SPINE An International Journal for the study of the spine, Publish Ahead of Print

DOI: 10.1097/BRS.0000000000003501

Evaluating the Minimal Clinically Important Difference of EQ-5D-3L in Patients with Degenerative Lumbar Spinal Stenosis

---

A Swiss prospective multi-center cohort study

Jakob M. Burgstaller, M.D., D.M.D., Ph.D.[1], Maria M. Wertli, M.D., Ph.D.[1,2], Nils H. Ulrich, M.D.[1], Giuseppe Pichierri, Ph.D.[1], Florian Brunner, M.D., Ph.D.[3], Mazda Farshad, M.D., M.P.H.[4], François Porchet, M.D.[5], Johann Steurer, M.D.[1], Isaac Gravestock, Ph.D.[1], on behalf of the LSOS Study Group

[1] Horten Centre for Patient Oriented Research and Knowledge Transfer, University of Zurich, Switzerland

[2] Division of General Internal Medicine, Bern University Hospital, Bern University, Switzerland

[3] Department of Physical Medicine and Rheumatology, Balgrist University Hospital, Forchstrasse 340, 8008 Zurich, Switzerland

[4] Spine Division, Balgrist University Hospital, Forchstrasse 340, 8008 Zurich, Switzerland

[5] Spine Unit, Department of Orthopedic- and Neurosurgery, Schulthess Clinic, Lengghalde 2, 8008 Zurich, Switzerland

Corresponding author

Jakob M. Burgstaller, Horten Centre Zurich, University of Zurich and University Hospital Zurich Pestalozzistrasse 24, CH-8032 Zürich, Email: jakob.burgstaller@usz.ch; Tel.: +41 44 255 87 09

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work.

Relevant financial activities outside the submitted work: grants.

Abstract

Study Design. Analysis of a prospective, multicenter cohort study.

Objective. The aim of our study was to compare thresholds of published minimal clinically important differences (MCID) for the 3-level EuroQol-5D health survey (EQ-5D-3L) summary index (range -0.53 to 1.00) with our anchor-based estimate and evaluate how useful these thresholds are in determining treatment success in patients undergoing surgery for degenerative lumbar spinal stenosis (DLSS).

Summary of Background Data. MCID values for EQ-5D-3L are specific to the underlying disease and only three studies have been published for DLSS patients reporting different values.

Methods. Patients of the multi-center Lumbar Stenosis Outcome Study (LSOS) with confirmed DLSS undergoing first-time decompression or fusion surgery with 12-month follow-up were enrolled in this study. To calculate MCID we used the Spinal Stenosis Measure (SSM) satisfaction subscale as anchor.

Results. For this study, 364 patients met the inclusion criteria; of these, 196 were very satisfied, 72 moderately satisfied, 43 somewhat satisfied and 53 unsatisfied 12-month after surgery. The MCID calculation estimated for EQ-5D-3L a value of 0.19. Compared to published MCID values (ranging from 0.30 to 0.52), our estimation is less restrictive.

Conclusions. In patients with LSS undergoing surgery, we estimated an MCID value for EQ-5D-3L summary index of 0.19 with help of the average change anchor-based method, which we find to be the most suitable method for assessing patient change scores.

Key Points

- Minimal clinically important differences (MCID) values for EQ-5D-3L are specific to the underlying disease and a useful concept to assess treatment success.
- Only three studies have been published yet calculating MCID of EQ-5D-3L summary index for patients with degenerative lumbar spinal stenosis undergoing surgery reporting different values.
- We estimated an MCID value of 0.19 based on the average change anchor method, which we find to be the most suitable method for assessing patient change scores.
- Compared to previously published MCID values (ranging from 0.30 to 0.52), our estimation is less restrictive.

Key Words: clinical meaningful improvement; decompression; degenerative lumbar spinal stenosis; EQ-5D; Fusion; meaningful clinical improvement; MCID; multi-center; quality of life; spine-surgery

Level of Evidence: 3

Introduction

Degenerative lumbar spinal stenosis (DLSS) is a highly prevalent condition and causes considerable pain and disability.[1] It has a remarkable impact on the patient's daily life and health. The treatment options range from conservative (pain management, physiotherapy, epidural steroid injections) to surgical approaches. Further, DLSS is the most frequent indication for spine surgery in patients older than 65 years of age.[2] Nevertheless, the amount of benefit of the surgery varies and one third of the patients did not report an improvement in pain, disability, or quality of life after surgery.[3-6] However, less is known about the surgery's impact on quality of life.

The EuroQol five-dimensional questionnaire (EQ-5D) is a standardized instrument to measure health-related quality of life (HRQOL).[7] The EQ-5D is a short self-administered patient-reported outcome measure (PROM) broadly applied in clinical studies to capture the health status from the patients' perspective.

To facilitate the interpretation of the evaluation of PROMs and hence the clinical relevance of a treatment effect, the "minimal clinically important difference" (MCID) is a useful concept.[8] It was proposed by Jaeschke et al.[9] in 1989 and it is defined as "the smallest difference in a score that is considered to be worthwhile or important".[10] Thus, the MCID determines a threshold for a clinically relevant change in a PROM − in contrast to the "mere" statistical significance threshold that does not automatically mean a clinically relevant benefit from therapy. If an average patient reaches or even exceeds the aforementioned MCID threshold, they consider their change as worthwhile and meaningful.[11]

MCID values are specific to the underlying disease[12,13] and have been already estimated for the 3-level version of EQ-5D (EQ-5D-3L) for various spinal disorders.[14-16] For patients with DLSS, we identified three studies that calculated an MCID, however, the values proposed varied considerably between 0.30 and 0.52 using the anchor-based mean change approach.[17-19] Considering an EQ-5D-3L score range of -0.53 to 1.00, an MCID of 0.52 would require an enormous change, which could hardly be classified as minimal. Further, all three studies were performed in the USA and consequently used another value set for the analysis due to different population norms. Hence, studies conducted in Europe may find the proposed MCIDs unsuitable.

The aim of our study is to compare these thresholds with our own anchor-based estimate and evaluate how useful these thresholds are in determining treatment success in patients undergoing surgery included in the Lumbar Stenosis Outcome Study (LSOS) performed in Switzerland.

Methods

Study design

For this retrospective analysis, we used data from the Lumbar Stenosis Outcome Study (LSOS). The LSOS is conducted as a prospective cohort study in the Rheumatology and Spine Surgery Units at eight medical centers (which service a region in Switzerland with approximately two million inhabitants). They were eligible for the LSOS study if they 1) were aged 50 years or more, 2) had a history of neurogenic claudication, 3) had lumbar spinal stenosis verified by Magnetic Resonance Imaging (MRI) or Computer Tomography (CT), 4) had no evidence of stenosis caused by tumor, infection, fracture, or significant deformity (>15° lumbar scoliosis, diagnosed on conventional x-ray with anterior-posterior and lateral views), and 5) had no clinical peripheral artery occlusive disease. The decision of the treatment strategy (conservative or surgical) was made by the patient and his attending physician. Additional information about LSOS is available elsewhere.[20]

Eligibility criteria for being included in this analysis

All patients were eligible who: met the above mentioned inclusion criteria, underwent first-time decompression surgery alone or additional fusion surgery within the first six months of enrollment, and had a 12-month follow-up.

Patients who had previous lumbar spine surgery, whose lumbar spine surgery was later than 6 months after study enrolment, who underwent revision surgery, or who received only conservative care and/or epidural injection were excluded.

Outcome Measures

EQ-5D: The EuroQol five-dimensional self-administered questionnaire (EQ-5D) is a standardized instrument to measure health-related quality of life and was developed by the EuroQol Group.[7] The EQ-5D descriptive system measures the health state in five dimensions

(mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), either with three levels of severity for each dimension (EQ-5D-3L), the most widely used version,[21] or with five levels (EQ-5D-5L).[7] The five dimensions can be reported as a single 5-digit number (from 11111 representing full health to 33333 [for EQ-5D-3L] or 55555 [for EQ-5D-5L], respectively, meaning worst health). With help of a value set (depending on population norms), the health state can be converted into a single summary index value. This value can range from -0.53 (for the French population) to 1, with 0 representing a health state equivalent to being dead and 1 indicating full health. The French value set was used to the calculation of summary index, as there is no Switzerland specific value set.[22] Further, the LSOS study only applied the EQ-5D-3L questionnaire.

The EQ-5D-3L was found to be an effective measure of health status in patients with degenerative lumbar spine pathologies (conservatively or surgically treated).[23] Further, its reliability was evaluated in patients undergoing low-back surgery and was determined to be "solid" (intraclass correlation coefficient (ICC) varied between 0.82 (95% confidence interval (CI) 0.67 to 0.90, $p<0.001$) and 0.87 (95% CI 0.72 to 0.94, $p<0.001$)).[24]

Spinal Stenosis Measure (SSM): The SSM, also known as Zurich Claudication Questionnaire (ZCQ), is a self-administered validated three-part questionnaire that was specifically designed for DLSS patients to measure their severity of symptoms and disability.[25] It is widely used in studies on DLSS[26-29] and recommended by the North American Spine Society (NASS). It consists of three different subscales; *the symptom severity subscale*, *the physical function subscale* and the *satisfaction subscale*. The subscale score ranges are 1−5, 1−4 and 1−4 (best-worst). The corresponding ICC values were 0.87 (95% CI 0.81 to 0.92), 0.81 (95% CI 0.71 to 0.87), and 0.90 (95% CI 0.84 to 0.93), respectively.[30] MCID in SSM symptoms is defined as an improvement (decrease) by at least 0.48 points, in SSM function by at least 0.52 points.[31]

Anchor Measures

SSM satisfaction subscale was used as anchor. This is a set of six questions asked following treatment to evaluate a patient's satisfaction with their treatment. The six questions have possible responses 1 (very satisfied), 2 (somewhat satisfied), 3 (somewhat dissatisfied), and 4 (very dissatisfied) which are averaged to give an overall satisfaction score. A score more than 2.5 represents a patient who is on average unsatisfied with the treatment. Stucki et al.[25] further

divided the range of satisfied scores into "very", "moderately", and "somewhat". Using this classification system, we use the patients who were somewhat satisfied to represent a minimal improvement like in previous studies.[25,31] The cut-offs for the somewhat group are less than or equal to 2.5 and greater than 2.0.

The other subscales of the SSM questionnaire (symptoms and function) were used as reference to validate the selected MCID.

Selection of MCID

We used the "mean change" method to calculate the MCID.[9] This identifies the average change from baseline to follow-up in the outcome of interest (EQ-5D-3L) in those patients with the smallest positive response in the anchor (SSM Satisfaction).

To show the influence of the different population normed value sets on the calculation of the summary index for the EQ-5D-3L and hence of the MCID, we additionally applied the value sets of the UK, US, and German population in our patient sample.

Statistical analysis

We summarized the patient characteristics at baseline with means and standard deviations (SD) and counts and percentages as appropriate. To test for differences between the groups of patients with different satisfaction levels we used Chi-squared tests for categorical variables and Welch's two-sided t-tests for continuous variables and reported p-values. As this was descriptive only, we did not set a threshold on the p-values.

We showed the relationship between SSM Satisfaction and change in EQ-5D-3L with a scatter plot and linear slopes and calculated the Pearson correlation coefficient.

To compare how the threshold derived in LSOS differs from the previously published thresholds, we compared the number and percentage of patients reaching MCID using each threshold and depicted this graphically.

All analyses were conducted with R for Windows.[32]

Results

Patient characteristics at baseline

Between December 2010 and December 2015, 841 patients agreed to participate in the LSOS study (**Figure 1**). For this analysis, 364 patients met the inclusion criteria; of these, 196 were very satisfied, 72 moderately satisfied, 43 somewhat satisfied and 53 unsatisfied after 12-month follow-up.

Table 1 presents the baseline characteristics of the somewhat satisfied and unsatisfied patients (see also **Appendix Table 1** http://links.lww.com/BRS/B563 for the baseline characteristics of all 364 patients). Mean age of patients in those both groups were 74 years, 60% were female, and 76% had symptoms longer than six months. In both groups, around 65% had a spondylolisthesis on the level of surgery. Mean baseline EQ-5D-3L summary index values were 0.40 (SD 0.30) for all patients, 0.45 (SD 0.28) in the somewhat satisfied group, and 0.35 (SD 0.31) in the unsatisfied group. None of the p-values are large and therefore do not suggest a difference between the groups.

Estimation of MCID

The mean change (improvement) in EQ-5D-3L summary index after 12-month follow-up was 0.19 for the somewhat satisfied, which corresponds to our MCID value, and 0.09 for the unsatisfied patients. **Figure 2** depicts a positive, albeit weak, relationship between mean SSM satisfaction and mean change in EQ-5D-3L summary index between 12-month follow-up and baseline in the unsatisfied group: as patient satisfaction improved (i.e. became less unsatisfied) changes in EQ-5D-3L became more positive. In the somewhat satisfied group, there was no evidence for a positive or negative relationship.

We fit the linear slopes and correlations separately for the unsatisfied and somewhat satisfied patients as we did not wish to assume a linear relationship between EQ-5D and SSM satisfaction over the qualitative shift from unsatisfied to satisfied.

Comparison with MCID of SSM symptoms and function

Figure 3 compares the change of SSM symptoms or function with the change of EQ-5D-3L between 12-month follow-up and baseline for all 364 patients. It shows that most patients who reached MCID in SSM symptoms or function (solid vertical line) are experiencing an improved

quality of life (66.6% and 74.6%, respectively) when applying our MCID value (solid horizontal line). The MCID values of Asher 2018[19]/Parker 2011[17] (dashed horizontal line) and Parker 2012[18] (dotted horizontal line) were more conservative (52.3% and 28.4% for SSM symptoms and 59.1% and 34.3% for SSM function, respectively) compared to our MCID value. The absolute percentage values for each quadrant can be found in **Appendix Table 2** http://links.lww.com/BRS/B563.

Ceiling effect

In our patient sample, there are many patients who reached the upper limit of the EQ-5D-3L instrument (1.00, best life quality) at 12-month follow-up, in particular in the highly satisfied group (53.1%). In the somewhat satisfied group only one patient (2.3%) has a summary index value of 1.00 at 12-month follow-up and no patient in the unsatisfied group. Therefore, the MCID estimate was not biased by a ceiling effect.

Comparison with other population value sets

When applying the value set for different populations, we estimated the following MCIDs for our patient sample (**Appendix Table 3** http://links.lww.com/BRS/B563): for the UK population 0.22, for the US population 0.14, and for the German population 0.20. The mean change (improvement) in EQ-5D-3L between baseline and follow-up was 0.19, 0.13, and 0.19, respectively.

Discussion

In our study investigating patients with lumbar spinal stenosis undergoing surgery, we have defined an MCID threshold of 0.19 which shows good agreement with other patient reported outcomes (SSM symptoms and function).

Unlike other studies,[17-19] we have only used an anchor-based method, as only anchor based methods can properly take into account the clinical aspect in to their estimates.[33] Specifically, we use the "average change" method based on the "somewhat satisfied" group.[9] Considering the definition of MCID ("the smallest difference in a score that is considered to be worthwhile or important"),[10] we need to differentiate between the before-after difference due to a therapy, and a difference between the health status of two groups patients. The "change difference" MCID

method is often attributed to Redelmeier et al.[34] but that publication describes a method for comparing patients at a single time point. Today, the "change difference" MCID refers to the difference between the change scores of the "slightly improved" patients and the changes scores of the "unchanged" patients, i.e. the difference of differences. This "change difference" MCID is therefore not suitable for classifying an individual patient's change score as meaningful or not. To see this, consider a study where the "slightly improved" group has an average change score of 0.19 and the "unchanged" group has an average change of 0.10. The "change difference" MCID estimate would be 0.19-0.10=0.09. Therefore, a patient with a score of 0.095 who rated themselves as having "no change" would be classified as having a meaningful change because they reached MCID (0.095>0.09)! The "change difference" MCID is more suitable for classifying the size of a treatment effect. The ROC method is also anchor-based but can be very sensitive to the patient sample used and estimates an optimal classification threshold, which does not necessarily correspond to the concept of MCID.[35] Estimates from distribution-based methods, such as estimate minimal detectable changes (MCD), half-SD, Cohen's effect size, etc., are statistical in nature and while may be useful in further calculations, are not calculated with any reference to clinical importance and so cannot be used by themselves as an MCID.

Further, we see in our study that the anchor and the measure are not fully in agreement. There are patients who report satisfaction but have a negative change in the measure. This may be due to nuances in the different questionnaires exposing different aspects of patients' quality of life. Alternatively, it may be due to a response shift phenomena where patients have a different conception of their condition after an intervention.[36]

Some aspects could have led to the differences in MCID values of the studies by Parker et al.[17,18] and Asher et al.[19] compared to our estimation; first, they included all "satisfied" patients whereas we only included "somewhat" satisfied patients. The minimal difference can only be estimated with patients in the somewhat group, otherwise the MCID is grossly overestimated,[37] and is rather the Average Clinically Important Difference. Further, different anchors (Parker 2011 and 2012: health transition item (HTI) of the SF-36; Asher 2018: NASS satisfaction scale) were used what might have also influenced the MCID values. Second, mean age was much lower in their studies than in our patient sample (50.9±11.8, 56.3±12.5 [Parker 2011 and 2012] and 62.0±11.2 [Asher 2018] vs 74.0±8.1 years in our study). It is well known that age influences recovery also in long-term follow-up.[38-40] All in all, these reasons could have led to greater mean

change in EQ-5D between baseline and follow-up (0.43, 0.41 [Parker 2011 and 2012] and 0.24 [Asher 2018] vs 0.13 in our study) and therefore to higher MCID values. The surgical procedure, decompression or fusion surgery, should not have affected the results, as Asher et al.[19] showed in their subgroup analysis separately for laminectomy and fusion patients. They concluded that MCID values are rather more specific to the diagnosis than to the procedure type. Therefore, we also included decompression as well as fusion procedures in our analysis.

The studies by Parker et al.[17,18] and Asher et al.[19] were all performed in the USA and therefore used another population value set. We were able to show that when using the US population value set it led to differences up to almost 50% in the estimation of the MCID, compared to using European population value sets. Consequently, it is important that the most appropriate population value set is used for calculations and for the evaluation of patient outcomes.

The EQ-5D-3L has some disadvantages regarding its instrument's sensitivity and the risk of encountering ceiling effects. This risk is defined as "the proportion of respondents scoring no problems on any of the five dimensions",[41] which equals 11111. In our sample, 114 (31%) patients reported the best possible score at 12 months. While this did not impact our estimation of MCID, it does highlight limitation of this instrument and the danger of blind application of a threshold on change scores. In our patients, 59% have a baseline score >0.48 (=1 - 0.52), which would prevent them from reaching Parker et al's[18] MCID threshold (and for Parker 2011[17]/Asher 2018[19] 34% are >0.70 (= 1 - 0.30) and for LSOS 13% >0.81 (= 1 - 0.19)). To counteract this challenge, the EuroQol Group introduced a revised version, the EQ-5D-5L, with five levels of severity for each dimension, allowing patients to be more precise in their health assessment. Its validity was first time tested by Herdman et al.[42] in 2011. Unfortunately, the EQ-5D-5L was not available at the start of the LSOS study in the year 2010.

A limitation of our study was that there is a lack of a "gold" standard measure of HRQOL to which the EQ-5D-3L can be compared, although a publication showed that the EQ-5D-3L was "the most valid and responsive measure of improvement" for HRQOL (compared to SF-12 and Zung Depression Scale) in patients undergoing lumbar fusion surgery.[43] Further, our patient sample included patients with and without degenerative spondylolisthesis − representing a typical DLSS population − which might have an influence on the outcome. However, a recent study showed no difference in mean EQ-5D after surgery regardless of the presence of a preoperative spondylolisthesis.[44]

It is crucial that studies using an MCID value from other publications for their own analyses consider the patient population the estimate is based on. For instance, Jansson et al.[45] who investigated patients with LSS who underwent surgery used the calculated MCID from the publication of Walters et al.[46] who compared in eight longitudinal studies with eleven different patient groups the MCID (mean value over all studies: 0.074). However, the patient groups were quite heterogeneous (the underlying diseases ranged from back pain through irritable bowel syndrome to acute myocardial infarction), and therefore, the patients' chance of improvement varies considerably. That might lead to under- or overestimated results. And as discussed above future studies should only use the new EQ-5D-5L to reduce potential ceiling effects and hence to prevent bias in the outcome.

When applying these MCID thresholds in future studies, researchers should take care that these thresholds are always an average estimate and that individual patients may have greater or lower thresholds for perceiving an important change. Given the risk of ceiling effects in EQ-5D-3L, it might be more appropriate to use a relative measure of improvement, such as the percentage of possible improvement (PPI) following Gilmer et al.[47] and Somerson et al.[48] in addition to the change score.

Conclusion

In patients with LSS undergoing surgery, we estimated an MCID value for EQ-5D-3L summary index of 0.19 with help of the average change anchor-based method, which we find to be the most suitable method for assessing patient change scores.

References

1. Lurie J, Tomkins-Lane C. Management of lumbar spinal stenosis. *BMJ* 2016;352:h6234.
2. Deyo RA. Treatment of lumbar spinal stenosis: a balancing act. *Spine J* 2010;10:625-7.
3. Atlas SJ, Keller RB, Wu YA, et al. Long-term outcomes of surgical and nonsurgical management of lumbar spinal stenosis: 8 to 10 year results from the maine lumbar spine study. *Spine* 2005;30:936-43.
4. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. *Spine* 2010;35:1329-38.
5. Held U, Burgstaller JM, Wertli MM, et al. Prognostic function to estimate the probability of meaningful clinical improvement after surgery - Results of a prospective multicenter observational cohort study on patients with lumbar spinal stenosis. *PLoS One* 2018;13:e0207126.
6. Rampersaud YR, Wai EK, Fisher CG, et al. Postoperative improvement in health-related quality of life: a national comparison of surgical treatment for focal (one- to two-level) lumbar spinal stenosis compared with total joint arthroplasty for osteoarthritis. *The Spine Journal* 2011;11:1033-41.
7. Group TE. EuroQol-a new facility for the measurement of health-related quality of life. *Health policy* 1990;16:199-208.
8. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol* 2002;14:109-14.
9. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15.
10. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics* 2000;18:419-23.
11. Burgstaller JM, Wertli MM, Steurer J, et al. The Influence of Pre- and Postoperative Fear Avoidance Beliefs on Postoperative Pain and Disability in Patients With Lumbar Spinal Stenosis: Analysis of the Lumbar Spinal Outcome Study (LSOS) Data. *Spine* 2017;42:E425-E32.
12. Zhou F, Zhang Y, Sun Y, et al. Assessment of the minimum clinically important difference in neurological function and quality of life after surgery in cervical spondylotic myelopathy patients: a prospective cohort study. *Eur Spine J* 2015;24:2918-23.
13. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11:171-84.
14. Parker SL, Adogwa O, Mendenhall SK, et al. Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis. *Spine J* 2012;12:1122-8.

15. Parker SL, Mendenhall SK, Shau D, et al. Determination of minimum clinically important difference in pain, disability, and quality of life after extension of fusion for adjacent-segment disease. *J Neurosurg Spine* 2012;16:61-7.

16. Johnsen LG, Hellum C, Nygaard OP, et al. Comparison of the SF6D, the EQ5D, and the oswestry disability index in patients with chronic low back pain and degenerative disc disease. *BMC Musculoskelet Disord* 2013;14:148.

17. 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;14:598-604.

18. 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;16:471-8.

19. Asher AL, Kerezoudis P, Mummaneni PV, et al. Defining the minimum clinically important difference for grade I degenerative lumbar spondylolisthesis: insights from the Quality Outcomes Database. *Neurosurg Focus* 2018;44:E2.

20. Steurer J, Nydegger A, Held U, et al. LumbSten: The lumbar spinal stenosis outcome study. *Bmc Musculoskeletal Disorders* 2010;11.

21. Nolan CM, Longworth L, Lord J, et al. The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. *Thorax* 2016;71:493-500.

22. Matter-Walstra K, Klingbiel D, Szucs T, et al. Using the EuroQol EQ-5D in Swiss cancer patients, which value set should be applied? *Pharmacoeconomics* 2014;32:591-9.

23. Mueller B, Carreon LY, Glassman SD. Comparison of the EuroQOL-5D with the Oswestry Disability Index, back and leg pain scores in patients with degenerative lumbar spine pathology. *Spine* 2013;38:757-61.

24. Solberg TK, Olsen J-A, Ingebrigtsen T, et al. Health-related quality of life assessment by the EuroQol-5D can provide cost-utility data in the field of low-back surgery. *Eur Spine J* 2005;14:1000-7.

25. Stucki G, Liang MH, Fossel AH, et al. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *J Clin Epidemiol* 1995;48:1369-78.

26. Fokter SK, Yerby SA. Patient-based outcomes for the operative treatment of degenerative lumbar spinal stenosis. *Eur Spine J* 2006;15:1661-9.

27. Hansraj KK, O'Leary PF, Cammisa FP, Jr., et al. Decompression, fusion, and instrumentation surgery for complex lumbar spinal stenosis. *Clin Orthop Relat Res* 2001:18-25.

28. Tuli SK, Yerby SA, Katz JN. Methodological approaches to developing criteria for improvement in lumbar spinal stenosis surgery. *Spine* 2006;31:1276-80.

29. Zucherman JF, Hsu KY, Hartjen CA, et al. A multicenter, prospective, randomized trial evaluating the X STOP interspinous process decompression system for the treatment of neurogenic intermittent claudication: two-year follow-up results. *Spine* 2005;30:1351-8.

30. Wertli MM, Steurer J, Wildi LM, et al. Cross-cultural adaptation of the German version of the spinal stenosis measure. *Eur Spine J* 2014;23:1309-19.

31. Stucki G, Daltroy L, Liang MH, et al. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. *Spine* 1996;21:796-803.

32. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019.

33. Copay AG, Subach BR, Glassman SD, et al. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;7:541-6.

34. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements: an illustration in rheumatology. *Archives of internal medicine* 1993;153:1337-42.

35. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol* 2017;82:128-36.

36. Schwartz CE, Finkelstein JA. Understanding inconsistencies in patient-reported outcomes after spine treatment: response shift phenomena. *The Spine Journal* 2009;9:1039-45.

37. Copay AG. Commentary: the proliferation of minimum clinically important differences. *Spine J* 2012;12:1129-31.

38. Makino T, Kaito T, Fujiwara H, et al. Risk Factors for Poor Patient-Reported Quality of Life Outcomes After Posterior Lumbar Interbody Fusion: An Analysis of 2-Year Follow-up. *Spine* 2017;42:1502-10.

39. Takenaka H, Kamiya M, Sugiura H, et al. Predictors of walking ability after surgery for lumbar spinal canal stenosis: A prospective study. *The Spine Journal* 2019.

40. Gautschi OP, Smoll NR, Joswig H, et al. Influence of age on pain intensity, functional impairment and health-related quality of life before and after surgery for lumbar degenerative disc disease. *Clin Neurol Neurosurg* 2016;150:33-9.

41. Janssen MF, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of Life Research* 2013;22:1717-27.

42. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727-36.

43. Godil SS, Parker SL, Zuckerman SL, et al. Accurately measuring the quality and effectiveness of lumbar surgery in registry efforts: determining the most valid and responsive instruments. *Spine J* 2014;14:2885-91.

44. Forsth P, Michaelsson K, Sanden B. Does fusion improve the outcome after decompressive surgery for lumbar spinal stenosis? A TWO-YEAR FOLLOW-UP STUDY INVOLVING 5390 PATIENTS. *Bone Joint J* 2013;95b:960-5.

45. Jansson KA, Nemeth G, Granath F, et al. Health-related quality of life (EQ-5D) before and one year after surgery for lumbar spinal stenosis. *J Bone Joint Surg Br* 2009;91:210-6.

46. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;14:1523-32.

47. Gilmer BB, Comstock BA, Jette JL, et al. The prognosis for improvement in comfort and function after the ream-and-run arthroplasty for glenohumeral arthritis: an analysis of 176 consecutive cases. *J Bone Joint Surg Am* 2012;94:e102.

48. Somerson JS, Sander P, Bohsali K, et al. What Factors are Associated With Clinically Important Improvement After Shoulder Hemiarthroplasty for Cuff Tear Arthropathy? *Clin Orthop Relat Res* 2016;474:2682-8.

Figure legend

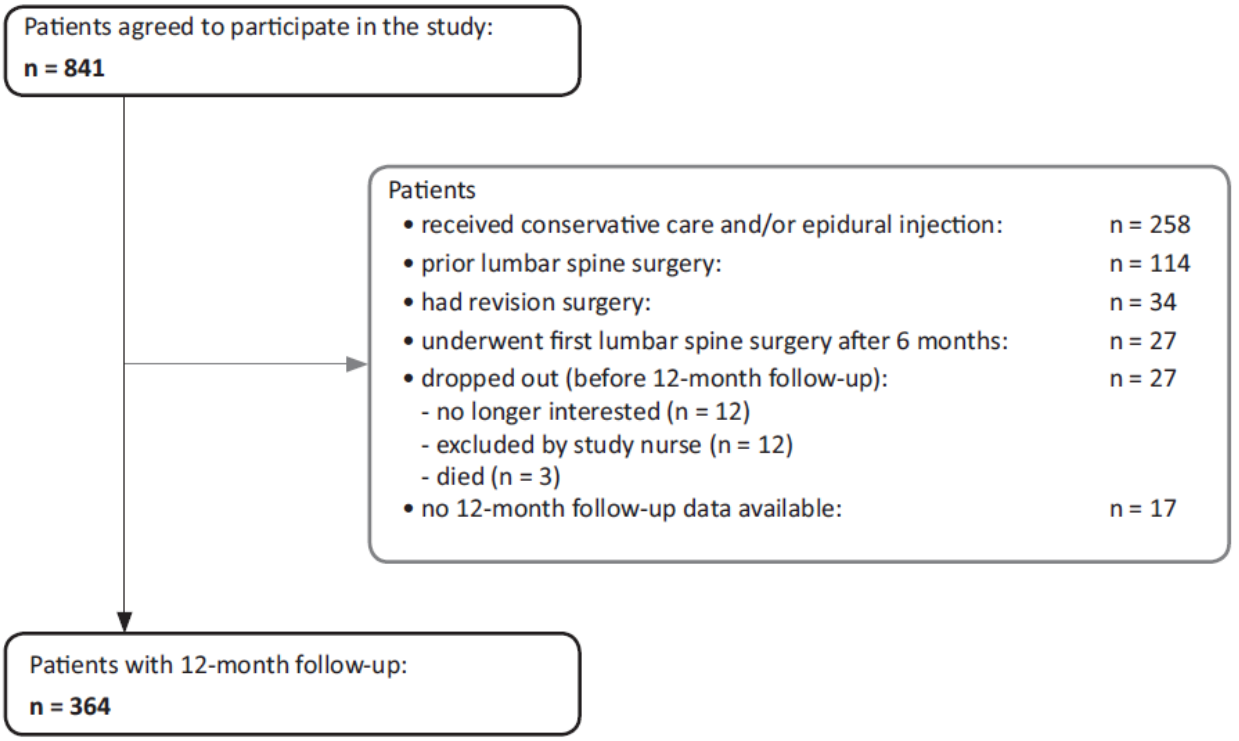

Figure 1: Study flow

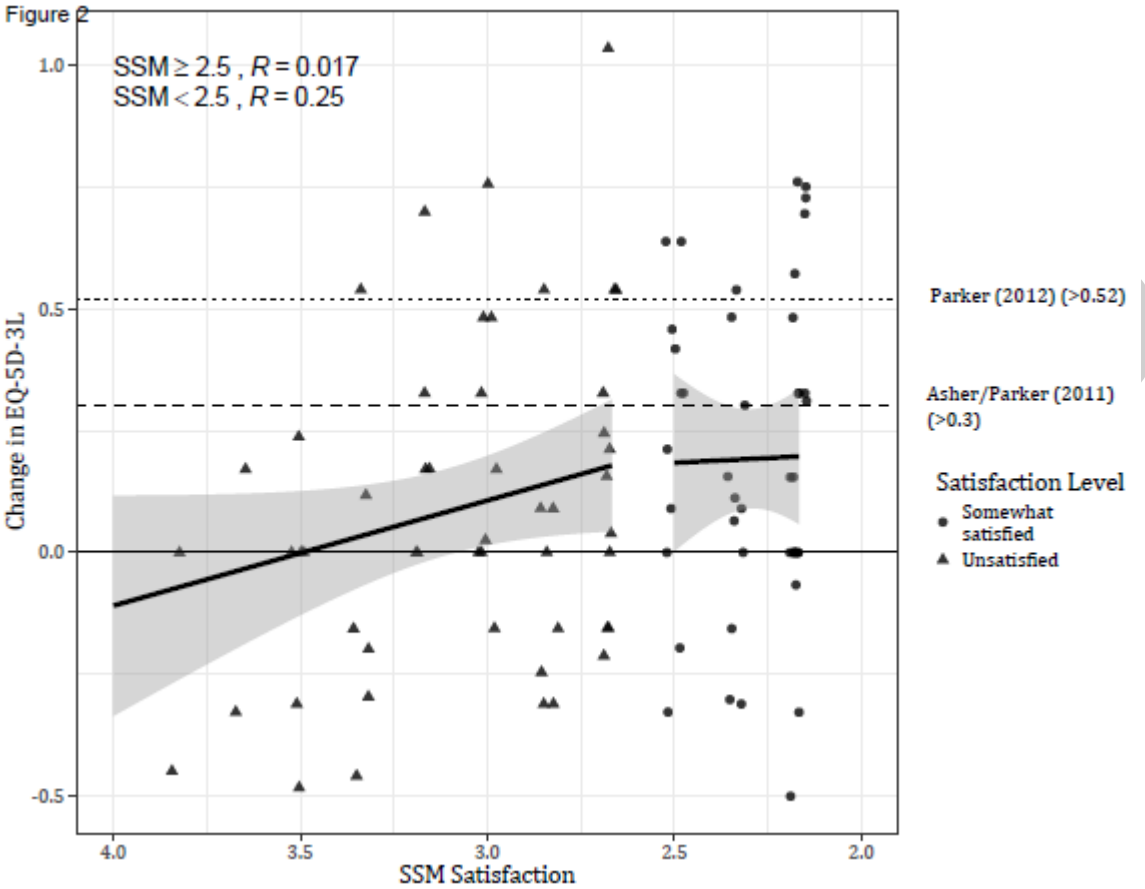

Figure 2: Patient reported SSM satisfaction versus EQ-5D-3L change

There is a positive, albeit weak, relationship (absolute correlation values 0.017 and 0.25) between mean SSM satisfaction and mean change in EQ-5D-3L summary index between 12-month follow-up and baseline. "Somewhat" satisfied patients have a mean SSM satisfaction value ≥2 and <2.5, "unsatisfied" patients ≥2.5. In the unsatisfied group: as patient satisfaction improved (i.e. became less unsatisfied) changes in EQ-5D-3L became more positive. In the somewhat satisfied group, there was no evidence for a positive or negative relationship..

Dashed horizontal line: MCID estimation of Parker 2011/Asher 2018 of 0.30; Dotted horizontal line: estimation of Parker 2012 of 0.52; R = Pearson correlation coefficient; SSM = Spinal Stenosis Measure

Figure 3

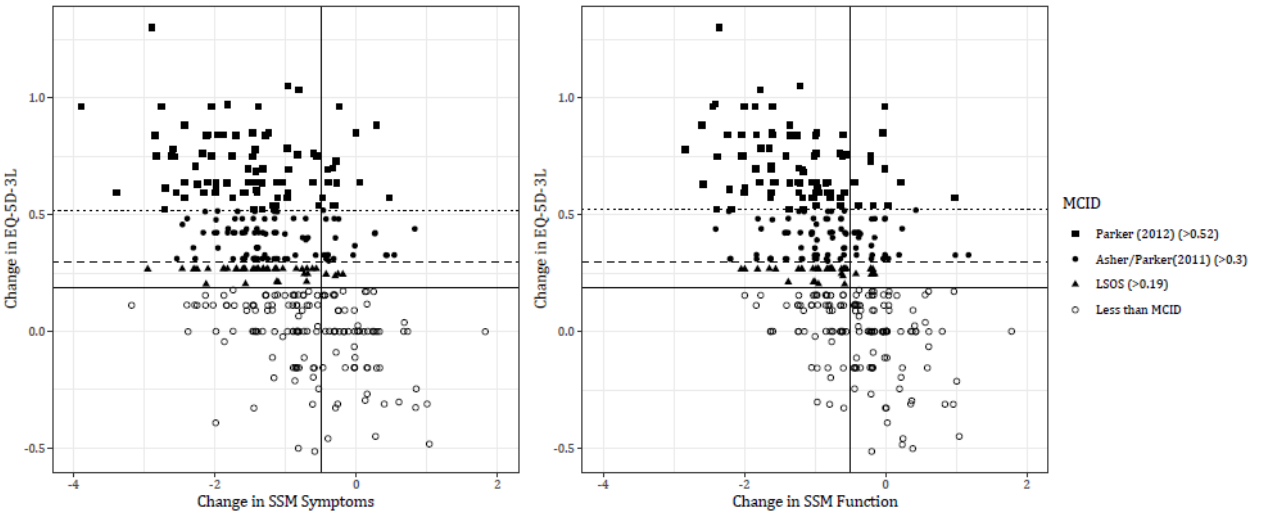

Figure 3: Comparison of various MCIDs of EQ-5D-3L to MCID of SSM symptoms and function applied to our full patient sample (n=364).

Solid horizontal line: our MCID estimation of 0.19; Dashed horizontal line: estimation of Parker 2011/Asher 2018 of 0.30; Dotted horizontal line: estimation of Parker 2012 of 0.52; Solid vertical line: MCID estimation of 0.48 (SSM symptoms) and of 0.52 (SSM function), respectively; SSM = Spinal Stenosis Measure

When a patient is above a horizontal line, he reached the respective MCID for EQ-5D-3L. When a patient is on the left side of the vertical line, he reached MCID for SSM symptoms or function, respectively.

Table 1: Patient characteristics at baseline

Group	Somewhat satisfied	Unsatisfied	p	Overall
n	43	53		96
Age, years, mean (SD)	74.95 (7.88)	73.83 (8.31)	0.50	74.33 (8.10)
Female, n (%)	26 (60.5)	32 (60.4)	1.00^	58 (60.4)
Social risk*, n (%)	17 (39.5)	18 (34.0)	0.73^	35 (36.5)
Education (compulsory school only), n (%)	11 (25.6)	14 (26.4)	1.00^	25 (26.0)
BMI, kg/m^2, mean (SD)	28.59 (4.66)	27.20 (4.81)	0.16	27.82 (4.77)
Smoker, n (%)	7 (16.3)	10 (18.9)	0.95^	17 (17.7)
CIRS, mean (SD)	10.58 (3.36)	10.75 (4.57)	0.84	10.68 (4.06)
Duration of symptoms >6 months, n (%)	31 (72.1)	42 (79.2)	0.57^	73 (76.0)
Spondylolisthesis[+] on operated level, n (%)	28 (65.1)	34 (64.2)	1.00^	62 (64.6)
EQ-5D-3L summary index, mean (SD)	0.45 (0.28)	0.35 (0.31)	0.10	0.40 (0.30)
Days between initial operation and 12-month follow-up, mean (SD)	340.7 (33.0)	341.7 (39.1)	0.90	341.2 (36.3)
Days between baseline and 12-month follow-up, mean (SD)	367.0 (8.1)	368.9 (10.8)	0.34	368.1 (9.7)

All p-values are from a Welch two-sided t-test unless otherwise noted.

CIRS, Cumulative Illness Rating Scale

[+] Meyerding listhesis grade ≥1

* Living alone, or single/divorced/widowed and living in a nursing/residential home.

[^] p-value from a chi-squared test.