cohort could choose to receive surgery or be a part of the nonsurgical observational group. Overall, 607 of 892 eligible patients (285 patients declined to participate for reasons not reported) were enrolled in the study (304 in the randomized cohort and 303 in the observational cohort). In the randomized cohort, 159 were randomized to receive surgery and 145 were randomized to the nonsurgical group. Of those assigned to surgery, 64 percent received it by 2 years and 66 percent by 4 years. Of those assigned to the nonsurgical group, 49 percent received surgery by 2 years and 54 percent by 4 years. In the observational cohort, 173 patients initially chose surgery and 130 chose nonsurgical care. Of those who initially chose surgery, 97% received it by 2 years with no additional surgeries at 4 years. Of those who initially chose nonsurgical care, 25 percent underwent surgery by 2 years and 33 percent by 4 years.

Due to the substantial amount of crossover in this study, the authors combined patients from both the randomized cohort and the observational cohort to create an as-treated study population that included at baseline 368 patients who underwent surgery and 233 patients who received

nonsurgical care. For the most part, this population was used to compare patient outcomes post-surgery. The authors did, however, perform an intention-to-treat analysis of the randomized cohort for the primary outcomes of the study.

Because the study violated randomization by combining randomized patients with nonrandomized patients in the as-treated cohort, the study was considered to have a high risk-of-bias. This is mainly due to significant differences on a number of important factors at baseline between the combined surgical group and the combined nonsurgical group. These factors included age, workers' compensation status, pain, physical function, disability, discomfort with symptoms, and preference for treatment. Overall, patients in the combined randomized and observational surgical group were younger and more likely to be on workers' compensation than patients in the combined nonsurgical group. They also had worse pain, function, disability, and symptoms than patients in the nonsurgical group. As a result, the authors adjusted for these factors in subsequent analyses of the study data. The intention-to-treat analysis is also potentially biased due to the high rate of treatment crossover between groups in this study. Under these circumstances, an intention-to-treat analysis generally underestimates between-treatment differences in patient outcomes. The risk-of-bias assessment is itemized in Table 47 of Appendix C.

3a Perioperative Outcomes: Fusion Versus Noninvasive Treatment

No perioperative outcomes were reported.

3b Patient-Centered Outcomes: Fusion Versus Noninvasive Treatment

As-treated change scores and treatment effects for pain and function were reported for the combined study population. Scores were adjusted to control for differences between the combined surgical and nonsurgical groups. The following variables were included in the adjusted models: age, gender, work status, depression, osteoporosis, joint problems, duration of current symptoms, reflex deficit, number of moderate or severe stenotic levels, baseline scores (for the SF-36, Oswestry Disability Index, and Stenosis Bothersomeness Index), and the center where patients were treated. The authors also reported results of the intention-to-treat analysis of the randomized cohort for the primary outcomes of pain and function. The results were reported as differences in mean changes from baseline between treatment groups including all time periods (3 months, 1 year, 2 years, and 4 years).

Pain: Pain was measured using the SF-36 subscale for physical pain. Higher scores on this scale indicate less severe symptoms. Results of the as-treated analysis indicated statistically significant treatment effects in favor of the combined surgical group across all time periods (mean difference at 2 years 18.1, 95% confidence interval (CI) 14.5 to 21.7, $p < 0.001$ and mean difference at 4 years 15.3, 95% CI 11.0 to 19.7, $p < 0.001$). However, the combined surgical group demonstrated little change in pain scores from baseline to 2 years and 4 years (baseline score 29.3 ± 16.8 , 2-year score 29.9 ± 1.2 , and 4-year score 31.1 ± 1.1). The results of the intention-to-treat analysis of the randomized cohort showed no statistically significant difference between treatment groups for the SF-36 physical pain scale (mean change from baseline to 2 years 1.5, 95% CI -4.2 to 7.3, $p = 0.52$ and 4 years -2.0, 95% CI -8.6 to 4.6, $p = 0.56$), but as noted earlier the high rate of treatment crossover between groups would tend to mask any differences between treatments in an intention-to-treat analysis. See Table 35 through Table 39 for results for all time periods.

Function: Function was measured using the SF-36 subscale for physical function (higher scores indicate less severe symptoms) and the ODI (lower scores indicate less severe symptoms). Results of the as-treated group indicated statistically significant treatment effects in favor of the combined surgical group across all time periods for both scales (mean difference on SF-36 function at 2 years 18.3, 95% CI 14.6 to 21.9, $p < 0.001$ and at 4 years 18.9, 95% CI 14.8 to 23.0, $p < 0.001$; mean difference in ODI at 2 years -16.7, 95% CI -19.5 to -13.9, $p < 0.001$ and at 4 years -14.3, 95% CI -17.5 to -11.1, $p < 0.001$). For ODI, an 8.2 to 20 point improvement would indicate a change that is potentially clinically significant. Function among the combined surgical group appeared to improve from baseline to 2 years and 4 years as evidenced by change in scores on the ODI, but not on the SF-36 physical function subscale. Again, the intention-to-treat analysis showed no statistically significant difference between treatment groups for either scale at 2 years (SF-36 function 1.9, 95% CI -3.7 to 7.5, $p = 0.71$ at 4 years -3.1, 95% CI -9.2 to 3.0, $p = 0.32$; ODI at 2 years 2.2, 95% CI -2.3 to 6.8, $p = 0.68$ and at 4 years 4.1, 95% CI -0.8 to 9.1, $p = 0.1$). See Table 7 for a description of strength-of-evidence ratings. See Table 40 through Table 44 for results for all time periods.

3c Adverse Events: Fusion Versus Noninvasive Treatment

The most common surgical complication was a dural tear (11 percent or 41 of 387 patients). The reoperation rate for recurrent stenosis or spondylolisthesis was 0.6 percent at 1 year, 3.0 percent at 2 years, and 5.0 percent at 4 years. Overall, at 4 years 7 deaths occurred in the nonsurgical group compared to 17 deaths in the fusion group (hazard ratio based on proportional-hazards model adjusted for age was 1.9, 95% CI 0.76 to 4.6, $p = 0.17$). According to the authors, all deaths were independently reviewed, and 18 were judged to be not related to treatment. Four deaths occurring between 621 and 1,379 days following surgery were considered to be of unknown causes. Two deaths, both in the surgical group, were considered to be potentially related to treatment—one patient died of respiratory distress 32 days post surgery and the other died of sepsis 82 days post surgery. See Table 45 and Table 46 for a complete listing of adverse events reported in the study.

Strength-of-Evidence Ratings: Fusion Versus Noninvasive Treatment

The strength-of-evidence ratings for Key Question 3 are summarized in Table 7, below.

Table 7. Key Question 3: strength-of-evidence ratings

Comparison	Outcome	Time	Number of Studies	Risk of Bias	Consistency	Directness	Precision ¹	Evidence Favors	Rating
Fusion vs. decompression	Pain, SF-36	2 and 4 YR	1	High	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient
	Function, SF-36	2 and 4 YR	1	High	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient
	Oswestry Disability Index	2 and 4 YR	1	High	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient

¹We considered this study to be imprecise for the presented outcomes due to inconsistencies between the findings of the as-treated and intention-to-treat analysis. The as-treated analysis indicated a statistically significant treatment effect in favor of the surgical group for pain and function outcomes, while the intention-to-treat analysis suggested there was no statistically significant differences between treatment groups on any of the primary outcomes. As indicated in previous commentary on this study, the high number of crossovers in the intention-to-treat analysis in the randomized trial is likely to underestimate the benefit of surgery, while the measured and unmeasured differences in the combined as-treated group may overestimate the benefit of surgery.⁸⁷

SF-36 = Short Form-36

Applicability

The applicability of the Weinstein et al. study is difficult to assess due to limited reporting of patients' health history and heterogeneity of the treatment interventions. The authors did not report if patients in the study had any comorbid conditions or a history of smoking. They reported only limited information about the type and duration of previous nonsurgical treatments.

Similarly, the authors reported limited information about the surgical intervention, indicating only that the surgical protocol consisted of standard posterior decompressive laminectomy with or without bilateral single-level fusion. Other than reporting the number of patients that underwent fusion with or without instrumentation, no further details were provided about the devices or techniques used in the fusion procedure. The authors did report that physicians did have the discretion to differ from the standard surgical protocol if they thought that the patient required a different procedure. This makes it difficult to attribute treatment success to any particular technique or approach.

Finally, patients in the nonsurgical group received a range of interventions that at minimum included physical therapy, education/counseling with home exercise instruction, and a non-steroidal anti-inflammatory drug. As a pragmatic trial, however, the heterogeneity of treatments in the surgical and non-surgical groups is probably representative of the range of treatments that similar patients would likely receive in clinical practice.

Summary: Fusion Versus Noninvasive Treatment

Presently, the data comparing fusion to noninvasive treatment for the management of degenerative spondylolisthesis with spinal stenosis is sparse and inconsistent. We identified one study that met inclusion criteria that compared the outcomes of fusion to nonsurgical care for this population. The study had a number of limitations that caused it to be of high risk-of-bias. These limitations include nonadherence to the randomized treatment, lack of baseline comparability among the as-treated study groups, and heterogeneity of the treatment interventions, particularly in the noninvasive control group. The strength-of-evidence rating for the study was considered insufficient because it was a single study with inconsistencies in the results of the as-treated and intention-to-treat analysis. The as-treated analysis indicated a statistically significant treatment effect in favor of the surgical group for pain and function outcomes, while the intent-to-treat analysis suggested there was no statistically significant differences between treatment groups on any of the primary outcomes. As indicated in previous commentary on this study, the high number of crossovers in the intention-to-treat analysis in the randomized trial is likely to underestimate the benefit of surgery, while the measured and unmeasured differences in the combined as-treated group may overestimate the benefit of surgery.⁸⁷ Therefore, the true difference in effect of fusion versus noninvasive treatment on clinical outcomes remains unclear in patients with degenerative spondylolisthesis.

Key Question 4: Spinal Fusion Compared to Other Invasive Procedures for Painful Lumbar Degenerated Disc(s)

Key Points: Fusion Versus Arthroplasty

- Limited evidence suggests that shorter surgical time, less blood loss, and shorter inpatient stays are associated with arthroplasty, and that disc recipients have better ODI functions

scores at 6 weeks postsurgery (strength of evidence: low). The difference in ODI functions were not observed at later followup times.

- For all other outcomes the data were insufficient to support any conclusions, typically because the mean difference was uninformative (i.e., captured the possibility of either being superior or equivalent) or because only one study addressed that outcome or duration of followup.

Detailed Synthesis

Two multicenter RCTs from the same research group in the United States that compared circumferential fusion with femoral ring allograft, pedicle screws, and autologous iliac crest bone graft with arthroplasty using the second-generation ProDisc-L (Synthes Spine, West Chester PA) met inclusion criteria. The surgical approach was anterior for arthroplasty and anterior and posterior for fusion. For clarity, we refer to them as the single-level study (in which only one vertebral level was treated per patient)^{16,88,89} and the two-level study (in which each patient was treated at two vertebral levels).^{17,90} In the single-level study, most patients were treated at L4/L5. In the two-level study, most patients were treated from L4 to S1. Study description information, including general characteristics, study inclusion criteria, and patient characteristics, is presented in Table 48 through Table 52 of Appendix C.

Both studies were investigational device exemption (IDE) studies for FDA for ProDisc-L. Author affiliations were with universities and spine specialty centers. The authors reported receiving direct funding for the two-level but not the single-level study, and they reported financial relationships with commercial interests in both studies.

Inclusion criteria were similar in both studies. Each enrolled adults with back and/or leg pain with evidence of disc degeneration shown on imaging and functional impairment indicated by an ODI score of at least 40. Patients in both studies had severe pain at baseline (a mean of about 75 on a 0–100-point scale). The single-level study required the adults to be between the ages of 18 and 60 years, and that they had tried nonsurgical treatments for at least 6 months without success. Both studies excluded patients with complicating comorbidities such as other spinal problems, metabolic bone disease, immune disease, or obesity; neither indicated how many patients were screened out.

In the single-level study, 236 patients were a mean of 40 years old, and 51 percent were women. Eighty-one percent were Caucasian, 4 percent African American, 12 percent Hispanic, and 1 percent Asian American. In the two-level study, 256 patients were a mean of 42 years old, 43 percent were women, and ethnic backgrounds were not reported. Neither study reported occupational status, workers compensation, or pending litigation. In the single-level study, the duration of symptoms and conservative treatment efforts were not reported (although to meet inclusion criteria, patients had to have tried conservative treatments for at least 6 months). Previous treatments included narcotics (81 percent), discectomy (16 percent), intradiscal electrothermal therapy (10 percent), laminectomy (8 percent), and laminotomy (3 percent). All patients had radicular pain. The only comorbidity reported was smoking (25 percent). In the two-level study, 8 percent of patients had pain starting 6 months to 1 year before enrollment, and 92 percent had pain starting at least a year prior. Previous nonsurgical treatments included injection (76 percent), physical therapy (83 percent), wearing a corset or brace (41 percent), and chiropractic care (37 percent); the durations for which these treatments were tried were not reported. Forty-one percent of patients had tried previous surgery, including discectomy

(19 percent), intradiscal electrothermal therapy (10 percent), laminectomy (17 percent), and laminotomy (3 percent). The only comorbidity reported was smoking (29 percent).

These studies merited a moderate risk-of-bias rating because neither patients nor outcome assessors were blinded to treatment received, the primary outcome measures were subjective, and for some outcomes and durations of followup, fewer than 85 percent of treated patients reported outcomes. This assessment is itemized in Table 76 of Appendix C.

Because one study assessed only single-level treatment of single-level disc degeneration and the other assessed two-level treatment of patients with at least two degenerated discs, these studies are too clinically dissimilar to combine in meta-analysis and expect to gain a meaningful single summary statistic to represent both studies. Therefore, the assessment for this key question is qualitative.

4a Perioperative Outcomes: Fusion Versus Arthroplasty

Surgical time: Both studies reported statistically significantly shorter surgery in the arthroplasty group. Surgery was a mean of 108 minutes shorter in the single-level study (mean \pm standard deviation 121 \pm 59 minutes for arthroplasty vs. 229 \pm 76 minutes for fusion) and a mean of 113 minutes shorter in the two-level study (160 \pm 73 minutes for arthroplasty vs. 273 \pm 82 minutes for fusion). Because these studies are both informative and have consistent findings, we conclude that arthroplasty requires less surgical time than fusion (strength of evidence: low).

Blood loss: Both studies reported statistically significantly less blood was lost in the arthroplasty group. Blood loss was a mean of 261 cc (about 8.7 fluid ounces) less in the single level study (204 \pm 231 cc for arthroplasty vs. 465 \pm 440 cc for fusion) and 171 mL (about 5.7 fluid ounces) less in the two-level study (398 \pm 451 mL for arthroplasty vs. 569 \pm 467 mL for fusion). Because these studies are both informative and have consistent findings, we conclude that arthroplasty results in less blood loss than fusion (strength of evidence: low); however, as the mean differences were much less than a typical blood donation (about 16 fluid ounces), the difference is unlikely to be clinically important.

Length of inpatient stay: Both studies reported statistically significantly shorter hospital stays for the arthroplasty group. In both studies, the stay was about a day less (mean \pm standard deviation single-level study 3.5 \pm 1.29 days for arthroplasty, 4.4 \pm 1.54 days for fusion; two-level study 3.8 \pm 1.5 days for arthroplasty vs. 5.0 \pm 1.9 days for fusion). Because these studies are both informative and have consistent findings, we conclude that arthroplasty is associated with a shorter duration of inpatient stay (strength of evidence: low).

All perioperative outcomes data are shown in Table 53 of Appendix C.

4b Patient-Centered Outcomes: Fusion Versus Arthroplasty

Pain: Mean pain VAS scores decreased from severe at baseline (about 75 out of 100 in all groups) to moderate at 6 weeks, 3 months, 6 months, 1 year, and 2 years in both studies (between 32 and 43 on the VAS in any group). Pain scores at followup were generally similar between treatment groups, and were only significantly different at 3 months in the single-level study. Neither study's findings suggest that pain levels are significantly different overall between groups. These data are shown in Table 54 and Table 55 of Appendix C. The single-level study also reported 5-year outcome data, at which time pain levels remained moderate and not significantly different between groups. This suggests that improvements are maintained, but a single study provides insufficient evidence to support any conclusions. These data are shown in

Table 60 of Appendix C. Furthermore, the mean differences for each of these time points are uninformative, so no evidence-based conclusions can be drawn.

Pain was also assessed in terms of proportion of patients taking narcotic pain medication. In the single-level study, similar proportions of patients took narcotic pain medication at baseline (76 percent fusion, 84 percent arthroplasty), 2 years (43 percent fusion, 45 percent arthroplasty), and 5 years (40 percent fusion, 38 percent arthroplasty); none of these differences were statistically significant. These data are shown in Table 59 and Table 60 of Appendix C. In the two-level study, the difference in proportion of patients taking narcotic pain medication was not significantly different at baseline (64 percent fusion, 69 percent arthroplasty), but statistically significantly more fusion-treated patients took narcotics at 2-year followup than arthroplasty-treated patients (59 percent vs. 36 percent, respectively). Because the findings from the two studies are inconsistent, no conclusions are possible regarding relative narcotic intake. These data are shown in Table 59 of Appendix C.

Function: Both the single and two-level treatment studies measured function using the ODI. Scores were similar in both groups and not statistically different at baseline for both studies. In the single-level study, ODI scores decreased from levels indicating “crippled” level of disability to “severe disability” at 6 weeks. The ODI still indicated severe disability at 3 months and 6 months in both studies, and severe disability at 18 months in the two-level study. All other scores indicated “moderate disability,”⁸⁵ including 2-year scores in both studies and 5-year scores in the single-level study. However, the differences between the groups’ scores were not statistically significantly different at 6-month, 1-year, 18-month, 2-year, or 5-year followups. Mean ODI scores were significantly lower at 6 weeks in the arthroplasty group in both studies. However, at a difference in means of 8.30 (95% CI, 3.12 to 13.45) in the single-level study and 5.90 (95% CI, 0.64 to 11.16) in the two-level study, this difference is unlikely to be clinically significant. In any event, the benefit would be only transient. Although the mean differences show the same direction of effect for each of the other time points, the variability in statistical significance and imprecision in CIs between the two studies means the evidence is inconsistent or uninformative, so no evidence-based conclusions can be drawn.

The two-level study also assessed function in terms of the proportion of patients working. The difference between groups was not statistically significantly different at either baseline (83 percent fusion vs. 79 percent arthroplasty) or 2-year followup (86 percent fusion, 80 percent arthroplasty). No conclusions are possible regarding return to work because of the imprecision in the effect estimate. Findings related to function are presented in Table 61 through Table 67 of Appendix C.

Quality of life: The two-level study assessed quality of life in terms of the SF-36 physical component score and mental component score. The physical and mental component scores were similar at baseline and clinically significantly below the normative score of 50 in both groups. The physical component score was statistically significantly higher in the arthroplasty group at 6 weeks, 3 months, 6 months, 18 months, and 2 years (all measured time points besides 1 year). Mean scores in neither group met or exceeded the normative standard of 50 points; quality of life remained impaired in both groups. The mental component scores were not statistically significantly different at any time point. The single-level study reported physical component scores at 5-year followup, and the scores in both groups had improved from “crippled” to “severe disability” and were similar and not statistically significantly different. This study suggests a small and likely nonclinically significant advantage for patients receiving arthroplasty compared

with those receiving fusion, but a single study provides an insufficient amount of evidence to draw conclusions. These data are summarized in Table 68 through Table 74 of Appendix C.

4c Adverse Events: Fusion Versus Arthroplasty

Few adverse events were reported in either study. Two (3 percent) fusion patients in each study had greater than 1,500 mL blood loss, as did one (1 percent) patient who had two-level arthroplasty. Three (4 percent) fusion patients in the two-level study and two (3 percent) in the single-level study had dural tear, as did one patient (0.6 percent) who had two-level arthroplasty. Two patients in each group of the two-level study had deep venous thrombosis (1 percent fusion, 3 percent arthroplasty). Two (3 percent) fusion patients in the single-level study had a wound infection. After surgery, one (1 percent) fusion patient and two (3 percent) arthroplasty patients reported retrograde ejaculation. In the two-level study, a total of six (8 percent) patients in the fusion group and four (2 percent) in the arthroplasty group had reoperation for any reason at up to 2 years. These data are shown in Table 75 of Appendix C.

Strength-of-Evidence Ratings: Fusion Versus Arthroplasty

The strength-of-evidence ratings for Key Question 4 are summarized in Table 8, below.

Table 8. Key Question 4: strength-of-evidence ratings

Comparison	Outcome	Time	Number of Studies	Risk of Bias	Consistency	Directness	Precision ¹	Evidence Favors	Strength of Evidence Rating
Fusion vs. artificial intervertebral disc	Surgical time	Peri-operative	2	Moderate	Consistent	Direct	Precise	Disc	Low
	Blood loss	Peri-operative	2	Moderate	Consistent	Direct	Precise	Disc	Low
	Inpatient stay	Peri-operative	2	Moderate	Consistent	Direct	Precise	Disc	Low
	Pain, VAS	6 WK – 2 YR	2	Moderate	Consistent	Direct	Imprecise	Inconclusive	Insufficient
		5 YR	1	Moderate	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient
	Pain, drugs	2 YR	2	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		5 YR	1	Moderate	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	6 WK	2	Moderate	Consistent	Direct	Precise	Disc	Low
		3 MO – 2 YR	2	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		5 YR	1	Moderate	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient
	Working	2 YR	2	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Quality of Life, SF-36	6 WK – 5 YR	1	Moderate	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient

¹ We considered a study to have an imprecise outcome when the intervention that was favored could not be determined.

ODI = Oswestry Disability Index; SF-36 = Short Form 36 questions; VAS = visual analog scale

Applicability

Both of these studies were multicenter studies in the United States, conducted under IDE status for FDA. What proportion of candidates was screened out was not reported; however, the strict inclusion criteria used in the studies appear similar to the FDA-approved indications. Both of these studies address the use of the ProDisc-L; applicability to other discs is unclear.

On average, enrolled patients were young, in their early forties. No information on their employment history was reported. Other than about a quarter of patients reporting smoking, no comorbidities were reported. However, because the primary purported advantage of an artificial disc is to retain range of motion and activity, these patients may not be unrepresentative of candidates for disc replacement in general.

Summary: Fusion Versus Arthroplasty

Perioperative outcomes suggest artificial intervertebral disc replacement may be less taxing on patients than fusion, but replication of these findings by researchers not affiliated with the manufacturer would be reassuring. The studies were generally inconsistent in finding significant differences in efficacy between artificial disc and fusion. Further, although the reported differences were not large, the CIs of the mean difference were too wide to conclude equivalence.

Key Question 5: Spinal Fusion Compared to Other Invasive Procedures for Painful Degenerative Lumbar Spinal Stenosis

Key Point: Fusion Versus Decompression

- Because only one study addressed fusion versus decompression, the evidence is insufficient to support evidence-based conclusions regarding the comparative effectiveness or safety of fusion and decompression in adults with low back pain due to stenosis with degenerative disc.

Detailed Synthesis

One RCT (Hallett et al. 2007)¹⁸ addressed this key question. The study compared single or bilateral foraminotomy with nerve root decompression to posterolateral instrumented pedicular fusion using a Moss-Miami pedicular fusion instrument (DePuy Ltd, Warsaw, IN) and transforaminal lumbar interbody fusion using an unspecified titanium interbody cage. Autologous iliac crest bone graft was used in both fusion groups. Most patients were treated at level L4/L5.

Hallett et al. enrolled 44 patients between 1998 and 2001 in a university spine center in Edinburgh, Scotland, and reported corporate or industry funding of their study. Disease in enrolled patients was diagnosed as foraminal stenosis with single-level disc degeneration evidenced by plain radiographs and MRI and visible evidence of nerve root compromise and radicular pain, and symptoms lasting for at least 5 years before study enrollment. Study participants had severe pain on average (7.75 on a scale of 0–10). All patients tried conservative treatments for at least 3 months before enrollment (most tried physical therapy and 1 or more of acupuncture, chiropractic, or massage), but the actual duration of nonsurgical treatments was not

reported. Patients with the following conditions were excluded: at least grade 2 degenerative spondylolisthesis at level to be treated or one adjacent; at least 1 cm vertebral translocation; at least 50 percent disc space in the level above or below the treated level, and cancer.

Of 48 potential participants, 44 (92 percent) were enrolled in the study. Patients had an average age of 57 years, 45 percent were women, and all were white. Thirty percent were retired, 23 percent were housewives, 27 percent were on sick leave, and the other patients worked in heavy manual or light industrial jobs. Two percent had pending workers compensation claims, and the proportion with pending litigation was not reported. Study description information is tabled in Table 77 through Table 81 of Appendix C.

The study merited a moderate risk-of-bias rating because allocation was not concealed, patients and outcomes assessors were not blinded, and the primary outcomes of interest were subjective. Five-year outcomes were reported but not reviewed because of insufficient numbers of patients per group (fewer than 10). Risk-of-bias assessment is itemized in Table 86 of Appendix C.

5a Perioperative Outcomes: Fusion Versus Decompression

No perioperative outcomes were reported.

5b Patient-Centered Outcomes: Fusion Versus Decompression

Pain, function, and quality-of-life mean scores and whether differences between groups were statistically significant were reported at baseline and 2-year followup. The proportion of patients for whom followup data was reported was high and similar across groups (13/14 decompression, 15/16 posterolateral interbody fusion, 12/14 transforaminal interbody fusion). Standard deviations or other measure of variance were not reported.

Pain: All three treatment groups experienced pain reduction on an 11-point VAS back pain scale from severe presurgical pain (mean 7.75 for all groups) to moderate levels at 2-year followup (mean 4.75 decompression, 4.5 posterolateral interbody fusion; 4 for transforaminal interbody fusion). Improvement was statistically significant in all three groups, as indicated by the authors. These VAS scores were similar and not significantly different at followup. These data are shown in Table 82 of Appendix C.

Function: Function was reported in terms of both the Low Back Outcomes Scale and the Roland Morris Disability Scale. Mean Low Back Outcome Scale scores improved from baseline to 2 years in all three groups (from 14 to 25 in decompression group; from 20 to 35 in posterolateral interbody fusion group; from 19 to 26 in transforaminal interbody fusion group). The improvement was statistically significant only in the decompression and posterolateral interbody fusion groups. The Roland Morris Disability Score improved from baseline in all three groups (from 16 to 13 decompression, from 13 to 9 posterolateral interbody fusion, from 14 to 12 transforaminal interbody fusion). No measure of variance was reported for either outcome. The authors reported the improvement was statistically significant only in the decompression group. The authors reported the difference between decompression and either fusion group was not statistically significantly different for either outcome. These data are shown in Table 83 of Appendix C.

Quality of life: Quality of life was reported in terms of the SF-36 physical functioning scale. Mean scores increased from baseline to 2 years in all three groups (from 34 to 45 in the decompression group; from 25 to 45 in the posterolateral interbody fusion group; from 31 to 35 in the transforaminal interbody fusion group). No measure of variance was reported. The authors

reported that the pre-post change per group was not statistically significant. The authors reported the difference between decompression and either fusion group was not statistically significantly different. These data are shown in Table 84 of Appendix C.

5c Adverse Events: Fusion Versus Decompression

One adverse event was reported: the need for reoperation. One patient in each fusion group underwent decompressive laminectomy (at 7 and 23 months, respectively), and one patient in the decompression group underwent instrumented lateral mass fusion. This information is shown in Table 85 of Appendix C.

Strength-of-Evidence Ratings: Fusion Versus Decompression

The strength-of-evidence ratings for Key Question 5 are summarized in Table 9, below.

Table 9. Key Question 5: strength-of-evidence ratings

Comparison	Outcome	Time	Number of Studies	Risk of Bias	Consistency	Directness	Precision ¹	Evidence Favors	Rating
Fusion vs. decompression	Pain, back	2 YR	1	Moderate	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient
	Function, LBOS	2 YR	1	Moderate	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient
	Function, RMDS	2 YR	1	Moderate	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient
	Quality of Life, SF-36	2 YR	1	Moderate	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient

¹ We considered a study to have an imprecise outcome when the intervention that was favored could not be determined.

LBOS = Low Back Outcomes Scale; RMDS = Roland Morris Disability Scale; SF-36 = Short Form-36

Applicability: Fusion Versus Decompression

Patients in this Scottish study had foraminal stenosis with single-level disc degeneration with radicular pain and symptoms lasting at least 5 years. No information on previous nonsurgical treatment efforts was reported. With a mean age of 57 years, patients are older than in the studies of patients with degenerated discs alone, and may be more similar to a Medicare enrolled population. Most patients were not working.

Summary: Fusion Versus Decompression

Presently, there is simply too little evidence available to support evidence-based conclusions for patient-centered or other outcomes, because only one small, single-center study meeting inclusion criteria was identified, and it has reporting limitations.

Key Question 6: Spinal Fusion Compared to Other Invasive Procedures for Painful Degenerative Lumbar Spondylolisthesis

We identified no studies that met inclusion criteria and addressed this key question.

Key Question 7: Spinal Fusion Approaches (e.g., Anterior, Posterior, Combined) and Techniques (e.g., Instrumentation, Graft Material) Compared to One Another for Painful Degenerated Lumbar Disc(s)

Key Points: Fusion With rhBMP-2 Versus Autogenous Bone Graft

- rhBMP-2 is associated with less blood loss than autogenous bone graft, while surgery time and length of hospital stay do not differ substantially for these products in adults undergoing fusion for low back pain due to degenerated disc(s) (strength of evidence: low).
- For all other outcomes, the data were insufficient to support any conclusions for fusion with rhBMP-2 and fusion with autogenous bone graft, mainly because of inconsistencies in the studies' findings and insufficient reporting of data, or because only one study addressed that outcome or duration of followup.

Detailed Synthesis

Three RCTs that addressed this comparison were all multicenter trials conducted at U.S. spine specialty centers.¹⁹⁻²¹ One study was an FDA IDE pilot trial,¹⁹ while devices in another were part of an ongoing FDA-approved IDE or corresponding national protocol.²⁰ Both these trials received manufacturer funding.^{19,20} One study, authored by Burkus et al. 2002, enrolled patients in 1998 with degenerated discs with low back pain and/or leg pain or sciatica.²⁰ Another study authored by Burkus et al. 2002 enrolled patients from 1998 to 1999 with degenerated discs with disabling low back and/or leg pain.²¹ Dawson et al. 2009¹⁹ enrolled patients from 2003 to 2004 with degenerated discs (L1–S1), low back pain, and radicular pain. Patients enrolled in all

three studies were required to undergo at least 6 months of nonoperative treatments. Mean age of patients was 46 years in two studies^{19,20} and 43 years in the other study.²¹ The percentage of women enrolled ranged from 48 percent to 59 percent. All three studies reported on percentage of patients receiving workers compensation or involved in litigation. See Table 87 through Table 90 of Appendix C for details on study design and patient characteristics.

In one study, 25 patients received rhBMP-2 (12 mg total dose) on collagen sponge with osteoconductive bulking agent while 21 patients received autogenous iliac crest bone graft and instrumentation for single-level posterolateral fusion.¹⁹ In a similar-sized study, patients underwent anterior lumbar interbody fusion (ALIF) and received threaded cortical allograft dowels with either rhBMP-2 (8–12 mL) or autogenous iliac crest bone graft.²⁰ One larger study enrolled 279 patients that underwent single-level ALIF with a tapered fusion device and either rhBMP-2 or autogenous bone graft.²¹

Medtronic Sofamor Danek (Memphis, TN) provided the following products: INFUSE/MASTERGRAFT, InFuse Bone Graft, LT-Cage Lumbar Tapered Fusion Device, and the Cotrel-Dubousset Horizon Spinal System. Two MD-II threaded cortical bone dowels per patient used in one study²⁰ were provided by Regeneration Technologies, Inc. (Alachua, FL). See Table 91 of Appendix C for details of treatment characteristics.

All three studies received a moderate risk-of-bias rating. The authors did not conceal allocation or report patient blinding to treatment allocation, and the primary outcomes were not objective. Full assessment is itemized in Table 114 of Appendix C. Of additional concern, the protocol described by Dawson et al.¹⁹ reported discarding local bone without using it for graft augmentation. According to a recent review by Carragee et al., usual practice is not to discard local bone but to use it for graft augmentation. Not following this usual practice may increase the risk of poor quality fusion and non-union among control patients, which could potentially increase reoperation rates in the autogenous bone graft groups but not the rhBMP-2 groups.⁹¹ If true, this could result in a biased comparison favoring rhBMP-2 over iliac crest bone grafting for some outcomes.

7a Perioperative Outcomes: Fusion With rhBMP-2 Versus Autogenous Bone Graft

All three studies reported on several perioperative outcomes. Dawson et al. 2009¹⁹ reported similar surgical time (2.4 hours [rhBMP-2] vs. 2.6 hours) and length of stay (4.0 days [rhBMP-2] vs. 4.1 days) between treatment groups. The other two studies^{20,21} also reported shorter length of hospital stay and shorter surgical time for the rhBMP-2 groups. Significant differences between groups were not reported for these outcomes, suggesting that the findings are consistent. Two studies did not report measures of dispersion for the difference between groups, so the overall precision is unknown. However, for both outcomes, the difference between groups appears small enough to infer that it is not substantial.

Less blood loss was reported for the rhBMP-2 group in all three studies; one study reporting significantly less blood loss at the $p=0.026$ level,²⁰ one was borderline ($p=0.055$),¹⁹ and one study did not report a p -value.²¹ Because no measures of dispersion were reported for mean blood loss in the largest study,²¹ the overall precision for the difference between groups is unknown for this outcome. However, the findings are consistent enough to infer that use of rhBMP-2 is associated with less blood loss in fusion procedures than use of autologous bone graft.

For perioperative outcomes, the risk of bias is moderate and the evidence is consistent and direct, but the precision is unknown. The uncertainty regarding precision results in a low overall strength of evidence for each of these outcomes (see Table 92 of Appendix C.)

7b Patient-Centered Outcomes: Fusion With rhBMP-2 Versus Autogenous Bone Graft

Two studies reported 2-year followup data on function, quality of life, and pain (leg and back);^{20,21} one study reported only on function.¹⁹

Back and leg pain: Two trials assessed back and leg pain using a 20-point numeric rating scale. The smaller study by Burkus et al. 2002,²⁰ reported significantly greater improvements in mean back pain scores for the rhBMP-2 group compared with autologous bone graft at 3 months (7.9±4.3 vs. 10.9±4.5; p=0.038), 6 months (6.8±4.3 vs. 9.9±5.1; p=0.034), and 2 years (7.4±6.0 vs. 10.9±6.0; p=0.047). Mean leg pain scores significantly improved at all time points in both groups. However, initial gains after 6 months (-8.5) in the autologous group were not maintained by 24 months (-3.1) compared with baseline. Based on analysis of variance, no significant differences were measured between groups at any time for leg pain.

The larger study by Burkus et al. 2002²¹ reported improvements in mean back and leg pain scores at all postoperative periods for both groups with no significant differences between groups at any time. Because of inconsistent results from the two studies, it is not possible to determine whether rhBMP-2 is associated with better back and leg pain control than autologous bone graft. See Table 93 to Table 102 of Appendix C for data on back and leg pain.

Function: Overall function was measured by ODI in two studies.^{20,21} Burkus et al. 2002,²¹ did not report any statistically significant differences between groups at any time point. We were unable to calculate p-values or mean differences for this study because of insufficient data.

In the smaller study by Burkus et al. 2002,²⁰ both groups improved significantly over baseline at all followup times. Based on an analysis of variance, differences between groups were statistically significant at 3 months (p=0.032), 6 months (p=0.039), and 24 months (p=0.039) favoring the rhBMP-2 arm. No statistically significant difference between groups was reported at 1 year. Significant findings from a single study provide an insufficient amount of evidence to support an evidence-based conclusion.

All three studies reported that both treatment groups had achieved clinically significant improvements in function (measured by ODI) at 2-year followup. In one study, a clinically significant improvement (≥15 point) at 1 year was similar at 2-year followup in 83 percent and 58 percent of patients in rhBMP-2 and autologous bone graft, respectively.²⁰ The other two studies reported a clinically significant improvement (≥15 percent) in both treatments arms by 2-year followup. No study reported significant differences between groups, and the effect sizes were imprecise (the CIs include the possibility of no difference or a substantial difference between groups). This data provides insufficient evidence to support a conclusion.

All three studies reported on return to work 2 years postoperatively. In all studies, the percentage of patients in the rhBMP-2 group who returned to work increased from baseline to 2 years postsurgery, while only one study reported a similar increase in the group who received autologous bone graft. However, no statistically significant between-group differences were reported at 2 years postoperatively for any study, and the effect sizes were imprecise. Baseline differences in the percentage of patients working before surgery may also have affected the percentage of patients working at 2 years postsurgery. Therefore, the return to work findings

do not support any conclusions. Data on function is shown in Table 105 to Table 107 of Appendix C.

Quality of life: One study by Burkus et al. 2002²⁰ reported on quality of life measured by two SF-36 subscales (physical and mental component). This study reported improvements in both treatment groups at all followup times, but did not report between-group differences. We were unable to calculate p-values or mean differences for these studies because of insufficient data (see Table 108 to Table 112 of Appendix C).

7c Adverse Events: Fusion With rhBMP-2 Versus Autogenous Bone Graft

Dawson et al. 2009¹⁹ reported intraoperative, perioperative, short-term, and intermediate adverse events. Two patients, one in each group, underwent intraoperative repair for incidental durotomies. Wound infections occurred at the surgical site (1 in each group) and the graft donor site (1 autologous bone graft). Reoperations due to bilateral malpositioned pedicle screws (1 patient) and hardware removal (1 patient) occurred in the rhBMP-2 group. Three patients in the autologous bone graft underwent reoperation because of residual disc material (1 patient), and pseudarthrosis (2 patients). According to a recent review,⁹¹ the FDA document summarizing this trial (no longer available on the FDA website) reported a higher rate of back and leg pain adverse events during the first three months after surgery in the rhBMP-2 group (16%) compared to the autologous control group (4.8%).

Burkus et al. 2002²⁰ reported additional posterior fixation was needed in the intermediate term in three (14 percent) patients who received autologous bone grafts. One patient in the rhBMP-2 group underwent reoperation in the short-to-intermediate term for the same reason.

The larger 2002 study by Burkus et al.²¹ reported that intraoperative vascular event rates did not differ between the rhBMP-2 group and the autograft group (4.2% vs. 3.7%). Minor graft site events related to iliac crest harvesting occurred in eight patients (5.9%) in the autograft group (not applicable to the rhBMP-2 group). Among male patients, retrograde ejaculation developed in five males (6.4%) in the rhBMP-2 group and one male (1.4%) in the autograft group; in four cases this was permanent. Reoperations were required in 11 patients (7.0%) in the rhBMP-2 group and 14 patients (10.3%) in the control group; almost all reoperations involved supplemental fixation, mostly for presumed pseudarthrosis (nonunion). The FDA summary of safety and effectiveness⁹² for this trial noted that “urogenital events occurred with greater frequency” in the rhBMP-2 group compared to the control group (11.5% vs. 7%). This document also noted that incidence of device-related adverse events was greater in the rhBMP-2 group. However, the FDA analysis included patients from an additional rhBMP-2 case series that had undergone a laparoscopic approach with a learning curve, which might partially explain the difference. Adverse events are listed in Table 113 of Appendix C.

Applicability: Fusion With rhBMP-2 Versus Autogenous Bone Graft

All three studies were multicenter trials conducted in the United States; two trials were conducted under IDE status for FDA. Both trials enrolled fewer than 50 patients but used strict inclusion criteria.^{19,20} One larger trial enrolled 279 patients and also used strict inclusion criteria.^{21,92}

On average, enrolled patients were young, ranging from 43 to 46 years of age. Because of limited reporting on comorbid diagnosis (smoking only) and no reporting on race and ethnicity, applicability of these patients to a Medicare enrolled population is unclear.

Summary: Fusion With rhBMP-2 Versus Autogenous Bone Graft

ALIF was performed in two studies;^{20,21} posterolateral fusion performed in one study.¹⁹ Studies were consistent in suggesting no substantial difference between treatment groups for surgical time and length of hospital stay, while rhBMP-2 was associated with less blood loss than autologous bone grafting.

All three studies reported treatment groups achieving clinically significant improvements defined as a 15 percent improvement or 15-point improvement in function (measured by ODI) at 2-year followup.

Adverse events were most often minimal and similar in these studies.¹⁹⁻²¹ However, certain adverse events that occurred with greater frequency in the rhBMP-2 groups (short-term pain in one trial, urogenital events in another) seemed to be underreported in the published journal reports of at least two of these studies compared to FDA documents summarizing the trial safety data.⁹¹ Also, the protocol of one of these studies¹⁹ may have increased the likelihood of reoperations in the control group relative to the rhBMP-2 group. Given these limitations, the true difference in risks of adverse events for rhBMP-2 versus autologous bone grafting, as well as the generalizability of the findings, remain unclear.

Key Points: Open Mini ALIF Versus Laparoscopic ALIF

- Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the relative perioperative outcomes or safety of open mini or laparoscopic surgery in patients undergoing fusion for disc degeneration.

Detailed Synthesis

One RCT, Chung et al. 2003,²² addressed the comparison of open mini and laparoscopic ALIF in patients with single-level degenerated discs. Twenty-two patients were treated in each group, all at level L5/S1 with a Brantigan carbon cage (DePuy Spine, Raynham, MA) and autologous bone graft. Patients in the open mini group were treated by an anterior midline extraperitoneal approach, and the laparoscopic group had an anterior transperitoneal approach. Treatment strategies are summarized in Table 119 of Appendix C.

This study was conducted in Korea at a spine hospital. Patients were enrolled between 1997 and 1999. For study design summary, see Table 115 of Appendix C. Chung and colleagues enrolled patients with degenerated discs as shown on plain radiograph, MRI, and discography with pain provocation, and symptoms lasting at least 6 months despite use of nonoperative treatment efforts for as long. Patients with other conditions including spondylolisthesis and stenosis were excluded. Patient inclusion criteria are summarized in Table 116 of Appendix C.

Of 54 patients screened for inclusion, all were enrolled and 51 (94 percent) were reported on for perioperative outcomes. Their mean age was about 50 years in both groups, with a range from 27 to 67 years. Baseline pain levels were very severe, over 9 out of 10 on average in each group. ODI scores were also high, at least 40 on average in each group. Seventy-seven percent of the laparoscopic group and 73 percent of the open mini group were women. No information on occupational status, workers compensation, or pending litigation were reported. Duration of

symptoms and previous nonsurgical treatments and nonsurgical treatments tried were not reported, nor were proportion with radicular pain or comorbid diagnoses. None of the patients had previous spinal fusion, but 20 percent had previous abdominal surgery and 7 percent had previous disc surgery. Patient demographic characteristics are summarized in Table 117 and patient health history is summarized in Table 118 of Appendix C.

The risk of bias of this study was rated as moderate because group allocation was not concealed and patients and outcome assessors were not blinded to administered treatments (see Table 122 of Appendix C).

7a Perioperative Outcomes: Open Mini ALIF Versus Laparoscopic ALIF

Surgical time: The mean surgical time was 158 minutes (range 90–330 minutes) in the laparoscopic surgery group and 83 minutes (range 40–150 minutes) in the mini open surgery group. The authors reported this difference was statistically significant, but data were insufficient to calculate a mean difference with a standard deviation or CI.

Blood loss: The mean blood loss was 85 mL (range 10–300 mL) in the laparoscopic surgery group and 68 mL (range 50–150 mL) in the mini open group. The study authors reported this difference was not statistically significant.

Length of inpatient stay: Patients in the laparoscopic surgery group stayed in the hospital a mean 3.9 days (range 2–7 days) after surgery, and patients in the mini open group stayed a mean of 3.4 days (range 2–6 days). The study authors reported this difference was not statistically significant. Perioperative data are listed in Table 120 of Appendix C.

7b Patient-Centered Outcomes: Open Mini ALIF Versus Laparoscopic ALIF

Patient-centered outcomes are not reviewed because of excessive differences between groups in duration of followup at time point reported (mean 43 months laparoscopic, mean 30 months mini open) and wide overall range of duration of followup (36–49 months).

7c Adverse Events: Open Mini ALIF Versus Laparoscopic ALIF

Only perioperative adverse events were reported. In the laparoscopic group, there was one (5 percent) patient with each of the following: cage malposition that led to symptomatic pseudarthrosis, retrograde ejaculation, and bladder dysfunction. In the mini open group, there was one (5 percent) patient each with bladder dysfunction and deep vein thrombosis. These events are summarized in Table 121 of Appendix C.

Applicability: Open Mini ALIF Versus Laparoscopic ALIF

Aside from patients being on average 50 years and mostly women, few characteristics were reported about patients in this study, so how applicable the findings are to U.S. populations or treatment practices is unclear. Patients with comorbid spondylolisthesis or stenosis were excluded.

Summary: Open Mini ALIF Versus Laparoscopic ALIF

Surgical time was significantly shorter in the mini open group, but no differences in blood loss or inpatient stay were reported and no patient-centered outcomes or adverse events were

reported, and the applicability of this information to U.S. patients and surgical practices is unclear. Laparoscopic surgical time is likely strongly influenced by surgeon experience and expertise and may differ among individuals based on ability and caseload, and may vary by surgical center type, with more advanced surgical strategies being used more frequently in specialty centers.

Key Points: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach

- Because only one study addressed this comparison for a single outcome, the evidence is insufficient to support evidence-based conclusions regarding the relative perioperative outcomes or safety of transperitoneal and retroperitoneal anterior surgical approaches in patients undergoing fusion for disc degeneration.

Detailed Synthesis: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach

One RCT, Sasso et al. 2003,²⁴ compared outcomes of men who had fusion by a transperitoneal anterior surgical approach with those who had fusion by a retroperitoneal anterior surgical approach. These patients are a subgroup of participants in the RCT by Burkus et al. 2002,²¹ but the comparison is not redundant. This was a U.S. multicenter study sponsored by Medtronic (although there is no commercial interest in the comparison of surgical approach). Study description is shown in Table 123 and treatment characteristics are summarized in Table 127; both tables can be found in Appendix C.

Sasso et al. enrolled patients with a single degenerated disc with disabling low back and/or leg pain lasting at least 6 months with nonsurgical treatment efforts lasting at least 6 months. Of the 279 patients assessed in the primary RCT, 146 were men and were assessed in Sasso et al. Of the entire population, 42 percent were working, 34 percent had workers compensation, and 14 percent had pending litigation. A third of patients smoked; no other information about health history was reported. Patient inclusion criteria and characteristics are reported in Table 124 through Table 126 of Appendix C.

This study was rated as moderate risk of bias because of lack of allocation concealment and blinding. This comparison may not have been prospectively planned, but it was included because it is a subgroup of an included study. For itemized assessment, refer to Table 129 of Appendix C.

7a Perioperative Outcomes: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach

No perioperative outcomes were reported.

7b Patient-Centered Outcomes: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach

No patient-centered efficacy outcomes were reported.

7c Adverse Events: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach

Temporary retrograde ejaculation lasting up to 1 year was experienced by two (2 percent) men in the retroperitoneal approach surgical group and four (13 percent) in the transperitoneal approach surgical group. Permanent retrograde ejaculation that was unresolved at 2 years was experienced by one (1 percent) man in the retroperitoneal group and three (15 percent) men in the transperitoneal group. Adverse events are listed in Table 128 of Appendix C.

Applicability: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach

Few characteristics about this patient subgroup were reported. The applicability of this comparison of surgical approaches may be limited to the type of fusion performed and instrumentation used.

Summary: Transperitoneal Versus Retroperitoneal Anterior Surgical Approach

This study found a higher rate of transient and permanent retrograde ejaculation among men who underwent transperitoneal anterior approach compared with retroperitoneal approach, but these results have not been confirmed in a second study.

Key Points: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion

- Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the relative benefit and harm of posterolateral or circumferential surgical approaches in patients undergoing fusion for disc degeneration.

Detailed Synthesis: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion

One RCT, Fritzell et al. 2002,²³ addressed this comparison in three treatment groups: posterolateral fusion without instrumentation, posterolateral fusion with variable screw placement (VSP), and circumferential fusion. A different publication of this study with the same lead author compared fusion to nonsurgical intervention in Key Question 1.¹² In the VSP group and circumferential fusion group, pedicle screws and open bendable plates were employed (DePuy Acromed, DePuy Spine, Raynham MA). In all groups, autologous iliac crest bone graft was used. Decompression was not performed. Treatment characteristics are summarized in Table 134 of Appendix C.

This study was conducted in 19 orthopedic departments in Sweden and enrolled patients between 1992 and 1998. It was funded by the Acromed Corp. and Ossano Scandinavia AB. General study characteristics are summarized in Table 130 of Appendix C.

Fritzell and colleagues enrolled patients with degenerated discs at one or two vertebral levels with back pain symptoms lasting at least 2 years. Diagnosis was made by consensus using imaging techniques and clinical examination. Patients had to be incapacitated enough to be on sick leave or equivalent. Enrollment criteria are summarized in Table 131 of Appendix C.

The number of patients screened for inclusion was not reported, but 222 were randomly assigned and 204 (94 percent) were treated. Patients had a mean age of 43 years (range 25–65 years), and half were women. All patients were on disability pension, and 61 percent were engaged in litigation or compensation pursuit. The mean duration of symptoms—which nonsurgical treatments were tried or the mean duration for which they were tried—and the proportion of patients with radicular pain were not reported. Eighteen percent of patients had previous successful herniated disc removal at least 2 years prior; patients with other previous spine surgeries were excluded. Thirty-nine percent of patients had some comorbidity, but they were not specified. Forty-one percent of patients smoked. Patient characteristics are summarized in Table 132 and Table 133 of Appendix C.

This study was rated as moderate risk of bias because of lack allocation concealment (especially because some patients dropped out in reaction to their treatment assignment) and lack of masking of patients and outcome assessors. Primary outcome measures were subjective. Risk-of-bias assessment is itemized in Table 139 of Appendix C.

7a Perioperative Outcomes: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion

Surgical time: In minutes, the mean \pm standard deviation surgical time was 110 ± 38 in the noninstrumented group, 194 ± 77 in the VSP group, and 335 ± 98 in the circumferential fusion group. Circumferential fusion took significantly longer than either of the posterolateral fusion groups. Posterolateral fusion took significantly more time with VSP than without.

Blood loss: In mL, the mean \pm standard deviation blood loss was 665 ± 895 in the noninstrumented group, 1283 ± 1125 in the VSP group, and 1433 ± 1236 in the circumferential group. Uninstrumented posterolateral fusion resulted in significantly less mean blood loss than either of the other groups. There was no significant difference between the VSP group and the circumferential fusion group.

Duration of inpatient stay: In days, the mean \pm standard deviation inpatient stay in days was 8.5 ± 2.6 for uninstrumented fusion, 9.7 ± 3.2 for VSP, and 11.5 ± 3.7 for circumferential fusion. The duration of stay was significantly shorter for the posterolateral groups than for the circumferential fusion group. There was no significant difference between instrumented and uninstrumented posterolateral fusion.

All perioperative outcomes data are shown in Table 135 of Appendix C.

7b Patient-Centered Outcomes: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion

Pain: The authors reported no statistically significant differences in change in pain levels among groups in back or leg pain at 2-year followup. In all groups, back pain decreased from moderate-to-severe levels to moderate levels. Leg pain decreased from moderate levels to low-to-moderate levels. These data are shown in Table 136 of Appendix C.

Function: As for pain levels, function improved in all groups. The change in function was not significantly different among groups at 2-year followup as measured using three scales: The ODI, the Million Scale, and the General Function Scale (GFS). Scores on the ODI, the primary function scale, decreased on average from indicating severe disability to indicating moderate disability.⁸⁵ See data in Table 137 of Appendix C.

7c Adverse Events: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion

Fritzell et al. reported many more harms than most studies in this report, especially perioperative harms. All of these events occurred in only one or two people per treatment group with the exception of new nerve root pain, which occurred in five (7 percent) of circumferential fusion patients, one (1.5 percent) VSP patient and no uninstrumented patients (see Table 138 of Appendix C.)

Applicability: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion

This study was conducted in Sweden and enrolled young patients (mean 43 years of age) who were on sick leave for at least 1 year. The mean duration of symptoms, which nonsurgical treatments were tried or the mean duration for which they were tried, and the proportion of patients with radicular pain were not reported.

Only 39 percent of patients had a comorbid diagnosis and 41 percent smoked. These findings may have limited applicability to older patients or those with many age-related comorbidities.

Summary: Posterolateral Fusion, With or Without Variable Screw Placement, or Circumferential Fusion

Results from intraoperative and perioperative outcomes suggest posterolateral fusion without instrumentation may be less taxing on patients than posterolateral fusion with VSP, or circumferential fusion, but replication of these findings by researchers not affiliated with the manufacturer would be reassuring. Because only one study addressed this comparison, the comparative efficacy of posterolateral fusion, with or without VSP, or circumferential fusion remains unclear primarily because of insufficient data for any one outcome.

Strength-of-Evidence Ratings: Spinal Fusion Approaches Compared to One Another for Painful Degenerated Lumbar Disc(s)

Strength-of-evidence ratings for Key Question 7 are summarized in Table 10, below.

Table 10. Key Question 7: strength-of-evidence ratings

Comparison	Outcome	Time	Studies	Risk of Bias	Consistency	Directness	Precision ¹	Evidence Favors	Strength of Evidence Rating
rhBMP-2 vs. autogenous bone graft	Surgical time	Peri-operative	3	Moderate	Consistent	Direct	Unknown	No substantial difference	Low
	Blood loss	Peri-operative	3	Moderate	Consistent	Direct	Unknown	rhBMP-2	Low
	Inpatient stay	Peri-operative	3	Moderate	Consistent	Direct	Unknown	No substantial difference	Low
	Back pain, analog	2 YR	2	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Leg pain, analog	2 YR	2	Moderate	Consistent	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	1 YR	2	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	2 YR	2	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Function, Return to work	2 YR	3	Moderate	Consistent	Direct	Imprecise	Inconclusive	Insufficient
Quality of life	2 YR	1	Moderate	Not applicable	Direct	Unknown	Inconclusive	Insufficient	
Mini open vs. laparoscopic surgery	Surgical time	Peri-operative	1	Moderate	Not applicable	Direct	Precise	Mini open	Insufficient
	Blood loss	Peri-operative	1	Moderate	Not applicable	Direct	Imprecise	Inconclusive	Insufficient
	Inpatient stay	Peri-operative	1	Moderate	Not applicable	Direct	Imprecise	Inconclusive	Insufficient
Posterolateral fusion (PLF) vs. posterolateral fusion with variable screw placement (VSP) vs. circumferential fusion	Surgical time	Peri-operative	1	Moderate	Not applicable	Direct	Precise	PLF (vs. either) VSP (vs. circumferential)	Insufficient
	Blood loss	Peri-operative	1	Moderate	Not applicable	Direct	Precise	PLF (vs. either)	Insufficient
	Inpatient stay	Peri-operative	1	Moderate	Not applicable	Direct	Precise	PLF (vs. either)	Insufficient
	Back pain	2 YR	1	Moderate	Not applicable	Direct	Imprecise	Inconclusive	Insufficient
	Leg pain	2 YR	1	Moderate	Not applicable	Direct	Imprecise	Inconclusive	Insufficient
	Function (ODI, GFS, Million)	2 YR	1	Moderate	Not applicable	Direct	Imprecise	Inconclusive	Insufficient

¹ We considered a study to have an imprecise outcome when the intervention that was favored could not be determined.

GFS = General Function Scale; Million = Million Visual Analog Scale; ODI = Oswestry Disability Index

Key Question 8: Spinal Fusion Approaches (e.g., Anterior, Posterior, Combined) and Techniques (e.g., Instrumentation, Graft Material) Compared to One Another for Painful Degenerative Lumbar Spinal Stenosis

Key Points: Posterolateral Fusion Versus Posterolateral Fusion Plus Transforaminal Interbody Fusion

- Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the comparative effectiveness or safety of posterolateral fusion versus posterolateral fusion plus transforaminal interbody fusion.

Detailed Synthesis: Posterolateral Fusion Versus Posterolateral Fusion Plus Transforaminal Interbody Fusion

One RCT (Hallett et al. 2007)¹⁸ addressed this key question. This study was described in Key Question 5. Study characteristics tables on the fusion groups only are shown in Table 140 through Table 144 of Appendix C.

8a Perioperative Outcomes: Posterolateral Fusion Versus Posterolateral Fusion Plus Transforaminal Interbody Fusion

No perioperative outcomes were reported.

8b Patient-Centered Outcomes: Posterolateral Fusion Versus Posterolateral Fusion Plus Transforaminal Interbody Fusion

Pain: Pain scores on an 11-point VAS were reduced from severe pain at baseline to moderate pain at 2-year followup in both groups. Data were insufficient to calculate p-values or mean differences; however, the findings from a single study provide an insufficient amount of evidence to support an evidence-based conclusion anyway. These data are shown in Table 145 of Appendix C.

Function: Function scores were measured in terms of the Low Back Outcome Scale and the Roland Morris Disability scale at 2-year followup. On the Low Back Outcome Scale, mean scores for transforaminal interbody fusion remained the same at 2-year followup and improved in the posterolateral interbody fusion; however, insufficient data were reported to determine p-values or mean differences between groups. These data are shown in Table 146 of Appendix C.

Quality of life: Mean SF-36 scores improved in both groups, more so in the posterolateral interbody fusion group, but because of insufficient data reporting, it is not possible to determine whether the difference between groups is either statistically or clinically important. These data are shown in Table 147 of Appendix C.

8c Adverse Events: Posterolateral Fusion Versus Posterolateral Fusion Plus Transforaminal Interbody Fusion

One (7 percent) patient in the transforaminal interbody fusion groups had a secondary decompressive laminectomy at 7-month followup, and one (6 percent) patient in the posterolateral interbody fusion group had the same at 23 months. No other harms were reported. Data are insufficient to support any evidence-based conclusions. These data are shown in Table 148 of Appendix C.

Key Points: Autograft Versus Coralline Hydroxyapatite Versus Both

- Because only one study addressed this comparison, there is an insufficient amount of evidence to support evidence-based conclusions regarding the comparative effectiveness of autograft versus coralline hydroxyapatite versus both.

Detailed Synthesis: Autograft Versus Coralline Hydroxyapatite Versus Both

One RCT addressed this comparison: Korovessis et al. 2005.²⁵ In this study, patients were randomly allocated to undergo fusion by a dorsal and lateral surgical approach involving laminotomy and undercutting facetectomy for decompression and employing as a graft material local autograft with coralline hydroxyapatite, local and iliac crest autograft with coralline hydroxyapatite, or iliac crest bone graft only. Treatment characteristics are shown in Table 154 of Appendix C.

Little information about the study design or enrolled patients was reported. This study was conducted in Greece at an orthopedic department of a general hospital (Table 150 of Appendix C). Patients with degenerative spinal stenosis with instability needing two or three level spinal fusion were enrolled. Patients with previous lumbosacral spine fusions, lumbar spine infection, osteoporotic vertebral fracture, or endocrine system disruption were excluded. No other information on patient inclusion criteria was reported (Table 151 of Appendix C). Sixty patients were enrolled, with a mean age of 61 years. No other information about the patient population was reported (Table 152 and Table 153 of Appendix C.)

This study was assigned a moderate risk-of-bias rating. Allocation was not concealed, there were differences between groups in baseline measure, there was no blinding to treatment allocation, and the primary outcomes were not objective. Full assessment is itemized in Table 158 of Appendix C.

8a Perioperative Outcomes: Autograft Versus Coralline Hydroxyapatite Versus Both

Surgical time: Surgical time was statistically shorter in the coralline hydroxyapatite group than the autograft only group (mean difference of 17 minutes less) and the group in which both coralline hydroxyapatite and iliac crest autograft were used (mean difference of 28 minutes less). The autograft only and both coralline hydroxyapatite group and autograft groups did not have statistically different surgical times.

Blood loss: The coralline hydroxyapatite group had significantly less blood loss than autograft only (mean difference of 183 cc) or coralline hydroxyapatite plus autograft (mean

difference of 133 cc). At only a few ounces, this mean difference is not likely to be clinically important. The difference between the groups that received both types of graft or autograft only was not significant.

These data are shown in Table 155 of Appendix C.

8b Patient-Centered Outcomes: Autograft Versus Coralline Hydroxyapatite Versus Both

Pain: At 2-year followup, none of the comparisons of mean pain levels were significantly different. These data are shown in Table 156 of Appendix C.

8c Adverse Events: Autograft Versus Coralline Hydroxyapatite Versus Both

Very few adverse events were reported. One patient in the autograft-only group had a superficial infection, and one patient in the group that had both graft types had a deep hematoma that required surgical intervention. Screw breakage occurred in two patients in the coralline hydroxyapatite group (1 at 6 months, 1 at 3 years) and in one patient who underwent two-level surgery in the autograft-only group (at 18 months). No other adverse events were reported. These harms are summarized in Table 157 of Appendix C.

Strength-of-Evidence Ratings: Spinal fusion approaches and techniques compared to one another for painful degenerative lumbar spinal stenosis

The strength-of-evidence ratings for Key Question 8 are presented in Table 11, below.

Table 11. Key Question 8: strength-of-evidence ratings

Comparison	Outcome	Time	Number of Studies	Risk of Bias	Consistency	Directness	Precision ¹	Evidence Favors	Strength of Evidence Rating
Posterolateral fusion vs. posterolateral fusion plus transforaminal interbody fusion	Pain, back, VAS	2 YR	1	Moderate	Not Applicable	Direct	Imprecise	Insufficient data	Insufficient
	Function, LBOS	2 YR	1	Moderate	Not Applicable	Direct	Imprecise	Insufficient data	Insufficient
	Function, RMDS	2 YR	1	Moderate	Not Applicable	Direct	Imprecise	Insufficient data	Insufficient
	Quality of life, SF-36	2 YR	1	Moderate	Not Applicable	Direct	Imprecise	Insufficient data	Insufficient
Autograft vs. coralline hydroxyapatite (CH) vs. both	Surgical time	Peri-operative	1	Moderate	Not Applicable	Direct	Imprecise	CH	Insufficient
	Blood Loss	Peri-operative	1	Moderate	Not Applicable	Direct	Imprecise	CH	Insufficient
	Pain, VAS	2 YR	1	Moderate	Not Applicable	Direct	Imprecise	Inconclusive	Insufficient

¹ We considered a study to have an imprecise outcome when the intervention that was favored could not be determined.

LBOS = Low Back Outcome Scale; RMDS = Roland Morris Disability Scale; SF-36 = Short Form-36; VAS = visual analog scale

Applicability: Spinal fusion approaches and techniques compared to one another for painful degenerative lumbar spinal stenosis

As described under Key Question 5, patients in Hallett et al. 2007¹⁸ had single-level foraminal stenosis with single-level disc degeneration with radicular pain and symptoms lasting at least 5 years. No information on previous nonsurgical treatment efforts was reported. With a mean age of 57 years, patients are older than in the studies of patients with degenerated discs alone, and may be more similar to a Medicare enrolled population. Most patients were not working.

Patients in Korovessis et al. had multilevel degenerative spinal stenosis with instability and a mean age of 61 years. Almost no other information about them was reported, so applicability of this patient population is unclear.

The studies were not conducted in the United States; Hallett et al. conducted their study in Scotland and Korovessis et al. conducted their study in Greece.

Summary: Spinal fusion approaches and techniques compared to one another for painful degenerative lumbar spinal stenosis

With only one study addressing each comparison, there is an insufficient amount of evidence to draw an evidence-based conclusion for either comparison. These studies were small, from single centers and were poorly reported.

Key Question 9: Spinal Fusion Approaches and Techniques Compared to One Another for Painful Degenerative Lumbar Spondylolisthesis

Key Points: Instrumentation Versus No Instrumentation

- Two studies addressed this comparison. The evidence is insufficient to support evidence-based conclusions regarding the comparative effectiveness and safety of instrumentation versus no instrumentation mainly due to dissimilarities in the reported outcome of the studies.

Detailed Synthesis: Instrumentation Versus No Instrumentation

Two RCTs (Fischgrund et al. 1997²⁶)(Abdu et al. 2009)²⁷ addressed this comparison. Abdu et al. 2009²⁷ is a subgroup analysis of the trial conducted by Weinstein et al.¹⁴ that is discussed under Key Question 3. In their subgroup analysis, Abdu et al. evaluated three fusion techniques: posterolateral in situ fusion (PLF); posterolateral instrumented fusion with pedicle screws (PPS); or PPS plus interbody fusion (360°). Because the proportion of patients who received single versus multilevel fusions in the 360° group differed substantially from the other two comparison groups (see Patient Criteria), we limited our discussion in this report to the PPS (instrumented) and PLF (noninstrumented) groups.

The Abdu analysis included 356 patients with degenerative spondylolisthesis and associated spinal stenosis: 213 patients in the instrumented group and 80 patients in the noninstrumented group. Patient characteristics included age of patients (mean 66 years), gender (66 percent

women), race (85 percent white), and occupational status (34 percent working full or part time). See Table 159 to Table 163 in Appendix C for more information about the characteristics of the patients included in this study.

This study received a high risk-of-bias rating. As noted for the full study publication, the comparisons were not randomized, the authors did not report allocation of concealment or patient blinding to treatment group, treatment choice was based on physician or patient preference, and the primary outcomes were not objective. Full risk-of-bias assessment is itemized in Table 181 of Appendix C.

The study by Fischgrund et al. was conducted in an orthopedic surgery department at a U.S. hospital and enrolled 68 patients (81 percent women) with a mean age of 68 years. A total of 35 patients were randomly assigned to the instrumented arm, which consisted of decompressive laminectomy and single-level autogenous bilateral lateral intertransverse process arthrodesis with transpedicular instrumentation using a posterior surgical approach. VSP pedicle screws and plates were provided by Depuy Spine, Inc. (formerly Acromed, Raynham, MA). The remaining 33 patients underwent similar surgery without instrumentation. Treatment characteristics are described in Table 163 of Appendix C.

All patients had radiographic presence of single-level degenerative lumbar spondylolisthesis with demonstrated spinal stenosis at the level of spondylolisthesis. The authors received no funding in support of this study. See Table 161 and Table 162 of Appendix C for additional patient characteristics. This study received a moderate risk-of-bias rating. The authors did not report allocation of concealment or patient blinding to treatment group, and the primary outcomes were not objective. Full assessment is itemized in Table 181 of Appendix C.

9a Perioperative Outcomes: Instrumentation Versus No Instrumentation

Several perioperative outcomes were reported in the study by Abdu et al.²⁷ The noninstrumented group benefited from significantly shorter mean surgical time (156.7 minutes vs. 212.2 minutes; $p < 0.001$); less blood loss (498.7 mL vs. 666.4 mL; $p = 0.021$), and fewer patients requiring transfusions (11 vs. 55; $p = 0.05$). This patient group also had a slightly shorter mean duration of hospital stay (4.2 days vs. 4.8 days), although this difference did not reach statistical significance. Perioperative outcomes were not reported in the Fischgrund study. See Table 179 for reports on perioperative outcomes.

9b Patient-Centered Outcomes: Instrumentation Versus No Instrumentation

Abdu et al.²⁷ reported on pain using the SF-36 bodily pain subscale and function using the SF-36 physical function subscale and the ODI. Six followup periods were reported on between six weeks and four years. Fischgrund et al. measured pain at 2- to 3-year followup using a 5-point VAS scale.

Pain: Abdu et al. reported no significant between-group differences at any time point for the SF-36 bodily pain subscale. Compared to baseline, results at 4-year follow-up indicated a lack of improvement for SF-36 bodily pain (-0.49 instrumented; -1.23 noninstrumented). At 2- to 3-year followup, Fischgrund et al. indicated reductions in pain measured by a five-point VAS (0 [no pain] to 5 [severe pain]) for both groups. For the instrumented group, preoperative versus postoperative back and lower limb pain mean scores were reduced from four points to one point. For the noninstrumented group, preoperative versus postoperative scores were reduced from four

points to two points (back pain) and to one point (lower limb). Data were insufficient to calculate p-values or mean differences.

Function: For ODI in the Abdu et al study, results at 6 months and 1 year indicated statistically significant differences in the mean changes from baseline between groups favoring the instrumented group (1 year, -26.33 vs. -20.92; $p < 0.02$). However, the mean difference between these groups is below the minimal clinically significant difference for ODI. No significant between-group differences appeared at any other time points. Results of the SF-36 physical function subscale indicated significant differences in the mean changes from baseline between groups favoring noninstrumented at 6 weeks (-18.47 vs. -24.18; $p = 0.0201$) and 3 months (-5.96 vs. -11.45; $p = 0.0247$). No significant between-group differences appeared at any other time points for the SF-36 physical function subscale. Compared to baseline, results at 4-year follow-up indicated a lack of improvement for SF-36 physical function subscale (-6.85 instrumented; -1.27 noninstrumented).

Due to the dissimilarities in the reported outcomes and instruments used to measure outcomes in the two studies that compared instrumentation to noninstrumentation, we are unable to make evidence-based conclusions. Complete data for all outcomes and timepoints for both studies is shown in Table 164 to Table 178 of Appendix C.

9c Adverse Events: Instrumentation Versus No Instrumentation

Abdu et al. reported numerous adverse events. Dural tears were the most commonly reported intraoperative event occurring in 12 percent (25/213) and 9 percent (7/80) of instrumented and noninstrumented patients, respectively. Two patients (one from each group) died within 3 months after surgery; one due to sepsis, the other due to respiratory distress. Reoperations rates were similar (6 percent) at 1 year but slightly higher for the noninstrumentation group at 2- (14 percent vs. 11 percent), 3- (16 percent vs. 12 percent) and 4-year followup (18 percent vs. 14 percent).

In the Fischgrund study, reoperation rates were similar between groups (8.6 percent (instrumented) vs. 6.1 percent (noninstrumented)) at 1- to 3-year followup. Two patients (1 from each group) underwent decompressive laminectomy at a different level. In the instrumented group, patients underwent repeat decompression and instrumented fusion (1 patient) and hardware removal for pain (1 patient). In the noninstrumented group, one patient underwent a second fusion with instrumentation. Screw failure occurred in one (2.8 percent) patient in the instrumented group. At 2 years, no new peripheral (lower motor neuron) neurological deficits were reported for either group. These data are insufficient to support any evidence-based conclusions. See Table 180 for complete listing of all adverse events reported in the studies.

Key Points: Bilateral Instrumentation Versus Unilateral Instrumentation

- Because only one study addressed this comparison, the evidence is insufficient to support evidence-based conclusions regarding the comparative effectiveness and safety of bilateral instrumentation versus unilateral instrumentation.

Detailed Synthesis: Bilateral Instrumentation Versus Unilateral Instrumentation

One RCT, Fernandez-Fairen et al. 2007,²⁸ conducted between April 1999 and September 2002 addressed this comparison. This study was conducted in Spain at one orthopedic surgery and trauma institute. Fusion was instrumented bilaterally in 42 patients and unilaterally in 40 patients. Surgery was carried out by one surgeon who use the Xia pedicular screw system (Stryker Spine, Allendale, NJ) in all cases (see Table 186 of Appendix C). A similar number of patients (25 [59.5 percent]) bilateral vs. 26 [65 percent] unilateral) also received decompression. In the nondecompressed cases, fusion mass was never placed centrally. The authors received no outside funding. Further details on study design and inclusion criteria can be found in Table 182 and Table 183 of Appendix C.

Patients with degenerative spondylolisthesis having persistent or recurrent lumbar pain lasting at least 6 months were enrolled. Mean age was approximately 61 years with each treatment group comprised of approximately 60 percent women. The authors reported no significant differences between groups for location and number of fusion at one level or at more than one level ($p>0.1$) or for the association with decompression ($p>0.5$). Patient characteristics are detailed in Table 184 and Table 185 of Appendix C.

This study received a moderate risk-of-bias rating. The authors did not report allocation concealment or patient blinding to treatment group, and the primary outcomes were not objective. Full assessment is itemized in Table 195 of Appendix C.

9a Perioperative Outcomes: Bilateral Instrumentation Versus Unilateral Instrumentation

Average surgical time was significantly shorter in the unilateral group compared with the bilateral group (168 ± 37 minutes vs. 203 ± 35 minutes; $p<0.001$). For the unilateral group, mean blood loss was less ($1,060\pm 270$ mL vs. $1,155\pm 207$ mL) and fewer transfusions were needed (14 times vs. 20 times). Mean duration of hospital stay, however, was shorter in the bilateral group (3.85 ± 0.54 days vs. 3.97 ± 1.01 days). No significant differences were reported between groups for these outcomes. These data are presented in Table 193 of Appendix C.

9b Patient-Centered Outcomes: Bilateral Instrumentation Versus Unilateral Instrumentation

Fernandez-Fairen et al. 2007²⁸ reported results for several subscales of the SF-36v2 (version 2), the international version of SF-36.

Pain: Compared with baseline, significant improvements were reported postoperatively at 1- and 3-year followup for the bodily pain subscale for both treatments. No significant differences between groups were reported for this subscale. Data is reported in Table 187 and Table 188 of Appendix C.

Function: Significant improvements at the $p = 0.001$ level were reported for both groups for the physical component subscale at 1- and 3-year followup. No significant differences between groups were reported. See Table 189 and Table 190 in Appendix C.

Quality of life: Compared with baseline, statistically significant improvements ($p = 0.001$) were reported postoperatively for both groups at 1- and 3-year followup for the physical component summary subscale. For the mental component summary subscale, the bilateral group demonstrated significant improvement ($p<0.01$) postoperatively at 1 and 3 years. For the

unilateral group, a statistically significant improvement was reported at 1-year followup compared with baseline (52.9 ± 7.96 vs. 44.43 ± 15.00 ; $p < 0.02$), and this level of improvement was maintained at the 3-year followup (52.34 ± 8.07 vs. 44.43 ± 15.00 ; $p < 0.02$). However, no significant differences between groups were reported for any subscale. These data, reported in Table 191 and Table 192 of Appendix C, are insufficient to support any evidence-based conclusions.

9c Adverse Events: Bilateral Instrumentation Versus Unilateral Instrumentation

Three (7.1 percent) patients in the bilateral group underwent reoperation because of nerve root irritation due to the violation of pedicle cortex by screw. No reoperation was reported for the unilateral group. This data is available in Table 194 of Appendix C.

Strength-of-Evidence Ratings: Spinal Fusion Approaches and Techniques Compared to One Another for Painful Degenerative Lumbar Spondylolisthesis

The strength-of-evidence ratings for Key Question 9 are presented in Table 12, below.

Table 12. Key Question 9: strength-of-evidence ratings

Comparison	Outcome	Time	Studies	Risk of Bias	Consistency	Directness	Precision ¹	Evidence Favors	Strength of Evidence Rating
Instrumentation vs. no instrumentation	Pain, bodily, SF-36	3 mos – 4 years	1	High	Not applicable	Direct	Imprecise	Inconclusive	Insufficient
	Physical function, SF-36	3 mos – 4 years	1	High	Not applicable	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	3 mos – 4 years	1	High	Not applicable	Direct	Imprecise	Inconclusive	Insufficient
	Surgical time	Peri-operative	1	High	Not applicable	Direct	Precise	Noninstrumented	Insufficient
	Blood loss	Peri-operative	1	High	Not applicable	Direct	Precise	Noninstrumented	Insufficient
	Duration of stay	Peri-operative	1	High	Not applicable	Direct	Imprecise	Inconclusive	Insufficient
	Transfusions	Peri-operative	1	High	Not applicable	Direct	Precise	Noninstrumented	Insufficient
	Pain, back, VAS	1–2 years	1	Moderate	Not applicable	Direct	Imprecise	Insufficient data	Insufficient
Pain, lower limb, VAS	1–2 years	1	Moderate	Not applicable	Direct	Imprecise	Insufficient data	Insufficient	
Bilateral instrumentation vs. unilateral instrumentation	Surgical time	Peri-operative	1	Moderate	Not applicable	Direct	Imprecise	Insufficient data	Insufficient
	Blood loss	Peri-operative	1	Moderate	Not applicable	Direct	Imprecise	Insufficient data	Insufficient
	Duration of stay	Peri-operative	1	Moderate	Not applicable	Direct	Imprecise	Insufficient data	Insufficient
	Transfusions	Peri-operative	1	Moderate	Not applicable	Direct	Imprecise	Insufficient data	Insufficient
	Pain, bodily, SF-36v2	1–3 years	1	Moderate	Not applicable	Direct	Imprecise	Insufficient data	Insufficient
	Physical function, SF-36vs2	1–3 years	1	Moderate	Not applicable	Direct	Imprecise	Insufficient data	Insufficient
	Physical component summary, SF-36v2	1–3 years	1	Moderate	Not applicable	Direct	Imprecise	Insufficient data	Insufficient
	Mental component summary, SF-36v2	1–3 years	1	Moderate	Not applicable	Direct	Imprecise	Insufficient data	Insufficient

¹ We considered a study to have an imprecise outcome when the intervention that was favored could not be determined.

SF-36v2 = Short Form-36, version 2; VAS = visual analog scale

Applicability: Spinal Fusion Approaches and Techniques Compared to One Another for Painful Degenerative Lumbar Spondylolisthesis

One of the three studies in this evidence base was conducted in the Midwestern United States.²⁶ In this study, enrolled patients were only required to fail a minimum of 3 months of nonoperative treatments. This study also required an associated diagnosis of stenosis and reported fewer than 10 percent of included patients were smokers.²⁶ One study, conducted in Spain, excluded patients who smoked or had other factors that could potentially influence healing.²⁸ The remaining study, conducted in the U.S., reported numerous comorbid diagnoses which makes it more applicable to the Medicare-enrolled population.²⁷

All three studies were conducted at orthopedic institutes, which limits their generalizability to general hospital settings. One study was published in 1997 but did not report enrollment years.²⁶ The other two studies, enrolled patients from 1999 to 2002²⁸ and 2000 to 2005.²⁷ Two studies were conducted more than a decade ago, so they may be dissimilar to surgeries performed today. These studies enrolled patients (mean age range 60–68 years) similar in age to a Medicare enrolled population.⁹³ However, in one of the studies, more than 80 percent of subjects were women.²⁶ Limitations also included enrolling small study populations (fewer than 85 patients).

Summary: Spinal Fusion Approaches and Techniques Compared to One Another for Painful Degenerative Lumbar Spondylolisthesis

The evidence base for this key question was comprised of three studies; two studies evaluating similar comparative treatments (instrumented vs. noninstrumented fusion). Abdu et al. 2009²⁷ reported benefits to noninstrumented patients for perioperative outcomes but minor differences between groups for pain and function. This study was a subgroup analysis of a larger study conducted by Weinstein et al¹⁴ and had a number of limitations which are fully described under Key Question 3 of this report.

Outcomes of interest reported in Fischgrund et al. 1997²⁶ were limited to average pain score measured by a five-point VAS. Although a small study, the authors indicated this was “the largest prospectively randomized study reporting on the use of pedicle screws for one diagnosis.” Fernandez-Fairen et al. 2007²⁸ reported on perioperative outcomes and several subscales of SF-36v2. However, significant differences between groups were reported for only one outcome, surgical time.

Key Question 10: Patient Characteristics Predictive of Outcomes After Spinal Fusion

Key Points

- Older age (65 years or older) appears to be associated with worse patient outcomes following spinal fusion.

Detailed Synthesis

There were insufficient data addressing Key Question 1–9 to address this key question using secondary analysis techniques such as meta-regression and subgroup analyses. We therefore performed a qualitative review of primary literature using patient-level data.

The evidence base for this question consisted of seven studies reported in eight separate publications (two studies were companion studies with one reporting data for up to 2 years followup¹⁴ and one reporting 4-year followup data.¹⁵ Of the studies, one was an RCT²⁹, one was a nonrandomized comparative trial^{14,15} and the other five were cohort studies (3 retrospective studies and 2 prospective studies).³⁰⁻³⁴ The studies examined a number of factors that could potentially affect patient outcomes following spinal fusion surgery. The most commonly assessed factors include age, gender, and workers compensation status. Table 13 lists the factors assessed in each of the studies addressing this key question.

Table 13. Patient and treatment variables

Study	Type of Study	Patient and Treatment Variables Assessed	Outcome(s) of Interest
Fukuta et al. 2011 ³⁰ Japan	Retrospective cohort study	Age, gender, body mass index, osteoporosis grader and spacer position (center vs. anterior)	Occurrence of spacer subsidence (sinking of spacer position)
Keorochana et al. 2011 ³¹ Thailand	Retrospective cohort study	Gender, age at time of surgery, onset of disease, income, comorbid conditions, smoking, work status, diagnosis, number of level of instrumentation, and SI fusion	Failed change defined as a reduction of Oswestry Disability Index score of less than 15%
Carreon et al. 2009 ³² United States	Prospective cohort study	Mental component summary (MCS), physical component summary (PCS), Oswestry Disability Index, back pain prominence, body mass index, age, smoking habits, and workers compensation status	Health-related quality of life
Weinstein et al. 2007 & 2009 ^{14,15}	Prospective nonrandomized comparative trial	Age, level of decompression, education level, gender, smoking history, severity of symptoms at baseline, duration of symptoms, treatment preferences, number of stenotic levels, severity of stenosis on imaging, number of coexisting conditions, neurogenic claudication, neurological deficit, and baseline SF-36 mental component score.	SF 36 bodily pain score, SF 36 physical function score, and Oswestry Disability Index score.
Okuda et al. 2006 ³³ Japan	Prospective cohort study	Patients were divided into 2 groups based on age: Group 1 included 31 patients 70 years of age or older (average age 74 years) and Group 2 included 70 patients younger than 70 years (average age 59 years). The clinical (rate of recovery) and radiological (results of fusion) outcomes of the groups were compared.	Rate of recovery and radiological results of fusion
Schuler et al. 2005 ³⁴ United States	Retrospective cohort study	Age, weight, sex, and workers compensation	Disability, quality of life, and back pain
Hagg et al. 2003 ²⁹ Sweden	Randomized controlled trial	Heavy job, workers compensation, disability pension, unemployment, sick leave due to lower back pain, gender, comorbidity, married/cohabit, smoking, prior surgery, continuous pain, personality traits, and psychological symptoms	Patient-rated improvement status and work status

Because the intent of this question is not to attribute cause, we did not formally assess risk of bias of the studies that addressed Key Question 10 or rate the strength of the evidence. The majority of studies that met inclusion criteria for this question were cohort studies, which are of higher risk of selection and other biases than controlled trials.

The average age of the patients enrolled in the studies ranged from 43 to 65.2 years. In all of the studies, at least half of the enrolled patients were female (range of female patients across studies 50 to 83 percent). All patients considered for surgery in the studies addressing this question had severe, disabling pain in the back and/or lower extremities that persisted for at least 12 weeks. The majority of patients across the studies had received a diagnosis of degenerative disc disease or degenerative spondylolisthesis. Only two studies reported on comorbidities. One study indicated that 34 percent (54/158) of the patients had comorbid conditions such as heart disease and arthritis of the lower extremities.³¹ The other study reported that 58 percent of patients had comorbidities, but did not describe the types of comorbidities.²⁹ The average length of followup across the studies was 2 years. See Table 196 of Appendix C for more information about the patients and treatment characteristics of these studies.

The fusion strategy varied across the studies. In the RCT, patients in the fusion group underwent posterolateral fusion with (74 patients) and without (73 patients) variable screw placement with pedicle screws and plates.²⁹ Patients in the comparison group (63 patients) received physical therapy primarily, with one or more additional treatments that ranged from acupuncture to cognitive functioning training. The patient population included in this study is also considered in one of the studies included in the evidence base for Key Question 1 of this report (Fritzell et al. 2001¹²).

The specific fusion strategy used in the nonrandomized controlled trial by Weinstein et al. was not clearly described.^{14,15} The study authors indicate that participating physicians agreed to use standard posterior decompressive laminectomy with or without bilateral single fusion (autogeneous iliac crest bone grafting at the level of the listhesis with or without posterior instrumentation using pedicle screws) for treating patients with degenerative spondylolisthesis (368 patients, 6% of which underwent decompressive laminectomy without fusion). Patients in the nonsurgical comparison (233 patients) received at minimum active physical therapy, education/counseling with home exercise instruction, and non-steroidal anti-inflammatory drugs if tolerated. However, patients in this group could have received any additional non-surgical therapy deemed appropriate by their physician.

Fifty-four patients in the retrospective cohort study conducted by Fukuta et al. underwent transforaminal interbody fusion (TLIF) using kidney-type intervertebral spacers.³⁰ In another retrospective study by Keorochana et al., 158 patients underwent decompressive laminectomy and instrumented fusion with pedicular screw.³¹ Okuda et al.³³ studied 101 patients who had undergone posterior lumbar interbody fusion, and Schuler et al.³⁴ studied 392 patients who had anterior lumbar interbody fusion. Finally, the majority of patients (76.1 percent of 546 patients) in the study by Carreon et al. underwent posterolateral fusion.³²

Findings

In general, the studies used multivariate regression analysis to examine the association between prognostic factors and patients' postfusion outcomes. The RCT by Hagg et al. used multivariate logistic regression to examine the impact of a variety of patient characteristics, including personality traits, personality disorders, and symptoms of depression, on patient-rated improvement and work status (part- or full-time work) at 2-year followup (see Table 13 for the full list of patient characteristics).²⁹

Overall, nonimprovement in the fusion group was significantly associated with a highly neurotic personality type, although the association was relatively weak. Personality traits were assessed using the Karolinska Scales of Personality (KSP). According to this scale, patients with

neurotic personality traits (“neurotic personality”) tend to be tense and stiff, restless, uneasy, panicky, easily fatigued, remorseful, and experience tremor and palpitations under stress. For patients in the nonsurgical group, more depressive symptoms were associated with greater reports of improvement. Again, although statistically significant, the effect size is relatively small, suggesting a weak association. The authors of the study indicate that this finding is difficult to interpret because it is contrary to most other studies that report that patient improvement is associated with a reduction in depressive symptoms. Thus, the authors make no conclusions regarding this finding. For both the fusion group and the nonsurgical group, patients with short sick leave were more likely to be working at followup. Younger age was also associated with working at a job for patients in the fusion group. All of these effect sizes were also small. See Table 196 of Appendix C for further details about the findings of this and the other studies considered under this key question.

Weinstein et al. fit regression models for selected subgroups to evaluate factors that could potentially have an impact on the effect of treatment.¹⁴ They considered a number of factors, all of which are listed in Table 13. The findings of their analyses indicated that patients less than 65 years old at baseline had larger treatment effects in favor of surgery at 3 months (21.3 vs. 14.6 for bodily pain, $p = 0.02$), but not at 1 or 2 years. They also found that patients with a degenerative spondylolisthesis at L3-L4 had a larger treatment effect than patients with a level of L4-L5 (33.1 vs. 16.8 for bodily pain, $p = 0.01$) at 2 years (not a 3 months or 1 year). Finally, patients with no more than a high school education had smaller treatment effects for surgery at 3 months (12.8 vs. 20.5 for bodily pain, $p = 0.002$) and at 2 years (11.5 vs. 21.6, $p = 0.01$). Subsequent analysis in the 4-year followup study indicated that patients with neurogenic claudication had larger treatment effects favoring surgery on both bodily pain and function.¹⁵

Three of the five cohort studies also divided patients into subgroups to compare patient outcomes. Fukuta et al. divided patients in several subgroups based on age, gender, body mass index, and primary disease.³⁰ Okuda et al. divided patients into two groups according to age— younger patients (younger than 70 years of age) and older age (older than 70 years of age)—and compared the clinical and radiographic outcomes of the two groups.³³ Schuler divided patients based on disc space: collapsed (<5 mm, 38 patients); intermediate collapsed (5–10 mm; 120 patients); intermediate (10–15 mm, 149 patients); and tall (>15 mm, 85 patients).³⁴

Fukuta et al. assessed the risk factors for spacer subsidence (SS) or sinking/displacement of the spacer using kidney-type spacers. The findings of their subgroup analysis indicated that center position of the kidney-shaped spacer and age (older than 65 years of age) were significant risk factors for SS 2 mm or more.³⁰ Further multivariate analysis indicated that age and spacer position had a significant synergistic effect on increasing SS. Okuda found that the rate of collapsed union or delayed union was significantly higher for patients 70 years of age or older than for patients younger than 70 years of age.³³ Patients in the collapsed disc group (<5 mm) in the Schuler et al. study had significant improvement in functioning (as measured by the ODI) at 6-month, 1-year, and 2-year followups.³⁴ These patients also demonstrated statistically significant improvement on the physical component summary of the SF-36 and VAS for low back pain at 6-month, 1-year, and 2-year followups.

Results of regression analysis in the Keorochana et al. study showed statistically significant associations between age at time of surgery (older than 65 years of age), onset of disease (longer than 24 months), and number of levels of instrumentation (more than 4 levels) and failed clinical outcome change.³¹ Failed change in this study was defined as a reduction of less than 15 percent on the ODI. Finally, the results of the study by Carreon et al. suggest that patients with better

preoperative mental component summary scores and worse preoperative disability scores (as measured by the overall ODI score) achieved greater improvement in disability after receiving fusion surgery. In this study, patients receiving workers compensation appeared to have less improvement in overall disability.

Summary: Patient Characteristics Predictive of Outcomes After Spinal Fusion

Overall, older age (65 years or older) appears to be associated with worse patient outcomes following spinal fusion. Three of the six studies considered for this question showed a statistically significant association between older age (65 years of age or older) and poor patient outcomes. The remaining three studies had, on average, a younger patient population with fewer patients aged 65 years or older, so they were less likely to detect this association. In one study, patients less than 65 years old at baseline had larger treatment effects in favor of surgery at short-term followup (3 months), but not at longer followup times. In another study, younger age was associated with working at least part-time at followup for patients who underwent fusion.

Discussion

Key Findings and Strength of Evidence

The purpose of this report was to assess whether previous research is sufficient to support evidence-based conclusions about the benefits and harms of lumbar fusion relative to nonsurgical treatments or other invasive treatments or to support conclusions about outcomes following the use of different fusion strategies. We also considered, which, if any, patient characteristics (e.g., pain severity, type or duration of treatment) the evidence suggests are associated with better or worse outcomes after spinal fusion. The report focused on outcomes for adults suffering from moderate-to-severe or severe pain due to degenerated disc(s), degenerative stenosis, or degenerative spondylolisthesis of the lumbar spine who underwent fusion surgery. The primary outcomes of interest in this report were patient-centered and included function, quality of life, and pain. We also considered adverse events (e.g., reoperation, neurological injury, blood clots, and infection) and perioperative outcomes such as surgery time, blood loss, and length of hospital stay.

The overall evidence base for this report consisted of 25 studies (2 studies addressed more than 1 question): 5 studies compared fusion surgery with continued noninvasive treatment, 3 compared fusion surgery with other invasive procedures (e.g., decompressive laminectomy), 10 compared different spinal fusion approaches and techniques, and 7 studies considered patient or treatment factors associated with patient outcomes following spinal fusion. For all but one of the comparative studies that met the inclusion criteria for this report, the overall risk-of-bias rating was moderate. The moderate rating was largely because of lack of concealment of allocation and/or blinding of patients or outcome assessors to treatment received or not reporting if concealment or blinding took place in the study. One study earned a high risk-of-bias rating due to high treatment crossover and other limitations that compromised the randomization of the study and introduced selection and other biases.

Because of the small evidence base (fewer than 2 studies) for some key questions, clinical heterogeneity (i.e., differences in patient characteristics, surgical approach, and control condition) or insufficient data reporting for other comparisons, meta-analysis was not used to determine summary effect size estimates. Instead, qualitative analysis in which the evidence for each key question was described, compared, and contrasted was used to draw conclusions where the evidence permitted (e.g., informative statistically or clinically significant findings from more than a single study).

Our analyses of the evidence led to the following conclusions:

- Limited data suggests that fusion leads to greater improvement in back pain relief and function than physical therapy at 2-year followup (strength of evidence: low); however, whether the difference is clinically significant is unclear (the confidence intervals overlap with what is considered a clinically significant difference), and findings at 1 year are insufficient to allow conclusions.
- Limited evidence suggests that shorter surgical time, less blood loss, and shorter inpatient stays are associated with arthroplasty, and that disc recipients have better ODI functions scores at 6 weeks postsurgery (strength of evidence: low). The difference in ODI functions were not observed at later followup times.

- rhBMP-2 is associated with less blood than autogenous bone graft, and surgery time and length of hospital stay do not differ substantially for these products for adults undergoing fusion for low back pain due to degenerated disc(s). Strength of evidence: low.

For all other comparisons, either no study was identified that met the inclusion criteria for this report, only one small study made up the evidence base, or evidence from more than one study was inconsistent and/or imprecise. All of these scenarios resulted in insufficient evidence to support conclusions. Because of insufficient reporting and variation in surgical methods used in the different studies, the incidence of adverse events (serious and minor) associated with fusion could not be determined conclusively. The table below summarizes the conclusions and strength of evidence for comparisons in which the evidence base consisted of more than a single study. Our analysis used the same definitions of risk-of-bias, consistency, directness, and precision as described by Owens et al., and our assessment of overall strength of evidence also followed the guidance presented in this article.⁹

Table 14. Summary of conclusion and strength-of-evidence ratings

Comparison	Outcome	Time	Number of Studies (Total N)	Risk of Bias	Consistency	Directness	Precision1	Evidence Favors	Strength of Evidence Rating
Fusion vs. physical and exercise therapies (Key Question 1)	Pain, back, VAS	1 YR	3 (N=153)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		2 YR	2 (N=299)	Moderate	Consistent	Direct	Precise	Fusion	Low
	Pain, leg, VAS	1 YR	2 (N=118)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Pain, drugs	1 YR	2 (N=118)	Moderate	Insufficient data	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	1 YR	3 (N=153)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
		2 YR	2 (N=299)	Moderate	Consistent	Direct	Precise	Fusion	Low
	Function, GFS	1 YR	2 (N=118)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
Function, return to work	1 YR	2 (N=118)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient	
Fusion vs. artificial intervertebral disc (Key Question 4)	Surgical time	Peri-operative	2 (N=472)	Moderate	Consistent	Direct	Precise	Disc	Low
	Blood loss	Peri-operative	2 (N=470)	Moderate	Consistent	Direct	Precise	Disc	Low
	Inpatient stay	Peri-operative	2 (N=473)	Moderate	Consistent	Direct	Precise	Disc	Low
	Pain, VAS	6 WK–2 YR	2 (N=465)	Moderate	Consistent	Direct	Imprecise	Inconclusive	Insufficient
	Pain, drugs	2 YR	2 (N=469)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	6 WK	2 (N=467)	Moderate	Consistent	Direct	Precise	Disc	Low
3 MO–2 YR		2 (N=467)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient	
rhBMP-2 vs. autogenous bone graft (Key Question 7)	Surgical time	Peri-operative	3 (N=371)	Moderate	Consistent	Direct	Unknown	No substantial difference	Low
	Blood loss	Peri-operative	3 (N=371)	Moderate	Consistent	Direct	Unknown	rhBMP-2	Low
	Inpatient stay	Peri-operative	3 (N=371)	Moderate	Consistent	Direct	Unknown	No substantial difference	Low

Comparison	Outcome	Time	Number of Studies (Total N)	Risk of Bias	Consistency	Directness	Precision ¹	Evidence Favors	Strength of Evidence Rating
	Back pain, analog	2 YR	2 (N=271)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Leg pain, analog	2 YR	2 (N=271)	Moderate	Consistent	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	1 YR	2 (N=298)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Function, ODI	2 YR	2 (N=271)	Moderate	Inconsistent	Direct	Imprecise	Inconclusive	Insufficient
	Function, Return to work	2 YR	3 (N=316)	Moderate	Consistent	Direct	Imprecise	Inconclusive	Insufficient

¹ We considered a study to have an imprecise outcome when the intervention that was favored could not be determined.

GFS = General Function Scale; ODI = Oswestry Disability Index; VAS = visual analog scale

The last key question considered in this review focused on which, if any, patient characteristics (e.g., pain severity, type or duration of prior treatment) were associated with better or worse outcomes after spinal fusion. The evidence addressing the other nine comparative effectiveness questions in the review were insufficient to perform secondary analysis techniques such as meta-regression and subgroup analyses to answer this question. We therefore performed a qualitative review of primary literature using patient-level data to address this key question. Overall, the evidence suggests that older age (65 years or older) appears to be associated with worse patient outcomes following spinal fusion. Although a number of other factors were considered in the studies that assessed prognostic factors, the results, unlike those for age, were largely inconsistent.

Findings in Relationship to What is Already Known

The findings of our review are similar to previous systematic reviews on the same topic. In a relatively recent review on the benefits and harms of surgery for nonradicular back pain with common degenerative changes, Chou et al. (2009) concluded that fusion is “slightly to moderately more effective than standard (nonintensive) nonsurgical therapy for improvement in pain and function.”⁹⁴ These authors, however, found only fair evidence that fusion is no more effective than intensive rehabilitation with a cognitive behavioral emphasis. Chou et al. indicated that further conclusions about the benefits of fusion compared to nonsurgical treatments could not be drawn due to differences between studies in the nonsurgical comparator treatments. In our report, we note that the nonsurgical comparator treatments differed across studies in the duration and intensity of the physical therapy component and in the supplemental treatments received (e.g., acupuncture, injections, advice, and/or cognitive therapy).

Also similar to the findings of our report is that Chou et al. found that the evidence was insufficient to determine optimal fusion methods. According to these authors, “instrumentation and electrical stimulation appear to enhance fusion rates, but effects on clinical outcomes are not established.”⁹⁴ These authors reported that the major complications of surgery included deep wound infections, major bleeding during surgery, thrombosis, acute respiratory distress syndrome, pulmonary edema, and heart failure. However, like in our report, Chou et al. indicated that complication rates following fusion vary widely across studies and are difficult to interpret due to differences in techniques, study population, and methodological shortcomings.

In another previous systematic review prepared for the Washington State Health Care Authority, ECRI Institute (2007) concluded that the evidence was insufficient to determine the relative benefits of lumbar fusion compared to conventional physical therapy or to intensive exercise/rehabilitation in patients with or without prior back surgery.⁹⁵ Discrepancies between this review and the ECRI Institute review are due to differences in the study inclusion criteria and number of studies included in the evidence base for each review. This review also includes a more recently published study comparing fusion to conventional physical therapy (Ohtori et al. 2011) that was not included in the ECRI review. The review by ECRI also concluded that lumbar fusion leads to higher rates of both early and late adverse events compared to non-intensive physical therapy or intensive exercise/rehabilitation plus CBT.

Finally, the Duke University Evidence-Based Practice Center (McCrorry et al. 2006) prepared a previous systematic review of spinal fusion for degenerative disease that was sponsored by AHRQ’s Technology Assessment Program.⁷ This review is intended, in part, to serve as an update to the review prepared by Duke University. The Duke review (currently available only in draft form) was primarily focused on outcomes of lumbar fusion in patients age 65 or older with

degenerative disc disease compared to nonsurgical management or other surgical strategies. The report evaluated four RCTs (three of which are included in this review) that directly compared lumbar fusion to nonsurgical therapies. The tentative conclusions of the draft report were that the evidence does not conclusively demonstrate short-term or long-term benefits compared with nonsurgical treatment for degenerative disc disease. Again, differences in the study inclusion criteria and number of included studies are what explain the difference in conclusions between this report and the Duke review.

Applicability

Our overall assessment of applicability follows the framework developed by Atkins et al., which evaluates applicability along the following categories: population, intervention, comparator, outcomes, and setting.⁹⁶ Our assessment of applicability begins with an evaluation of the age of patients in the studies included in this review for Key Questions 1 to 9. The use of spinal fusion has been on the rise in general, but particularly among people over the age of 60.⁷ Of particular interest in this report is the potential applicability of the findings to patients over the age of 60, as these individuals are likely to be Medicare beneficiaries. Information captured on age indicates that the average age of the patients enrolled in the majority of the included studies (10 studies) was ≤ 43 years. Of the seven studies conducted in the U.S., four enrolled patients with an average age ranging from 40 to 43 years.^{16,17,20,21}

Recent statistics indicate that slightly more than 50 percent of current Medicare beneficiaries range in age from 65 to 74 years.⁹³ However, only four of the included studies enrolled similar aged patients. The studies include one conducted in Scotland¹⁸ in which the average age of patients was 57 years, one conducted in Greece²⁵ in which the average age was 61 years, and two studies conducted in the U.S. in which the average age was 56 and 69 years, respectively.^{19,26}

Other patient characteristics evaluated as part of our assessment of applicability include gender, race and ethnicity, duration of symptoms, comorbid diagnoses, and previous treatments. Most of the included studies reported on gender. Overall, seven studies reported enrolling <50 percent women (range 43% to 49%),^{11,13,17,18,21,23,24} and one study enrolled only males.²⁴ Race or ethnicity was reported in only two studies. One of the studies enrolled only Caucasians¹⁸ and the other enrolled a slightly more diverse population that included 81 percent Caucasians, 12 percent Hispanics, 4 percent African-Americans, and 1 percent Asian-Americans.¹⁶ African heritage has been described as a general risk factor for degenerative spinal conditions.⁴⁰ However, this population was largely underrepresented in the studies included in this review. Further, the limited description of ethnicity in the included studies limits the potential applicability of the findings of this review for the Medicare population, which is currently described as mostly women of white, black and Hispanic descent.⁹³ Further, current forecasts anticipate a large increase in minority populations among the Medicare-enrolled population.⁹³

Reports on duration of symptoms and comorbidities were limited. While inclusion criteria required enrollment of symptomatic patients, only six studies reported on duration of symptoms. Two of the six studies indicated that patients were symptomatic for less than six months,^{26,28} while two other studies indicated that patients were symptomatic for an average of seven years.^{12,13} Three studies reported on both the overall percent of patients with comorbidities and the percent of patients who smoked.^{11,12,23} Nine studies reported only on use of tobacco, which ranged from 25 percent to 67 percent of patients,^{10,16-21,24,26} and three studies made no report of comorbidities or smoking history.^{13,22,28} One study excluded patients who smoked or had other factors that could potentially influence healing.²⁸ Such limited reporting on comorbidities makes

it difficult to generalize the findings of these studies to the majority (83%) of Medicare beneficiaries who currently suffer from at least one chronic condition⁹⁷ or to a minority of beneficiaries (over 20%) who suffer from five or more chronic conditions.⁹⁷

Use of previous non-surgical treatments was more consistently reported than comorbidities or other health related factors among the included studies. Previous non-surgical treatments were described in five studies, four of which were conducted in the U.S. These studies described use of narcotics, injections, physical therapy, chiropractic, aerobic conditioning and nonsteroidal anti-inflammatory drugs.^{16,17,20,26}

Applicability of interventions used in the included studies may be somewhat limited due to the modernization of techniques and devices. Although devices in this report have been evaluated for inclusion by members of the technical expert panel for this report, many of the included studies were conducted over a decade ago, which may limit their applicability to currently used fusion instrumentation (e.g., rods, pedicle and facet screws) and techniques. Seven studies enrolled patients prior to 2000,^{10-12,20-23} while two studies, one conducted in 1997²⁶ and one in 2003,²⁴ did not report enrollment years.

This report included comparisons of fusion to non-invasive (Key Question 1 through 3) and other invasive treatments (Key Question 4 through 6). Four studies included in Key Question 1 compared fusion to physical therapy. Two of these studies were conducted in Scandinavia^{10,11} and patients in the non-surgical group received complex and time-consuming exercise and cognitive therapy programs that are dissimilar from typical outpatient physical therapy in the U.S. In one study, conducted in Japan,¹³ the nonsurgical treatment was administered for over 2 years, which also may not be representative of U.S. practices. Patients in the noninvasive comparator group of the one study that addressed Key Question 3 received a range of therapies that at minimum included active physical therapy, education/counseling with home exercise instruction and a non-steroidal anti-inflammatory drug.¹⁴

The above study, however, was designed to be a pragmatic trial. As such, the heterogeneity of treatments in the non-surgical group is probably at least somewhat representative of the range of treatments that patients would likely receive in clinical practice. Thus, the findings of the study may be more representative of what is likely to occur in clinical practice. However, such heterogeneity makes replication of the findings of pragmatic studies difficult because the exact mix of the alternative or comparator treatment typically varies among patients within the studies and is likely to vary from one clinical practice to another.

Use of total disc replacement was assessed in Key Question 4. We identified two multicenter studies^{16,17} comparing fusion to the same device, the ProDisc-L (Synthes Spine, West Chester, PA) making applicability to other discs unclear. There are currently two devices (ProDisc-L and the Charite) approved for total disc replacement in the U.S. However, this procedure is currently not covered for Medicare beneficiaries older than 60 years of age.⁶⁴

The outcomes of interest in this review were peri-operative, patient-centered and adverse events. Validated scales were consistently used across the included studies to measure important patient-centered outcomes. Pain was typically assessed using the Visual Analog Scale. Function was measured using several scales including the Oswestry Disability Index, Japanese Orthopaedic Association back pain questionnaire, and General Function Scale. The Short Form 36 was used in several studies to assess quality of life. However, outcome reporting was restricted in two of the included studies. One study that compared open mini and laparoscopic ALIF in patients with single-level degenerated discs only reported peri-operative outcomes.²² The other study compared outcomes of men who had fusion by a transperitoneal anterior surgical

approach with those who had fusion by a retroperitoneal anterior surgical approach and only reported adverse events.²⁴ Studies, however, consistently reported long-term follow-up. A majority reported outcomes at 2 years, and one study reported a outcomes at 5-years following surgery.²⁸

Lastly, we consider the applicability of setting of included studies. Nine (56%) studies were conducted outside the U.S.,^{10-13,18,22,23,25,28} and many of the studies were conducted in spine specialty centers and orthopedic centers. Such factors are likely to limit the applicability of the findings of these studies for both U.S. Medicare beneficiaries and other primary care populations.

Implications for Clinical and Policy Decisionmaking

Most spine surgery is performed electively. Thus, any treatment decision involves considering the trade-offs between potential benefits and harms. This review found limited evidence to support the conclusion that fusion leads to greater improvement in back pain and function than conventional (nonintensive) physical therapy for adults with low back pain due to degenerated disc(s). In this review the evidence was considered insufficient to determine the benefits of lumbar fusion compared to more intensive rehabilitation programs. However, in their review, Chou et al. concluded that there “is fair evidence that fusion is no more effective than intensive rehabilitation with a cognitive behavioral emphasis.”⁹⁴ Thus, those making decisions regarding surgery for low back pain due to degenerated discs may want to consider the availability of more intensive noninvasive treatments.

Decisionmakers also need to consider the potential for adverse events. Spinal fusion, as with all back surgeries, is associated with a number of adverse events ranging from relatively minor events such as postsurgical bleeding and pain to major events such as deep wound infections and heart failure. However, because of insufficient reporting and variation in surgical methods used in the different studies, the incidence of adverse events (serious and minor) associated with fusion could not be determined conclusively in this report. In a review from ECRI Institute, the authors concluded that lumbar fusion leads to higher rates of both early and late adverse events compared to non-intensive physical therapy or intensive exercise/rehabilitation plus CBT.⁹⁵ In their review, Chou et al. indicated that the most frequently reported major adverse events among the studies they reported included deep wound infections, major bleeding during surgery, thrombosis, acute respiratory distress syndrome, pulmonary edema, and heart failure.⁹⁴

The associated risk of adverse events is likely to vary depending on factors such as patient age and surgical method. Our qualitative review of studies for prognostic factors associated with patient outcomes following fusion surgery suggests that older age (65 years or older) appears to be associated with worse patient outcomes following spinal fusion. Furthermore, limited evidence from one of the included studies in this review found a higher occurrence of complications with more technically difficult procedures.²³ In their study, which evaluated different fusion techniques, Fritzell et al. indicated that more overall complications occurred with circumferential fusion than with instrumented and noninstrumented posterolateral fusion. Finally, while based on limited evidence, the findings of this review indicate that arthroplasty is associated with less blood loss in adults with low back pain due to degenerated disc(s).

Limitations of the Comparative Effectiveness Review Process

Fusion methods no longer used in the United States or not commercially available in the United States were not considered relevant and were, therefore, not reviewed in this report. However, determining which fusion methods were current and available within the United States was challenging. To overcome this challenge, we solicited TEP expertise to help us determine which methods and devices are relevant to current U.S. practices. Although the devices covered in this report have been evaluated for inclusion by members of the TEP, many of the included studies were conducted over a decade ago, which may still limit their applicability to currently used fusion instrumentation (e.g., rods, pedicle and facet screws) and techniques.

Limitations of the Evidence Base

The primary limitations of this review were that no studies were identified for several comparisons of interest and only one or two studies made up the evidence base for other comparisons. Our searches did not identify any studies that met the inclusion criteria for this report that compared spinal fusion and continued noninvasive treatment for patients with degenerative stenosis. Likewise, no studies were identified that met inclusion criteria that compared spinal fusion to other invasive procedures for patients with spondylolisthesis. For most of the questions in this review that considered the comparative effectiveness of different approaches or techniques used in spinal fusion surgery (e.g., open mini ALIF versus laparoscopic ALIF) only one study was identified.

Further, this review was limited to a qualitative analysis of the available evidence. We planned to perform meta-analysis whenever the evidence base for a key question met the following minimum criteria: it consisted of at least two studies addressing the same outcome at the same duration of followup and the studies were clinically similar in terms of patient characteristics, surgical approach and strategy, and comparability of control groups. However, none of the evidence bases met the minimum criteria, so we did not attempt to use meta-analysis to determine summary effect sizes estimates. Our ability to draw evidence based conclusions from our qualitative analysis was limited for many comparisons of interest because the evidence from more than one study was inconsistent and/or imprecise.

Research Gaps

Through our review of the evidence, we identified a number of gaps in the literature that need further research. In particular, research is needed on the benefits and harms of fusion for individuals over 60 years of age. The number of fusion surgeries in this population is growing despite a lack of evidence that surgery is more beneficial than other noninvasive treatments for individuals over 60. In only a few of the studies included in this review was the mean age of the patients over 50 years. Further, our qualitative review of studies evaluating prognostic factors associated with patient outcomes following surgery suggest that older age (>65 years) is associated with poorer patient outcomes.

In general, more studies are needed that focus on identifying patients who are more or less likely to benefit from fusion. Our searches identified only a handful of prognostic studies, and the patient and treatment characteristics evaluated in those studies varied. Further, patient characteristics, particularly patient comorbidities and other health related factors were poorly reported in many of the comparative trials included in the review.

Poor reporting along with variation in the surgical methods used among the comparative trials that addressed key questions 1 to 9 limited our ability to conclusively determine the incidence of adverse events associated with fusion. Thus, more complete reporting of all adverse events (serious and minor) associated with fusion and its comparators is needed in future research. Further, sufficient followup to capture late adverse events is also needed in studies comparing fusion to other invasive procedures.

One overarching problem with the evidence base in this report is the variation in the therapies used in the noninvasive comparator group among the studies that compared fusion to noninvasive alternatives. For instance, in the studies that compared fusion to physical therapy, the physical therapy component varied considerably in terms of intensity, duration, and use of supplemental therapies such as acupuncture, cognitive behavioral therapy, or injections. Efforts to develop a standardized approach to defining and delivering physical therapy would make replication of comparisons between fusion and physical therapy possible. Similarly, clearly describing what therapies patients received in all treatment groups is important to replication of comparisons of fusion to other treatments. In at least one study that made up the evidence base for this report, the specific surgical tools and techniques used in the fusion group were not clearly described and the therapies provided in the nonsurgical group varied and were not fully described.

Overall, more studies are needed that compare fusion to other noninvasive therapies, such as exercise therapy or cognitively-oriented therapies. More studies that compare spinal fusion and noninvasive treatment(s) for patients with degenerative stenosis are also needed. Our searches did not identify any studies that met the inclusion criteria for this report that compared spinal fusion and continued noninvasive treatment for patients with degenerative stenosis. Ideally, future studies would compare certain types of fusion for certain indications to non-operative care.

Our searches did not identify any studies that met inclusion criteria that compared spinal fusion to other invasive procedures for patients with spondylolisthesis. Thus, more studies are needed in this area. Finally, more studies that compare different fusion methods and techniques are needed to clarify optimal surgical procedures. Because implantable devices are frequently replaced by new products and product generations, either by product line updates or withdrawal of previous implants and instrumentation from the market because of adverse events, ongoing clinical studies of new devices and materials are needed. Many of the studies retrieved for our review were ultimately excluded for lack of relevance to modern treatment practices in the United States.

Lastly, for most comparisons considered in this review only one or two studies were identified. However, to support an evidence-based conclusion, replication of findings is generally needed. Replication of comparisons in clinically comparable populations across multiple studies also enables meta-analysis, increases the power of the evidence base to detect a difference between treatments overall, and decreases the likelihood that the overall findings will be imprecise. Future studies ideally would perform randomized assignment of patients to treatment arms and, while blinding of patients and practitioners is not always practical, outcome assessors would be blinded to treatment assignment if possible.

References

1. Manchikanti L. Epidemiology of low back pain. *Pain Physician* 2000;3(2):167-92.
2. Pleis JR, Ward BW, Lucas JW. Summary health statistics for U.S. adults: National Health Interview Survey, 2009. *Vital Health Stat* 10 2010 Dec;(249):1-207. PMID: 21905346
3. National Center for Health Statistics. Health, United States, 2010: with special feature on death and dying. Hyattsville (MD): National Center for Health Statistics; 2011. 547 p.
4. Hu SS, Tribus CB, Tay BK, et al. Chapter 5: Disorders, diseases, and injuries of the spine. Osteomyelitis of the spine. In: Skinner HB, editor. *CURRENT diagnosis and treatment in orthopedics*. 4th ed. New York: McGraw-Hill; 2006. Also available: <http://www.accessmedicine.com/content.aspx?aID=2319334>.
5. Maghout Juratli S, Franklin GM, Mirza SK, et al. Lumbar fusion outcomes in Washington State workers' compensation. *Spine (Phila Pa 1976)* 2006 Nov 1;31(23):2715-23. PMID: 17077741
6. Resnick DK. Spinal fusion for discogenic back pain: patient selection, operative techniques, and outcomes. *Tech Neurosurg* 2003;8(3):176-190.
7. McCrory DC, Turner DA, Patwardhan MB, et al. Spinal fusion for treatment of degenerative disease affecting the lumbar spine [draft]. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2006 Nov 1. 93 p.
8. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 2008 Jan 1;33(1):90-4. PMID: 18165753
9. Owens D, Lohr K, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions. In: Agency for Healthcare Research and Quality (AHRQ). *Methods guide for comparative effectiveness reviews*. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2009 Jul. Also available: http://effectivehealthcare.ahrq.gov/ehec/products/122/328/2009_0805_grading.pdf.
10. Ivar Brox J, Sorensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine (Phila Pa 1976)* 2003 Sep 1;28(17):1913-21. PMID: 12973134
11. Brox JI, Reikeras O, Nygaard O, et al. Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: a prospective randomized controlled study. *Pain* 2006 May;122(1-2):145-55. PMID: 16545523
12. Fritzell P, Hagg O, Wessberg P, et al. 2001 Volvo Award Winner in Clinical Studies: lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine (Phila Pa 1976)* 2001 Dec 1;26(23):2521-34. PMID: 11725230
13. Ohtori S, Koshi T, Yamashita M, et al. Surgical versus nonsurgical treatment of selected patients with discogenic low back pain: a small-sized randomized trial. *Spine (Phila Pa 1976)* 2011 Mar 1;36(5):347-54.
14. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med* 2007 May 31;356(22):2257-70. PMID: 17538085

15. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *J Bone Joint Surg Am* 2009 Jun;91(6):1295-304. PMID: 19487505
16. Zigler J, Delamarter R, Spivak JM, et al. Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine (Phila Pa 1976)* 2007 May 15;32(11):1155-62; discussion 1163. PMID: 17495770
17. Delamarter R, Zigler JE, Balderston RA, et al. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement compared with circumferential arthrodesis for the treatment of two-level lumbar degenerative disc disease: results at twenty-four months. *J Bone Joint Surg Am* 2011 Apr 20;93(8):705-15. PMID: 21398574
18. Hallett A, Huntley JS, Gibson JN. Foraminal stenosis and single-level degenerative disc disease: a randomized controlled trial comparing decompression with decompression and instrumented fusion. *Spine (Phila Pa 1976)* 2007 Jun 1;32(13):1375-80. PMID: 17545903
19. Dawson E, Bae HW, Burkus JK, et al. Recombinant human bone morphogenetic protein-2 on an absorbable collagen sponge with an osteoconductive bulking agent in posterolateral arthrodesis with instrumentation: a prospective randomized trial. *J Bone Joint Surg Am* 2009 Jul 1;91(7):1604-13.
20. Burkus JK, Transfeldt EE, Kitchel SH, et al. Clinical and radiographic outcomes of anterior lumbar interbody fusion using recombinant human bone morphogenetic protein-2. *Spine (Phila Pa 1976)* 2002 Nov 1;27(21):2396-408. PMID: 12438990
21. Burkus JK, Gornet MF, Dickman CA, et al. Anterior lumbar interbody fusion using rhBMP-2 with tapered interbody cages. *J Spinal Disord Tech* 2002 Oct;15(5):337-49. PMID: 12394656
22. Chung SK, Lee SH, Lim SR, et al. Comparative study of laparoscopic L5-S1 fusion versus open mini-ALIF, with a minimum 2-year follow-up. *Eur Spine J* 2003 Dec;12(6):613-7. PMID: 14564558
23. Fritzell P, Hagg O, Wessberg P, et al. Chronic low back pain and fusion: a comparison of three surgical techniques: a prospective multicenter randomized study from the Swedish lumbar spine study group. *Spine (Phila Pa 1976)* 2002 Jun 1;27(11):1131-41.
24. Sasso RC, Burkus JK, LeHuec JC. Retrograde ejaculation after anterior lumbar interbody fusion: transperitoneal versus retroperitoneal exposure. *Spine (Phila Pa 1976)* 2003 May 15;28(10):1023-6. PMID: 12768143
25. Korovessis P, Koureas G, Zacharatos S, et al. Correlative radiological, self-assessment and clinical analysis of evolution in instrumented dorsal and lateral fusion for degenerative lumbar spine disease. Autograft versus coralline hydroxyapatite. *Eur Spine J* 2005 Sep;14(7):630-8. PMID: 15789231
26. Fischgrund JS, Mackay M, Herkowitz HN, et al. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal instrumentation. *Spine (Phila Pa 1976)* 1997;22(24):2807-2812.
27. Abdu WA, Lurie JD, Spratt KF, et al. Degenerative spondylolisthesis: does fusion method influence outcome? Four-year results of the spine patient outcomes research trial. *Spine (Phila Pa 1976)* 2009 Oct 1;34(21):2351-60. PMID: 19755935

28. Fernandez-Fairen M, Sala P, Ramirez H, et al. A prospective randomized study of unilateral versus bilateral instrumented posterolateral lumbar fusion in degenerative spondylolisthesis. *Spine (Phila Pa 1976)* 2007 Feb 15;32(4):395-401. PMID: 17304127
29. Hagg O, Fritzell P, Ekselius L, et al. Predictors of outcome in fusion surgery for chronic low back pain. A report from the Swedish Lumbar Spine Study. *Eur Spine J* 2003 Feb;12(1):22-33.
30. Fukuta S, Miyamoto K, Hosoe H, et al. Kidney-type intervertebral spacers should be located anteriorly in cantilever transforaminal lumbar interbody fusion: analyses of risk factors for spacer subsidence for a minimum of 2 years. *J Spinal Disord Tech* 2011 May;24(3):189-95. PMID: 20634726
31. Keorochana G, Laohacharoensombat W, Wajanavisit W, et al. Functional outcome after decompression and instrumented arthrodesis in degenerative lumbar spinal stenosis: factors influencing unsuccessful outcome change. *J Med Assoc Thai* 2011 Dec;94(12):1487-94. PMID: 22295737
32. Carreon LY, Glassman SD, Djurasovic M, et al. Are preoperative health-related quality of life scores predictive of clinical outcomes after lumbar fusion? *Spine (Phila Pa 1976)* 2009 Apr 1;34(7):725-30. PMID: 19333106
33. Okuda S, Oda T, Miyauchi A, et al. Surgical outcomes of posterior lumbar interbody fusion in elderly patients. *J Bone Joint Surg Am* 2006 Dec;88(12):2714-20.
34. Schuler TC, Burkus JK, Gornet MF, et al. The correlation between preoperative disc space height and clinical outcomes after anterior lumbar interbody fusion. *J Spinal Disord Tech* 2005 Oct;18(5):396-401.
35. Modic MT, Ross JS. Lumbar degenerative disk disease. *Radiology* 2007 Oct;245(1):43-61. PMID: 17885180
36. Patel RK. Lumbar degenerative disk disease clinical presentation. In: Medscape Reference [internet]. New York (NY): WebMD LLC; 2010 Nov 5 [accessed 2012 Aug 14]. [12 p]. Available: <http://emedicine.medscape.com/article/309767-clinical>.
37. Vernon-Roberts B, Moore RJ, Fraser RD. The natural history of age-related disc degeneration: the pathology and sequelae of tears. *Spine (Phila Pa 1976)* 2007 Dec 1;32(25):2797-804. PMID: 18246000
38. Hsiang J, Furman MB, Keenan ME, et al. Spinal stenosis. In: eMedicine: Medscape Reference [internet]. WebMD, LLC [updated 2011 Jun 22]. [accessed 2011 Jul 7]. [53 p]. Available: <http://emedicine.medscape.com>.
39. Froese BB. Lumbar spondylolysis and spondylolisthesis. In: eMedicine: Medscape Reference [internet]. WebMD, LLC [updated 2009 Jul 5]. [accessed 2011 Jul 7]. [16 p]. Available: <http://emedicine.medscape.com>.
40. Kalichman L, Hunter DJ. Diagnosis and conservative management of degenerative lumbar spondylolisthesis. *Eur Spine J* 2008 Mar;17(3):327-35. PMID: 18026865
41. Licciardone JC. The epidemiology and medical management of low back pain during ambulatory medical care visits in the United States. *Osteopath Med Prim Care* 2008;2:11. PMID: 19025636
42. Patel RK, Slipman CW. Lumbar degenerative disk disease. In: eMedicine: Medscape Reference [internet]. WebMD, LLC [updated 2010 Nov 5]. [accessed 2011 Jul 7]. [49 p]. Available: <http://emedicine.medscape.com>.
43. Institute for Clinical Systems Improvement (ICSI). Adult low back pain. Bloomington (MN): Institute for Clinical Systems Improvement (ICSI); 2010 Nov. 74 p.

44. Chou R, Qaseem A, Snow V, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007 Oct 2;147(7):478-91. PMID: 17909209
45. North American Spine Society (NASS). Diagnosis and treatment of degenerative lumbar spondylolisthesis. Burr Ridge (IL): North American Spine Society (NASS); 2008. 133 p.
46. North American Spine Society (NASS). Diagnosis and treatment of degenerative lumbar spinal stenosis. Burr Ridge (IL): North American Spine Society (NASS); 2007 Jan. 262 p.
47. van Middelkoop M, Rubinstein SM, Verhagen AP, et al. Exercise therapy for chronic nonspecific low-back pain. *Baillieres Best Pract Res Clin Rheumatol* 2010 Apr;24(2):193-204. PMID: 20227641
48. Mayer J, Mooney V, Dagenais S. Evidence-informed management of chronic low back pain with lumbar extensor strengthening exercises. *Spine J* 2008 Jan-Feb;8(1):96-113. PMID: 18164458
49. Ferreira ML, Smeets RJ, Kamper SJ, et al. Can we explain heterogeneity among randomized clinical trials of exercise for chronic back pain? A meta-regression analysis of randomized controlled trials. *Phys Ther* 2010 Oct;90(10):1383-403. PMID: 20671101
50. Standaert CJ, Weinstein SM, Rumpeltes J. Evidence-informed management of chronic low back pain with lumbar stabilization exercises. *Spine J* 2008 Jan-Feb;8(1):114-20. PMID: 18164459
51. Mayer JM, Mooney V, Matheson LN, et al. Continuous low-level heat wrap therapy for the prevention and early phase treatment of delayed-onset muscle soreness of the low back: a randomized controlled trial. *Arch Phys Med Rehabil* 2006 Oct;87(10):1310-7. PMID: 17023239
52. van Middelkoop M, Rubinstein SM, Kuijpers T, et al. A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. *Eur Spine J* 2011 Jan;20(1):19-39. PMID: 20640863
53. van Geen JW, Edelaar MJ, Janssen M, et al. The long-term effect of multidisciplinary back training: a systematic review. *Spine (Phila Pa 1976)* 2007 Jan 15;32(2):249-55. PMID: 17224822
54. Gatchel RJ, Rollings KH. Evidence-informed management of chronic low back pain with cognitive behavioral therapy. *Spine J* 2008 Jan-Feb;8(1):40-4. PMID: 18164452
55. Gatchel RJ, Robinson RC. Pain management. In: O'Donohue W, Fisher JE, Hayes SC, editors. *Cognitive behavior therapy: applying empirically supported techniques in your practice*. New York (NY): John Wiley & Sons; 2003. p. 273-9.
56. Vlaeyen JW, Morley S. Cognitive-behavioral treatments for chronic pain: what works for whom? *Clin J Pain* 2005 Jan-Feb;21(1):1-8. PMID: 15599126
57. Back pain facts & statistics. [internet]. Arlington (VA): American Chiropractic Association; [accessed 2012 Jun 22]. [3 p]. Available: http://www.acatoday.org/pdf/back_pain.pdf.
58. New study shows chiropractic is cost-effective in treating chronic back pain. [internet]. Arlington (VA): American Chiropractic Association; 2005 Nov 15 [accessed 2012 Jun 22]. [1 p]. Available: http://www.acatoday.org/content_css.cfm?CID=636.
59. Furlan AD, Imamura M, Dryden T, et al. Massage for low-back pain. *Cochrane Database Syst Rev* 2008;(4):CD001929. PMID: 18843627
60. Imamura M, Furlan AD, Dryden T, et al. Evidence-informed management of chronic low back pain with massage. *Spine J* 2008 Jan-Feb;8(1):121-33. PMID: 18164460

61. Gatchel RJ, Mayer TG. Evidence-informed management of chronic low back pain with functional restoration. *Spine J* 2008 Jan-Feb;8(1):65-9. PMID: 18164455
62. Szpalski M, Gunzburg R, Mayer M. Spine arthroplasty: a historical review. *Eur Spine J* 2002 Oct;11 Suppl 2:S65-84. PMID: 12384726
63. van den Eerenbeemt KD, Ostelo RW, van Royen BJ, et al. Total disc replacement surgery for symptomatic degenerative lumbar disc disease: a systematic review of the literature. *Eur Spine J* 2010 Aug;19(8):1262-80. PMID: 20508954
64. CMS Manual System, Pub 100-03 Medicare National Coverage Determinations Manual. Baltimore (MD): Centers for Medicare & Medicaid Services (CMS); 2007 Sep 11. Lumbar artificial disc replacement (LADR).
65. Laminectomy. In: MedlinePlus [internet]. Bethesda (MD): National Library of Medicine (U.S.) [accessed 2012 Jun 25]. [3 p]. Available: <http://www.nlm.nih.gov/medlineplus/ency/article/007389.htm>.
66. Lumbar spondylolysis and spondylolisthesis rehabilitation program. In: Medscape Reference [internet]. New York (NY): WebMD, LLC; 2012 Jan 18 [accessed 2012 Jun 25]. [3 p]. Available: <http://emedicine.medscape.com/article/310235-treatment>.
67. Lumbar zygapophysial (facet) joint injections. [internet]. Burr Ridge (IL): North American Spine Society; [accessed 2012 Jun 25]. [3 p]. Available: http://www.knowyourback.org/Pages/Treatments/InjectionTreatments/LZJ_Injections.aspx.
68. Henschke N, Kuijpers T, Rubinstein SM, et al. Injection therapy and denervation procedures for chronic low-back pain: a systematic review. *Eur Spine J* 2010 Sep;19(9):1425-49. PMID: 20424870
69. Health Net. National medical policy no. NMP48. Intradiscal electrothermal therapy (IDET) percutaneous intradiscal radiofrequency thermocoagulation. [internet]. Los Angeles (CA): Health Net; 2011 Jan [accessed 2012 Jun 25]. [14 p].
70. Narine L, Yee DS, Einarson TR, et al. Quality of abstracts of original research articles in CMAJ in 1989. *CMAJ* 1991 Feb 15;144(4):449-53. PMID: 1993292
71. Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. *JAMA* 1999 Mar 24-31;281(12):1110-1. PMID: 10188662
72. Lilford R, Braunholtz D, Harris J, et al. Trials in surgery. *Br J Surg* 2004 Jan;91(1):6-16. PMID: 14716788
73. Carragee EJ, Cheng I. Minimum acceptable outcomes after lumbar spinal fusion. *Spine J* 2010 Apr;10(4):313-20. PMID: 20362247
74. Maughan EF, Lewis JS. Outcome measures in chronic low back pain. *Eur Spine J* 2010 Sep;19(9):1484-94. PMID: 20397032
75. Lauridsen HH, Hartvigsen J, Manniche C, et al. Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. *BMC Musculoskelet Disord* 2006;7:82. PMID: 17064410
76. Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)* 2000 Nov 15;25(22):2940-52; discussion 2952. PMID: 11074683
77. Copay AG, Glassman SD, Subach BR, et al. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. *Spine J* 2008 Nov-Dec;8(6):968-74. PMID: 18201937
78. AHRQ EHCP methods guide for comparative effectiveness reviews of medical interventions. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2010 Nov. 207 p.

79. Treadwell JR, Tregear SJ, Reston JT, et al. A system for rating the stability and strength of medical evidence. *BMC Med Res Methodol* 2006;6:52. Also available: <http://www.biomedcentral.com/1471-2288/6/52>. PMID: 17052350
80. Armitage P, Berry G. *Statistical methods in medical research*. 3rd ed. Oxford, England: Blackwell Scientific; 1994. Significance tests and confidence intervals. p. 94-9.
81. Braitman LE. Confidence intervals assess both clinical significance and statistical significance [editorial]. *Ann Intern Med* 1991 Mar 15;114(6):515-7. PMID: 1994799
82. Fairbank J, Frost H, Wilson-MacDonald J, et al. Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *BMJ* 2005 May 28;330(7502):1233. PMID: 15911537
83. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med* 2008 Feb 21;358(8):794-810. PMID: 18287602
84. Birkmeyer NJ, Weinstein JN, Tosteson AN, et al. Design of the Spine Patient outcomes Research Trial (SPORT). *Spine (Phila Pa 1976)* 2002 Jun 15;27(12):1361-72. PMID: 12065987
85. Longo UG, Loppini M, Denaro L, et al. Rating scales for low back pain. *Br Med Bull* 2010;94:81-144. PMID: 20064820
86. Fukui M, Chiba K, Kawakami M, et al. JOA Back Pain Evaluation Questionnaire (JOABPEQ)/JOA Cervical Myelopathy Evaluation Questionnaire (JOACMEQ). The report on the development of revised versions. April 16, 2007. *J Orthop Sci* 2009 May;14(3):348-65. Also available: http://www.joa.or.jp/english/oat/pdf/user_guide.pdf. PMID: 19499305
87. Bigos SJ. Commentary & perspective on “Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis: four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts”. *Journal of Bone and Joint Surgery, Inc.*; 2009 Jun. 5 p.
88. Zigler J, MD. (Texas Back Institute). Personal communication. Single-level ProDisc IDE/FDA study unpublished data. 2012. 1 p.
89. Zigler JE. Five-year results of the ProDisc-L multicenter, prospective, randomized, controlled trial comparing ProDisc-L with circumferential spinal fusion for single-level disabling degenerative disk disease. *Semin Spine Surg* 2012 Mar;24(1):25-31.
90. Ziegler J. (Texas Spine Institute). Personal communication. 2012. 1 p.
91. Carragee EJ, Hurwitz EL, Weiner BK. A critical review of recombinant human bone morphogenetic protein-2 trials in spinal surgery: emerging safety concerns and lessons learned. *Spine J* 2011 Jun;11(6):471-91. PMID: 21729796
92. Medtronic Sofamor Danek, Inc. Summary of safety and effectiveness data. InFUSE bone graft/LT-CAGE lumbar tapered fusion device. P000058. Rockville (MD): U.S. Food and Drug Administration (FDA); 2002 Jul 2. 40 p. Also available: http://www.accessdata.fda.gov/cdrh_docs/pdf/P000058b.pdf.
93. Umans B, Nonnemaker KL. The Medicare beneficiary population. Washington (DC): AARP Public Policy Institute; 2009 Jan. 4 p. (Fact sheet; no. 149). Also available: http://assets.aarp.org/rgcenter/health/fs149_medicare.pdf.
94. Chou R, Baisden J, Carragee EJ, Resnick DK, Shaffer WO, Loeser JD. Surgery for low back pain: a review of the evidence for an American Pain Society Clinical Practice Guideline. *Spine (Phila Pa 1976)* 2009 May 1;34(10):1094-109. PMID: 19363455

95. ECRI Institute. Spinal fusion and discography for chronic low back pain and uncomplicated lumbar degenerative disc disease (Prepared by ECRI Institute for Washington State Health Care Authority under Contract No. 6020-001788). Olympia (WA): Washington State Health Care Authority; 2007 Oct 19. 219 p. Also available: <http://www.hta.hca.wa.gov>.
96. Atkins D, Chang S, Gartlehner G, et al. Assessing the applicability of studies when comparing medical interventions [AHRQ Publication No. 11-EHC019-EF]. In: Methods guide for effectiveness and comparative effectiveness reviews. Rockville (MD): Agency for Healthcare Research and Quality; 2011 Jan.
97. Anderson GF. Medicare and chronic conditions. *N Engl J Med* 2005 Jul 21;353(3):305-9. PMID: 16034019
98. AcroMed Corporation. 510(k) summary for AcroMed Pedicle Screw (4.75 mm). K965102. [internet]. Silver Spring (MD): U.S. Food and Drug Administration (FDA); 1997 May 8 [accessed 2012 Jun 21]. [2 p].
99. DePuy AcroMed, Inc. Summary of safety and effectiveness data for Lumbar I/F Cage with VSP Spine System. P960025. Rockville (MD): U.S. Food and Drug Administration (FDA); 1999 Feb 2. 29 p. Also available: http://www.accessdata.fda.gov/cdrh_docs/pdf/P960025b.pdf.
100. Medtronic Sofamor Danek, Inc. USA. Summary of safety and effectiveness. CD Horizon spinal system. K060615. Silver Spring (MD): U.S. Food and Drug Administration (FDA); 2006 Oct 25. 7 p. Also available: http://www.accessdata.fda.gov/cdrh_docs/pdf6/K060615.pdf?utm_campaign=Google2&utm_source=fdaSearch&utm_medium=website&utm_term=cd%20horizon%20spinal%20system&utm_content=1.
101. Premarket Approval (PMA) for Infuse Bone Graft, filler, recombinant human bone morphogenetic protein, collagen scaffold with metal prosthesis, osteoinduction. P000058. [internet]. Silver Spring (MD): U.S. Food and Drug Administration (FDA); 2002 Jul 2 [accessed 2012 Jun 21]. [2 p]. Available: <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMA/pma.cfm?id=11126>.
102. Important information on the Infuse/Mastergraft posterolateral revision device. Memphis (TN): Medtronic Sofamor Danek, USA, Inc.; 26 p. Also available: http://www.accessdata.fda.gov/cdrh_docs/pdf4/H040004c.pdf.
103. DePuy Inc. 510(k) summary of safety and effectiveness for DePuy AcroMed MOSS Miami spinal system. K983583 . Silver Spring (MD): U.S. Food and Drug Administration (FDA); 4 p. Also available: <http://www.fda.gov>.
104. Synthes Spine. Summary of safety and effectiveness for PRODISC-L total disc replacement. P050010 . Silver Spring (MD): U.S. Food and Drug Administration (FDA); 2006 Aug 14. 23 p. Also available: <http://www.fda.gov>.
105. Interpore Cross International. 510(k) summary for Pro Osteon 500R resorbable bone graft substitute. K990131. [internet]. Silver Spring (MD): U.S. Food and Drug Administration (FDA); 1999 Jan 13 [accessed 2012 Jun 21]. [5 p]. Available: http://www.accessdata.fda.gov/cdrh_docs/pdf/K990131.pdf.
106. Howmedica Osteonics Corp. Special 510(k) summary of safety and effectiveness: line extension to the Xia spine system. Silver Spring (MD): U.S. Food and Drug Administration (FDA); 2001 Dec 19. 5 p. Also available: <http://www.fda.gov>.
107. Parker SL, Mendenhall SK, Shau DN, et al. Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. *J Neurosurg Spine* 2012 Feb 10. PMID: 22324801

108. Hagg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003 Feb;12(1):12-20. PMID: 12592542
109. Parker SL, Adogwa O, Paul AR, et al. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. *J Neurosurg Spine* 2011 May;14(5):598-604. PMID: 21332281
110. Spratt KF. Patient-level minimal clinically important difference based on clinical judgment and minimally detectable measurement difference: a rationale for the SF-36 physical function scale in the SPORT intervertebral disc herniation cohort. *Spine (Phila Pa 1976)* 2009 Jul 15;34(16):1722-31. PMID: 19561562
111. Scott NA, et al. Systematic review of off-pump coronary artery bypass surgery with the aid of the Octopus Tissue Stabilizer: update and re-appraisal. Adelaide, South Australia: Australia Safety & Efficacy Register of New Interventional Procedures - Surgical (ASERNIP-S); 2001 Oct. (ASERNIP-S Report ; no. 20).