

Wide Variability in Patient-Reported Outcomes Measures After Fusion for Lumbar Spinal Stenosis: A Systematic Review

James P. Winebrake, BS¹ , Francis Lovecchio, MD², Michael Steinhaus, MD², James Farmer, MD², and Andrew Sama, MD²

Global Spine Journal
2020, Vol. 10(2) 209-215
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2192568219832853
journals.sagepub.com/home/gsj



Abstract

Study Design: Systematic review.

Objectives: The purpose of this study is to review outcomes reporting methodology in studies evaluating fusion for lumbar spinal stenosis.

Methods: A systematic review of PubMed and Embase databases was conducted from January 2007 to June 2017 for English language studies with minimum of 2 years postoperative follow-up reporting outcomes after fusion for lumbar spinal stenosis. Two reviewers assessed each study; those meeting inclusion criteria were examined for pertinent data. Outcome measures were categorized into relevant domains: pain/symptomatology, function/disability, and surgical satisfaction. Return to work reporting was also recorded.

Results: Of 123 studies meeting inclusion criteria, 76% included posterior-only fusion, 32% included posterior/transforaminal interbody fusion, and 5% included anterior/lateral interbody fusion (non-mutually exclusive). There was significant variation in patient-reported outcomes (PROs) used—studies reported 31 unique PROs assessing at least one domain: 22 evaluating pain, 23 evaluating function, and 3 evaluating surgical satisfaction. Most commonly utilized PROs were the Oswestry Disability Index (73% of studies), Visual Analog Scale (55%), and 36-Item Short Form Survey (32%). The remaining 28 measures were used in 14% of studies or fewer. PROs specific to symptoms of lumbar spinal stenosis, such as the Zurich Claudication Questionnaire, were only used rarely (7/123 studies). Only 14% of studies reported on time to return to work.

Conclusions: The literature surrounding fusion in the setting of lumbar stenosis is characterized by substantial variability in outcomes reporting. Very few studies utilized measures specific to lumbar spinal stenosis. Efforts to standardize outcomes reporting would facilitate comparisons of surgical interventions.

Keywords

spinal stenosis, visual analog scale, return to work, pain measurement, lumbosacral region, fusion, decompression, lumbar interbody fusion

Introduction

The number of patients with lumbar spinal stenosis (LSS) treated with fusion has been rising rapidly. A recent epidemiological study of the Nationwide Inpatient Sample found that the rate of decompression and fusion for a diagnosis of LSS rose from 21.5% to 31.5% between 2004 and 2009, while the rate of decompression alone dropped from 58.5% to 49.2%.¹ While surgery has been shown to be superior to conservative management in many studies,²⁻⁴ other investigations have not been as

¹ Weill Cornell Medicine, New York, NY, USA

² The Hospital for Special Surgery, New York, NY, USA

Corresponding Author:

James P. Winebrake, Cornell University Joan and Sanford I Weill Medical College, 420 East 70th Street, 13K-2, New York, NY 10021, USA.
Email: jpw2005@med.cornell.edu



definitive in their results.^{5,6} The ability to measure the effectiveness of surgery is likely hampered by a wide variety of surgical techniques used to treat LSS and assortment of outcomes measures used to define results.⁷ A 2016 Cochrane review was unable to definitively conclude that surgery was superior, due to the heterogeneity of interventions and lack of standardized outcomes measures.⁸

The future of health care delivery is based on value—a measure of health outcomes achieved per dollar spent.⁹ Value will dictate both whether surgery for LSS is superior to conservative management and whether one surgical intervention is superior to another. Thus, the accurate determination of value must be based on standardized outcomes. Patient-reported outcomes (PROs) form the basis on which value will be determined.⁹ The National Institute of Health's effort to define PROs through its PROMIS initiative is just one example of how powerful entities in health care are emphasizing PROs to define value.¹⁰

Multiple high-quality analyses on surgical interventions for LSS have employed PROs in their determination of effectiveness.^{3,4,11-13} Unfortunately, a wide variety of PROs are used in spine surgery research,⁷ limiting our ability to compare results between surgical techniques. As payer systems move toward value-based care, the need to categorize and understand PRO instruments used to evaluate the postoperative results of common interventions becomes ever more pressing. The purpose of this study is to identify and characterize the PROs utilized in evaluating the efficacy of fusion for LSS. We hypothesize that there are a wide variety of PROs used, and that the majority of utilized PROs are too broad, and thus inadequate, to best assess results after lumbar fusion for spinal stenosis.

Materials and Methods

The PubMed and Embase computerized databases were systematically searched to identify all literature in the last decade (January 2007 through June 2017) reporting outcomes after spinal fusion for lumbar stenosis (Table 1). All articles were retrieved by an electronic search of Medical Subject Headings and keyword terms and their respective combinations (Figure 1).¹⁴⁻¹⁶ Inclusion criteria consisted of any study recording patient-reported clinical outcomes after lumbar fusion for spinal stenosis. Studies that compared fusion cohorts with those undergoing nonsurgical management, decompression alone, or dynamic stabilization were also included. Exclusion criteria included follow-up of less than 24 months; animal, biomechanics, cadaveric, and basic science studies; review articles; surgical technique guides; and case reports. Also excluded were studies for which English full text was not available.

The literature search is outlined in Figure 1. The initial title search yielded a subset of possible articles that were then further included or excluded according to the contents of the article's abstract, wherein articles were again selected based on inclusion and exclusion criteria. The full text was reviewed of articles selected in both the title and abstract phase. Appropriate studies for final inclusion were then selected. The title, abstract, and full-text selection process with assessment of bias at the study level was

Table 1. Database Search for Systematic Review^a.

| Database | Search Terms |
|----------------|--|
| PubMed, Embase | Keyword: “((spinal stenosis) OR lumbar stenosis) AND fusion” |

^aSearch terms entered into PubMed and Embase search engines to identify English language studies from January 2007 to June 2017.

performed independently by 2 study authors (JPW and FL) with any discrepancies discussed and resolved by mutual agreement.

Several metrics were collected from each study to describe the patient population, including level of evidence per Oxford Center for Evidence-Based Medicine (OCEBM) criteria,¹⁷ study design, number of patients at baseline and follow-up, mean patient age, gender distribution, inclusion of patients with spondylolisthesis (Meyerding grade I or II only, grades III and above were not reported by any study), follow-up time, and type of fusion surgery (as well as use of nonfusion controls). When fusion method was not specified, it was assumed as posterolateral fusion given the prevalence of this method.¹⁸ Moreover, a majority of assessed studies specifying fusion method cited use of this technique as opposed to interbody or other approaches. Proportions were calculated based only on studies that reported the given metric, as several studies lacked number of patients at follow-up, gender distribution, and age.

Patient-Reported Outcomes

Outcomes of interest encompassed any validated PROs recorded in the included studies. PROs were further classified based on multiple domains: pain/symptomatology, function/disability, and surgical satisfaction (Figure 2). Several PROs were questionnaires that captured outcomes in 2 or all domains. Documentation of return to work and/or baseline activity was recorded as a marker of function/disability.

Results

Study Characteristics

Study characteristics are reported in Table 2. This systematic review included 123 published works: 4 Level I studies (3%), 25 Level II studies (20%), 48 Level III studies (39%), and 46 Level IV studies (37%) per OCEBM criteria.¹⁷ The mean number of patients in each study was 373.9 (median, 94; range, 17-8142) at baseline and 311.7 (median, 91; range, 17-5390) at most recent postoperative follow-up. Minimum follow-up time averaged 37.1 months (range, 24-120). Several studies reported multiple methods of fusion—76% of studies assessed PROs following posterior or posterolateral fusion, while 32% assessed PROs following posterior or transforaminal interbody fusion. Only 5% assessed PROs following anterior or lateral interbody fusion. Overall, 31 distinct PROs were used among the 123 studies (Figure 2).

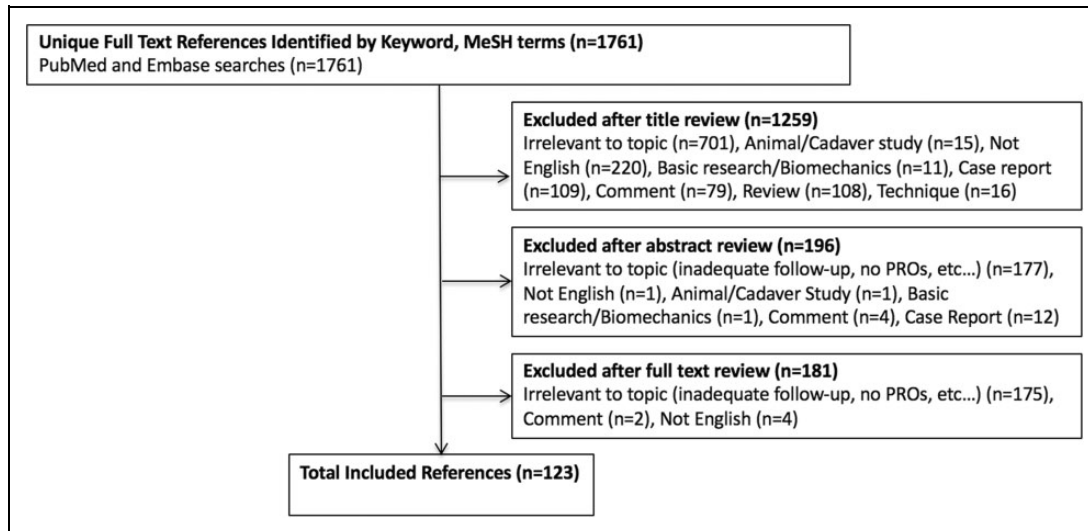


Figure 1. Flow diagram representing the systematic review process used in this study. A total of 123 studies were included for final analysis.

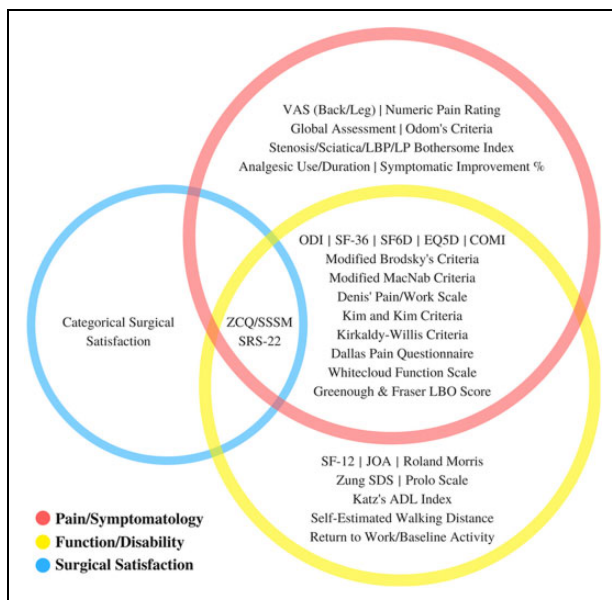


Figure 2. All (31) identified patient-reported outcomes measures by domain reported in studies (n = 123) included for review. Abbreviations: VAS, Visual Analog Scale; LBP, low back pain; LP, leg pain; ODI, Oswestry Disability Index; SF-36, 36-Item Short Form Survey; EQ5D, European Quality of Life-5 Dimensions; COMI, Core Outcome Measures Index; LBO, Low Back Outcome; SF-12, 12-Item Short Form Health Survey; JOA, Japanese Orthopedic Association Score; Zung SDS, Zung Self-rating Depression Scale; ADL, activities of daily living; ZCQ/SSSM, Zurich Claudication Questionnaire/Swedish Spinal Stenosis Measure; SRS-22, Scoliosis Research Society Outcomes Questionnaire.

PRO Measures

Figure 3 portrays the distribution of PRO measure utilization among studies. The mean number of reported measures was 2.76 (range, 1-8). Twenty percent of studies reported a single PRO measure and 30% reported 2 measures. The remaining

Table 2. Study Characteristics^a.

| Characteristic | Mean (Range) or n (%) |
|--|-----------------------|
| Patients at baseline per study, n (n = 123) | 373.9 (17-8142) |
| Patients at follow-up per study, n (n = 123) | 311.7 (17-5390) |
| Patient age, years (n = 117) | 62.8 (43.2-82.5) |
| Male patients (n = 115) | 15 294 (39.9%) |
| Method of fusion (n = 123) | |
| PLF | 93 (76%) |
| PLIF/TLIF | 39 (32%) |
| ALIF/LLIF | 6 (5%) |
| Dynamic stabilization | 12 (10%) |
| Spondylolisthesis (Grade I or II) (n = 123) | |
| Yes | 83 |
| No | 7 |
| Unspecified | 33 |
| Level of evidence (n = 123) | |
| I | 4 |
| II | 25 |
| III | 48 |
| IV | 46 |
| Study design (n = 123) | |
| Prospective | 40 |
| Retrospective | 83 |
| Years (n = 123) | |
| 2007-2011 | 51 |
| 2012-2017 | 72 |
| Minimum follow-up time, months (n = 124) | 37.1 (24-120) |

Abbreviations: PLF, posterior/posterolateral lumbar fusion; PLIF, posterior lumbar interbody fusion; TLIF, transforaminal lumbar interbody fusion; ALIF, anterior lumbar interbody fusion; LLIF, lateral lumbar interbody fusion.

^a Includes demographics of the patients included in the studies reviewed as well as the overall characteristics of the studies included in the review.

50% of studies presented 3 or more PRO measures for assessment.

The 10 most frequently cited PRO measures among the 123 reviewed studies are depicted in Figure 4. The Oswestry Disability Index (ODI), Visual Analog Scale (VAS), and 36-Item Short Form Survey (SF-36) were utilized most often—73%,

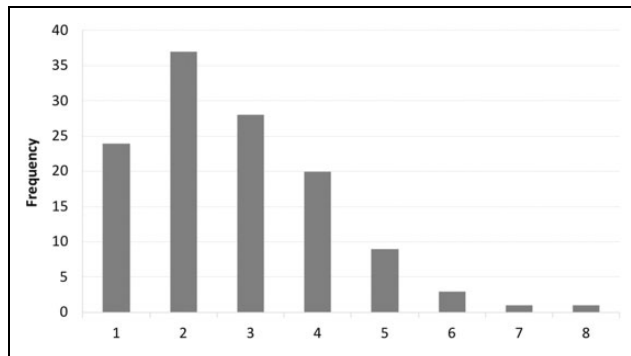


Figure 3. Number of PRO measures reported per study.

55%, and 32% of studies, respectively. No other measure was used in more than 14% of the analyzed studies. VAS was the most prominent measure reporting pain/symptomatology alone (back pain, leg pain, or both), whereas the Japanese Orthopedic Association Score (JOA) was the most prominent measure reporting solely function/disability (12%). Similarly, surgical satisfaction was usually assessed via categorical approach using 2-point, 5-point, or other Likert-type scales (13%).

Temporal Trends

Overall, the number of publications per year citing PRO measures as a means of evaluating outcomes following fusion for lumbar stenosis increased from 10.2 (2007 to 2011) to approximately 13.1 (2012 to mid-2017). Table 3 stratifies studies by the aforementioned time intervals and ranks measures by frequency of use. Furthermore, the mean number of PRO measures reported per study was significantly greater in studies published between 2012 and 2017 (3.08) compared to those published between 2007 and 2011 (2.31) ($P = .003$).

Discussion

Understanding the value of a surgical intervention depends on the outcome measures used. While radiographic and objective outcomes play a role in determining value, trends in health care payment models suggest that PROs will be increasingly important. Surgical interventions for LSS are already highly criticized in the medical community given conflicting data on outcomes,¹⁹ so the ability to evaluate these interventions is increasingly important and dependent on standardized and validated PROs. On the subject of said clinical outcomes, we recognize that the predominant limitation to this review is based on potential for publication bias across and selective reporting bias within individual studies. That said, we expected each to be at least partially mitigated by the fact that we were assessing for variability in reporting methods and measures, rather than the clinical outcomes themselves.

We hypothesized that there would be a wide variety of PROs used, and that many of the PROs would be too general to account for outcomes after lumbar spine fusion. Our review

of 123 recently published long-term studies on outcomes after fusion for LSS confirms the first part of our hypothesis. Even though we purposefully narrowed our search to a specific type of surgical intervention (fusion) for a specific diagnosis (LSS), and only included studies with greater than 2 years of follow-up, there were an overwhelming number of PROs used. In addition, the number of PROs reported per study has also increased over time. With regard to the second part of our hypothesis, the most popular PROs were not specific to outcomes after lumbar spine surgery.

Prior work has demonstrated the wide variability of PROs in spinal surgery. Guzman et al reviewed the frequency, trends, and methods of utilization of various spine-related PROs from 2004 to 2013, and also came to the conclusion that an extensive variety of PROs are utilized in spine surgery. Unlike our review, they chose articles on any form of spine surgery (lumbar, thoracic, cervical) for any diagnosis, and included articles from only 5 orthopedic journals (neurosurgical journals were excluded), identifying 206 unique PROs in 1079 spine surgery articles.⁷ Despite the use of a much broader search, the most commonly used PROs were the same as those identified in our study—the VAS, ODI, and SF-36. Similarly, Yadla et al performed a systematic review of outcomes studies on adult spinal deformity (ASD), finding that research on ASD is similarly plagued by a wide variety of PROs without standardization of outcomes instruments.²⁰ Finally, Ueda et al assessed PROs in degenerative cervical spine surgery and reported comparable heterogeneity, with 53 total outcomes reported.²¹ The authors of these studies conclude that there is a need for greater standardization and guidelines to specify which instruments should be used for a given spinal disease or treatment. While it is less surprising that there are heterogeneous outcomes reported across various spinal pathologies in the study by Guzman et al, our findings of similar heterogeneity in a focused search in LSS confirms these concerns and highlights the need for standardization of outcomes.

Similarly to the findings in Guzman et al,⁷ the 3 most popular PROs used to evaluate lumbar spine fusion for spinal stenosis were the ODI, SF-36, and VAS pain scale, utilized in 73%, 55%, and 32% of studies, respectively. The most popular of the PROs, the Oswestry Disability Index, was first published and validated in 1980 by Fairbank et al as a measurement of disability in patients with chronic back pain.²² The ODI is available in multiple languages, and has only undergone one minor update since its inception—this popularity likely explains its wide use. The most detailed evaluation of the tool's face-content (ie, the instrument's effectiveness in measuring outcomes as intended or claimed) was conducted in patients with chronic back pain²³ and was found to be nonspecific to spine pathology. Multiple trials have demonstrated that leg pain, rather than chronic back pain, is more predictably relieved by surgery for LSS. Thus, the common use of a PRO designed to measure disability secondary to chronic back pain may very well underestimate the effect of surgery for LSS and therefore may not be the most appropriate measure.

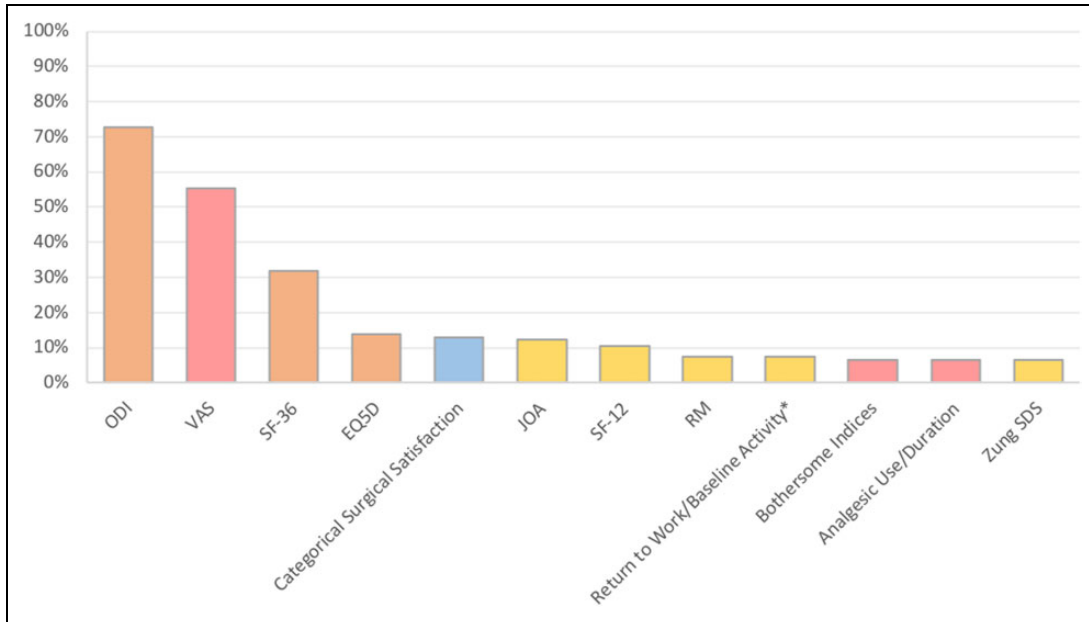


Figure 4. Top 10 most frequently reported PRO measures (of 124 studies). Red, yellow, and blue denote categorization among domains of pain/symptomatology, function/disability, and surgical satisfaction, respectively. Orange denotes measures assessing both pain and function criteria combined. *Reflects instances of reporting as a discrete, freestanding measure.

Table 3. PRO Utilization Stratified Over Time^a.

| 2007-2011 (n = 51 Studies) | 2012-2017 (n = 72 Studies) |
|---|---|
| 1. ODI (30) | 1. ODI (60) |
| 2. VAS (20) | 2. VAS (48) |
| 3. SF-36 (20) | 3. SF-36 (19) |
| 4. JOA (9) | 4. EQ5D (14) |
| 5. RM (5) | 5. Categorical surgical satisfaction (11) |
| 6. Categorical surgical satisfaction (5) | 6. ZCQ/SSSM (7) |
| 7. Bothersome indices (4) | 7. JOA (6) |
| 8. Return to work/baseline (4) ^a | 8. Return to work/baseline (5) ^a |
| 9. EQ5D (3) | 9. Zung SDS (5) |
| 10. Zung SDS (3) | 10. Analgesic use/duration (5) |

Abbreviations: ODI, Oswestry Disability Index; VAS, Visual Analog Scale; SF-36, 36-Item Short Form Survey; JOA, Japanese Orthopedic Association Score; RM, Roland Morris Disability Questionnaire Score; EQ5D, European Quality of Life-5 Dimensions; Zung SDS, Zung Self-rating Depression Scale; ZCQ/SSSM, Zurich Claudication Questionnaire/Swedish Spinal Stenosis Measure.

^aReflects instances of reporting as a discrete, freestanding measure.

The SF-36 originated from the efforts of the Medical Outcomes Study (MOS), a longitudinal study conducted in a general population of adult patients with medical conditions to determine the effect of various system-related factors that influence care.²⁴ The survey relies on 8 concepts of health: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, general mental health, social functioning, energy/fatigue or vitality, and general health perceptions. The questionnaire has been used across all types of pathology and captures a broad range of

domains, and its wide applicability suggests that the SF-36 is a validated PRO to measure a patient’s overall quality of life. It is reasonable to suggest that that surgical intervention for spinal stenosis may well affect some of the domains more so than others, or in more nebulous ways. In particular, general mental health, vitality, and role limitations due to personal or emotional problems may contribute to both surgical eligibility and post-operative outcome, not to mention the patient perception of such.

The third most commonly utilized PRO in our review, the Visual Analog Scale for pain, was developed in 1972 as a measurement of pain intensity.²⁵ Popular for its simplicity and ease of use, it is almost ubiquitous to measure a patient’s pain level, but recent studies cast wide doubt on whether this tool can capture the complexity and idiosyncratic nature of pain experience.^{26,27} For example, while 2 patients experience a moderate level of pain intensity, due to personal circumstances/daily requirements one patient may be extremely bothered by this level of pain while another patient may be minimally affected. Other measures of pain, such as the Sciatica Bothersome Index (SBI), which rates the “bothersomeness” of paresthesias, weakness, and leg or back pain, may be more specific to symptoms of LSS and more accurately measure the effect of this condition, and subsequent intervention.^{28,29} This outcome instrument has been used by well-known trials on interventions for lumbar spine surgery, including the Maine Lumbar Spine Study and the Spine Patient Outcomes Research Trial.^{3,4} However, our review found that out of all included studies, this PRO was utilized by only 4 articles published between 2007 and 2011, and even fewer published between 2012 and 2017.

Similarly to the rare utilization of the SBI, our review found few investigations that utilized PROs developed specifically to measure symptoms associated with LSS. The Zurich Claudication Questionnaire/Swedish Spinal Stenosis Measure (ZCQ/SSSM) was developed at 3 academic hospitals in 1996 as a patient questionnaire used to measure the effect of decompressive surgery on patients with LSS.³⁰ Since that time it has been validated in multiple languages.^{31,32} The questions selected for inclusion were chosen based on consensus opinion of an expert panel (rheumatologists, orthopedic surgeon, behavioral scientist) combined with a literature review. Each question corresponds to symptom severity, physical function, or satisfaction with the results of the back operation. The ZCQ/SSSM and the SRS-22 (developed for use in spine deformity surgery) were the only 2 PROs that encompassed the 3 PRO classifications we defined in our search: pain/symptomatology, function/disability, and surgical satisfaction. The most widely known application of the ZCQ/SSSM was in trials to measure the effectiveness of the X-STOP (Medtronic, Switzerland) dynamic stabilization device.^{33,34} Despite its more focused nature, in our review the ZCQ/SSSM was only used in 7 total studies.

In our increasingly cost-conscious environment, we must select treatments and interventions that deliver the greatest value for a given cost. On a disease-specific level, this requires standardized outcome instruments validated for a specific pathology to enable accurate comparison across studies. The current LSS literature, with heterogeneous reporting and sub-optimal PRO use (eg, ODI), limits our ability to draw such comparisons. We must reach consensus on outcomes reporting in LSS, whether it be an existing disease-specific measure, such as the ZCQ/SSSM, or an entirely new measure developed by an international panel of experts (anecdotally, we found that researchers in different regions of the world were especially prone to using varying methods, despite meeting in the same congress). Though it is difficult to truly demonstrate the superiority of one measure versus another without additional dedicated randomized trials, previously cited traits of a unifying measure would include disease-specific outcomes reporting, measures of general health, pain, satisfaction, and employment.^{7,21,35} We agree that reporting a general health measure should be standard practice in studies of LSS given that payers are increasingly forced to restrict payments to the most cost-effective interventions both within and across pathologies. The SF-36 is the most commonly reported general instrument seen in our study but has limitations, including significant floor/ceiling effects in the cervical population.³⁶ An alternative instrument is the National Institutes of Health PROMIS, which provides consistent, reliable, and responsive PROs that are generalizable across a variety of diseases and have been validated across a variety of orthopedic conditions.³⁷⁻⁴¹ PROMIS assesses physical, mental, and social health and uses item response theory, which is more responsive to changes than traditional testing, as well as computer adaptive testing, which decreases burden on the patient and allows for improved integration into electronic medical record systems. PROMIS has the potential to be an efficient, reliable, and responsive

instrument to measure general health and could be adopted in the LSS as well as other spinal literature.

Conclusions

Here we demonstrate that the LSS literature is characterized by substantial variability in outcomes reporting, with 31 total PROs and 90% of instruments reported in fewer than 15% of articles. Regarding impact on employment, only 14% of studies reported on return to work. Efforts to standardize outcomes reporting, including both a disease-specific and general health outcome measure, would facilitate comparison across the literature and improve our understanding of the prognosis of this disease. As the cost of care becomes increasingly scrutinized, such standardization will enable value comparisons across disciplines.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

James P. Winebrake  <https://orcid.org/0000-0001-9158-6438>

References

1. Bae HW, Rajae SS, Kanim LE. Nationwide trends in the surgical management of lumbar spinal stenosis. *Spine (Phila Pa 1976)*. 2013;38:916-926.
2. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus non-operative treatment for lumbar spinal stenosis four-year results of the spine patient outcomes research trial. *Spine (Phila Pa 1976)*. 2010;35:1329-1338.
3. Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE. Long-term outcomes of surgical and nonsurgical management of lumbar spinal stenosis: 8 to 10 year results from the Maine lumbar spine study. *Spine (Phila Pa 1976)*. 2005;30:936-943.
4. Weinstein JN, Tosteson TD, Lurie JD, et al; SPORT Investigators. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med*. 2008;358:794-810.
5. Aleem IS, Drew B. Cochrane in CORR®: surgical versus non-surgical treatment for lumbar spinal stenosis. *Clin Orthop Relat Res*. 2017;475:2632-2637.
6. Lurie JD, Tosteson TD, Tosteson A, et al. Long-term outcomes of lumbar spinal stenosis: eight-year results of the spine patient outcomes research trial (SPORT). *Spine (Phila Pa 1976)*. 2015;40:63-76.
7. Guzman JZ, Cutler HS, Connolly J, et al. Patient-reported outcome instruments in spine surgery. *Spine (Phila Pa 1976)*. 2016; 41:429-437.
8. Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus non-surgical treatment for lumbar spinal stenosis. *Cochrane Database Syst Rev*. 2016;(1):CD010264.

9. Porter ME. What is value in health care? *N Engl J Med*. 2010;363:2477-2481.
10. National Institutes of Health. PROMIS: Patient-Reported Outcomes Measurement Information System-Overview. <http://com.monfund.nih.gov/promis/index>. Accessed April 10, 2018.
11. Sielatycki JA, Parker SL, Godil SS, McGirt MJ, Devin CJ. Do patient demographics and patient-reported outcomes predict 12-month loss to follow-up after spine surgery? *Spine (Phila Pa 1976)*. 2015;40:1934-1940.
12. McGirt MJ, Parker SL, Hilibrand A, et al. Lumbar surgery in the elderly provides significant health benefit in the US health care system: patient-reported outcomes in 4370 patients from the N2QOD registry. *Neurosurgery*. 2015;77(suppl 4):S125-S135.
13. Armaghani SJ, Archer KR, Rolfe R, Demaio DN, Devin CJ. Diabetes is related to worse patient-reported outcomes at two years following spine surgery. *J Bone Joint Surg Am*. 2016;98:15-22.
14. Makhni EC, Padaki AS, Petridis PD, et al. High variability in outcome reporting patterns in high-impact ACL literature. *J Bone Joint Surg Am*. 2015;97:1529-1542.
15. Makhni EC, Steinhaus ME, Morrow ZS, et al. Outcomes assessment in rotator cuff pathology: what are we measuring? *J Shoulder Elbow Surg*. 2015;24:2008-2015.
16. Kahlenberg CA, Nwachukwu BU, Ferraro RA, Schairer WW, Steinhaus ME, Allen AA. How are we measuring patient satisfaction after anterior cruciate ligament reconstruction? *Orthop J Sport Med*. 2016;4:2325967116673971. doi:10.1177/2325967116673971
17. Oxford Center for Evidence-Based Medicine. 2011 Levels of evidence. <https://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf>. Accessed February 15, 2019.
18. Kepler CK, Vaccaro AR, Hilibrand AS, et al. National trends in the use of fusion techniques to treat degenerative spondylolisthesis. *Spine (Phila Pa 1976)*. 2014;39:1584-1589.
19. Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus nonsurgical treatment for lumbar spinal stenosis. *Spine (Phila Pa 1976)*. 2016;41:E857-E868.
20. Yadla S, Maltenfort MG, Ratliff JK, Harrop JS. Adult scoliosis surgery outcomes: a systematic review. *Neurosurg Focus*. 2010;28:E3.
21. Ueda H, Cutler HS, Guzman JZ, Cho SK. Current trends in the use of patient-reported outcome instruments in degenerative cervical spine surgery. *Global Spine J*. 2016;6:242-247.
22. Fairbank JC, Couper J, Davies JB, O'Brein JP. The Oswestry low back pain disability questionnaire. *Physiotherapy*. 1980;66:271-273.
23. Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)*. 2000;25:2940-2952.
24. Tarlov AR, Ware JE Jr, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The medical outcomes study: an application of methods for monitoring the results of medical care. *JAMA*. 1989;262:925-930.
25. Woodforde JM, Merskey H. Some relationships between subjective measures of pain. *J Psychosom Res*. 1972;16:173-178.
26. Hawker GA, Davis AM, French MR, et al. Development and preliminary psychometric testing of a new OA pain measure—an OARSI/OMERACT initiative. *Osteoarthritis Cartilage*. 2008;16:409-414.
27. Hush JM, Refshauge KM, Sullivan G, De Souza L, McAuley JH. Do numerical rating scales and the Roland-Morris Disability Questionnaire capture changes that are meaningful to patients with persistent back pain? *Clin Rehabil*. 2010;24:648-657.
28. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. *Spine (Phila Pa 1976)*. 1995;20:1899-1909.
29. Grøvle L, Haugen AJ, Keller A, Natvig B, Brox JI, Grotle M. The bothersomeness of sciatica: patients' self-report of paresthesia, weakness and leg pain. *Eur Spine J*. 2010;19:263-269.
30. Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossel AH, Katz JN. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. *Spine (Phila Pa 1976)*. 1996;21:796-803.
31. Kim HJ, Lee YK, Kim DO, Chang BS, Lee CK, Yeom JS. Validation and cross-cultural adaptation of the Korean version of the Zurich Claudication Questionnaire in patients with lumbar spinal stenosis. *Spine (Phila Pa 1976)*. 2018;43:E105-E110.
32. Hara N, Matsudaira K, Masuda K, et al. Psychometric assessment of the Japanese version of the Zurich Claudication Questionnaire (ZCQ): reliability and validity. *PLoS One*. 2016;11:0160183. doi:10.1371/journal.pone.0160183
33. Huddleston P. X-stop resulted in a higher reoperation rate than minimally invasive decompression, but both decreased symptoms of neurogenic intermittent claudication in patients with lumbar spinal stenosis. *J Bone Joint Surg Am*. 2015;97:1889.
34. Anderson PA, Tribus CB, Kitchel SH. Treatment of neurogenic claudication by interspinous decompression: application of the X STOP device in patients with lumbar degenerative spondylolisthesis. *J Neurosurg Spine*. 2006;4:463-471.
35. Müller U, Duetz MS, Roeder C, Greenough CG. Condition-specific outcome measures for low back pain. Part I: validation. *Eur Spine J*. 2004;13:301-313.
36. Baron R, Elashaal A, Germon T, Hobart J. Measuring outcomes in cervical spine surgery: think twice before using the SF-36. *Spine (Phila Pa 1976)*. 2006;31:2575-2584.
37. Cella D, Riley W, Stone A, et al; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010;63:1179-1194.
38. Hung M, Hon SD, Franklin JD, et al. Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. *Spine (Phila Pa 1976)*. 2014;39:158-163.
39. Hung M, Baumhauer JF, Latt LD, et al; National Orthopaedic Foot & Ankle Outcomes Research Network. Validation of PROMIS physical function computerized adaptive tests for orthopaedic foot and ankle outcome research. *Clin Orthop Relat Res*. 2013;471:3466-3474.
40. Tyser AR, Beckmann J, Franklin JD, et al. Evaluation of the PROMIS physical function computer adaptive test in the upper extremity. *J Hand Surg Am*. 2014;39:2047-2051.
41. Hung M, Clegg DO, Greene T, Saltzman CL. Evaluation of the PROMIS physical function item bank in orthopaedic patients. *J Orthop Res*. 2011;29:947-253.