The Spine Journal 19 (2019) 637−644

THE SPINE JOURNAL

Clinical Study

Patient-reported outcomes unbiased by length of follow-up after lumbar degenerative spine surgery: Do we need 2 years of follow-up?

Victor E. Staartjes, BMed[a,b,*], Alessandro Siccoli, BMed[b], Marlies P. de Wispelaere, MSc[c], Marc L. Schröder, MD, PhD[a]

[a] *Department of Neurosurgery, Bergman Clinics Amsterdam, Naarden, Rijksweg 69, 1411 GE, Naarden, The Netherlands*
[b] *Faculty of Medicine, University of Zurich, Zurich, Switzerland*
[c] *Department of Clinical Informatics, Bergman Clinics Amsterdam, Naarden, The Netherlands*

Received 23 April 2018; revised 2 October 2018; accepted 3 October 2018

Abstract

BACKGROUND: In modern clinical research, the accepted minimum follow-up for patient-reported outcome measures (PROMs) after lumbar spine surgery is 24 months, particularly after fusion. Recently, this minimum requirement has been called into question.
PURPOSE: We aim to quantify the concordance of 1- and 2-year PROMs to evaluate the importance of long-term follow-up after elective lumbar spine surgery.
STUDY DESIGN: Retrospective analysis of data from a prospective registry.
PATIENT SAMPLE: We identified all patients in our prospective institutional registry who underwent degenerative lumbar spine surgery with complete baseline, 12-month, and 24-month follow-up for ODI and numeric rating scales for back and leg pain (NRS-BP and NRS-LP).
OUTCOME MEASURES: Oswestry Disability Index (ODI) and NRS-BP and NRS-LP at 1 year and at 2 years.
METHODS: We evaluated concordance of 1- and 2-year change scores by means of Pearson's product-moment correlation and performed logistic regression to assess if achieving the minimum clinically important difference (MCID) at 12 months predicted 24-month MCID. Odds ratios (OR) and their 95% confidence intervals (CI), as well as model areas-under-the-curve were obtained.
RESULTS: A total of 210 patients were included. We observed excellent correlation among 12- and 24-month ODI ($r=0.88$), NRS-LP ($r=0.76$) and NRS-BP ($r=0.72$, all $p<.001$). Equal results were obtained when stratifying for discectomy, decompression, or fusion. Patients achieving 12-month MCID were likely to achieve 24-month MCID for ODI (OR: 3.3, 95% CI: 2.4−4.1), NRS-LP (OR: 2.99, 95% CI: 2.2−4.2) and NRS-BP (OR: 3.4, 95% CI: 2.7−4.2, all $p<.001$) with excellent areas-under-the-curve values of 0.81, 0.77, and 0.84, respectively. Concordance rates between MCID at both follow-ups were 87.2%, 83.8%, and 84.2%. A post-hoc power analysis demonstrated sufficient statistical power.
CONCLUSIONS: Irrespective of the surgical procedure, 12-month PROMs for functional disability and pain severity accurately reflect those at 24 months. In support of previous literature, our results suggest that 12 months of follow-up may be sufficient for evaluating spinal patient care in clinical practice as well as in research.

Keywords: Decompression; Discectomy; Degenerative; Fusion; Outcome measurement; Patient-reported outcomes.

FDA device/drug Note: Not applicable.

Author disclosures: ***VES:*** Nothing to disclose; ***AS:*** Nothing to disclose; ***MPW:*** Nothing to disclose; ***MLS:*** Nothing to disclose.

Conflict of interest: M.S. is a consultant to Mazor Robotics, Ltd. The other authors declare that the article and its content were composed in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

* Corresponding author. Department of Neurosurgery, c/o Bergman Clinics, Naarden, Rijksweg 69, 1411 GE, Naarden, Amsterdam, The Netherlands. Tel.: 0031 88 900 0500; fax 0031 88 900 0568.

E-mail address: victor.staartjes@gmail.com (V.E. Staartjes).

https://doi.org/10.1016/j.spinee.2018.10.004
1529-9430/

Introduction

Patient-reported outcome measures (PROMs) are the central tool for clinical research in spine surgery, as well as for effective follow-up in day-to-day patient care. For this reason, a large number of national and institutional registries have been initiated [1]. In contrast to clinical trials, these registries capture actual care in real-world clinical situations among large patient populations and heavily rely on PROMs. These instruments are validated, reliable, heavily correlated with patient satisfaction, and can even be adjusted to an individual patient's needs [2−5]. Over 70% of spine surgeons regularly use PROMs in their clinical practice [6].

In recent years, the consensus has been that a minimum follow-up of 2 years after surgery should be required for clinical research in degenerative spine care [7−9]. This applies foremost to outcomes after lumbar fusion surgery, where some journals and societies even require this minimum follow-up duration for clinical studies [7]. However, primary endpoints at 1-year are being reported at an increasing rate [10,11]. Furthermore, the maximum improvement after spine surgery is usually met at around 1-year postoperatively, which underlines the importance of this timepoint when comparing treatment effectiveness [9]. Longer follow-up unquestionably adds value to capture reoperations and adjacent segment degeneration, but multiple years of follow-up are often difficult to achieve in large studies. Additionally, at long-term follow-up, the validity of their results can be hampered by up to 79% of dropout [1,12].

As study methodologies evolve over the years, it may be suitable to adopt 1-year measures as primary endpoints in comparative clinical trials focused on PROMs [9,11,13]. Previous studies have shown that a period of 12 months is sufficient for a comprehensive assessment after deformity [7] and fusion [11,14] surgery. However, no analysis has been carried out for lumbar decompression, and whereas it has been established that short-term outcomes agree with 1-year outcomes after lumbar discectomy [15], it is currently unknown if 1 year of follow-up reflects long-term results in this patient population. For this reason, the purpose of this study was to quantify the agreement of 1- and 2-year PROMs after surgery for degenerative pathologies of the lumbar spine.

Materials and methods

Patient population

From a prospective institutional registry of lumbar and cervical spine interventions, we identified all patients who underwent lumbar spine surgery for degenerative diseases. From these, those who had a complete baseline, as well as 12-month follow-up and 24-month follow-up in PROMs were included in this study. All patients were operated on between December 2010 and January 2018 at a specialized spine center by a senior neurosurgeon. They underwent minimally invasive robot-guided transforaminal or posterior lumbar interbody fusion, transaxial lumbar interbody fusion, anterior lumbar interbody fusion, mini-open lumbar decompression, or tubular microdiscectomy [16,17]. This registry was authorized by the local institutional review board (Medical Research Ethics Committees United, Registration Number W16.065), and this study was carried out in accordance with the Declaration of Helsinki. All the patients included in this study provided written informed consent.

Data collection

For baseline measurement, patients completed a standardized questionnaire including numeric rating scales (NRS) for back pain and leg pain severity, and a validated Dutch version of the Oswestry Disability Index (ODI) to capture functional disability [3]. These instruments have been widely applied in spinal patient care, and have demonstrated good construct validity. At 12 months and 24 months after surgery, scheduled follow-up questionnaires containing the same outcome measurements were automatically dispatched to patients by use of a web-based tool designed for this purpose by the Department of Clinical Informatics of our institution. Clinical and surgical data were registered in separate databases.

Statistical analysis

Continuous data are given as mean ± standard deviation, and categorical data as numbers and percentages. We considered the correlation of 12- and 24-month ODI values our primary endpoint. A power analysis for our primary endpoint was performed to detect a correlation of 0.3 or greater at a significance level of $\alpha = 0.05$ and our sample size. The minimum clinically important difference (MCID) was defined as a 10 point difference among baseline and follow-up scores for the ODI, and as a two point difference for NRS scores, as defined by Ostelo et al. [18]. We evaluated the association of 12-month and 24-month change scores by Pearson's product-moment correlation. Pearson correlation coefficients (r) and significance values for convergent validity are provided. Binomial logistic regression models were trained and evaluated to determine the likelihood of patients achieving MCID at the 12-month follow-up to also achieve MCID at the 24-month follow-up. Odds ratios (OR) and their 95% confidence intervals (CI) are provided. The area under the receiver operating characteristics curve (AUC) was calculated for these models. Concordance rates for MCID at both follow-ups, defined as the proportion of patients without change in MCID state, were also obtained. All analyses were carried out in R version 3.4.4 (The R Foundation for Statistical Computing, Vienna, Austria). A $p \leq .05$ on a two-tailed test was considered statistically significant.

Results

The flow of patients throughout this analysis is shown in Fig. 1. Data from 4,329 patients was available, of which 3,279 (75.7%) underwent lumbar spine surgery. All patients had complete clinical, intra- and perioperative data. 210 (100%) patients had a complete baseline, 12-month and 24-month measurements for ODI, NRS back pain and NRS leg pain. An overview of demographic and clinical data of the included patients is provided in Table 1. Our post-hoc power analysis determined a statistical power of 0.994 for the analysis of our primary endpoint with a sample size of 210 patients.

Patient-reported outcome measures

All PROMs improved significantly from baseline to 12-month and 24-month follow-ups (all $p < .001$, Table 2). For the ODI, we measured a mean improvement of 30.3 ± 22.8 after 1 year and 29.8 ± 22.9 after 2 years. For NRS leg pain, the 1-year improvement amounts to 4.9 ± 3.4 and the 2-year improvement to 4.8 ± 3.6. NRS back pain improved by 2.5 ± 3.5 after 1 year and by 2.6 ± 3.4 after 2 years. MCID for ODI was reached in 166 patients (79%) after 1 year and in 165 patients (79%) after 2 years. MCID for NRS leg pain was achieved in 163 patients (78%) after 1 year and in 151 patients (72%) after 2 years, and in NRS back pain in 105 patients (50%) after 1 year and in 98 patients (47%) after 2 years. Density plots (Fig. 2) demonstrate the distribution of differences between the two PROM measurements.

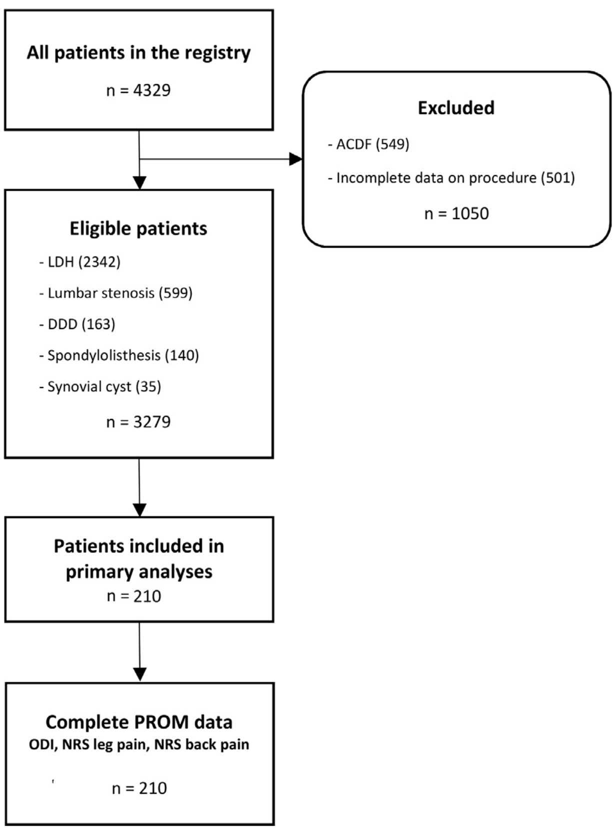

Fig. 1. The flow of patients throughout this analysis.

Table 1
Baseline patient characteristics

Characteristic	Value
Active Smoker	34 (16%)
Male gender	113 (54%)
Age	52.2 ± 12.4
BMI [kg/m^2]	25.6 ± 3.4
Indication	
LDH	156 (74%)
Lumbar stenosis	36 (17%)
DDD	10 (4.8%)
Spondylolisthesis	6 (2.9%)
Synovial cyst	2 (1.0%)
ASA score	
Class I	109 (53%)
Class II	95 (46%)
Class III	1 (0.5%)
Index level	
L1-L2	2 (1.0%)
L2-L3	8 (3.8%)
L3-L4	30 (14%)
L4-L5	86 (41%)
L5-S1	82 (39%)
Procedure	
tMD	150 (71.4%)
Decompression	40 (19%)
MI-TLIF	6 (2.9%)
MI-PLIF	7 (3.3%)
ALIF	6 (2.9%)
AxiaLIF	1 (0.5%)
Baseline PROM	
ODI	44.4 ± 17.5
NRS leg pain	6.9 ± 5.3
NRS back pain	5.3 ± 3.0

BMI, Body Mass Index; LDH, lumbar disc herniation; DDD, degenerative disc disease; ASA, American Society of Anesthesiologists; tMD, tubular microdiscectomy; MI-TLIF, minimally invasive transforaminal lumbar interbody fusion; MI-PLIF, minimally invasive posterior lumbar interbody fusion; ALIF, anterior lumbar body fusion; AxiaLIF, axial lumbar body fusion; PROM, patient-reported outcome measurement; ODI, Oswestry Disability Index, NRS, numeric rating scale.

Table 2
Change scores for ODI, NRS leg pain and NRS back pain. p Values for longitudinal differences from the baseline measurement to the 12-month and 24-month follow-ups are provided, respectively

Measure	Value	p
ODI		
12-month change score	30.3 ± 22.8	<.001*
24-month change score	29.8 ± 22.9	<.001*
NRS leg pain		
12-month change score	4.9 ± 3.4	<.001*
24-month change score	4.8 ± 3.6	<.001*
NRS back pain		
12-month change score	2.5 ± 3.5	<.001*
24-month change score	2.6 ± 3.4	<.001*

ODI, Oswestry Disability Index; NRS, numeric rating scale.
* $p \leq .05$

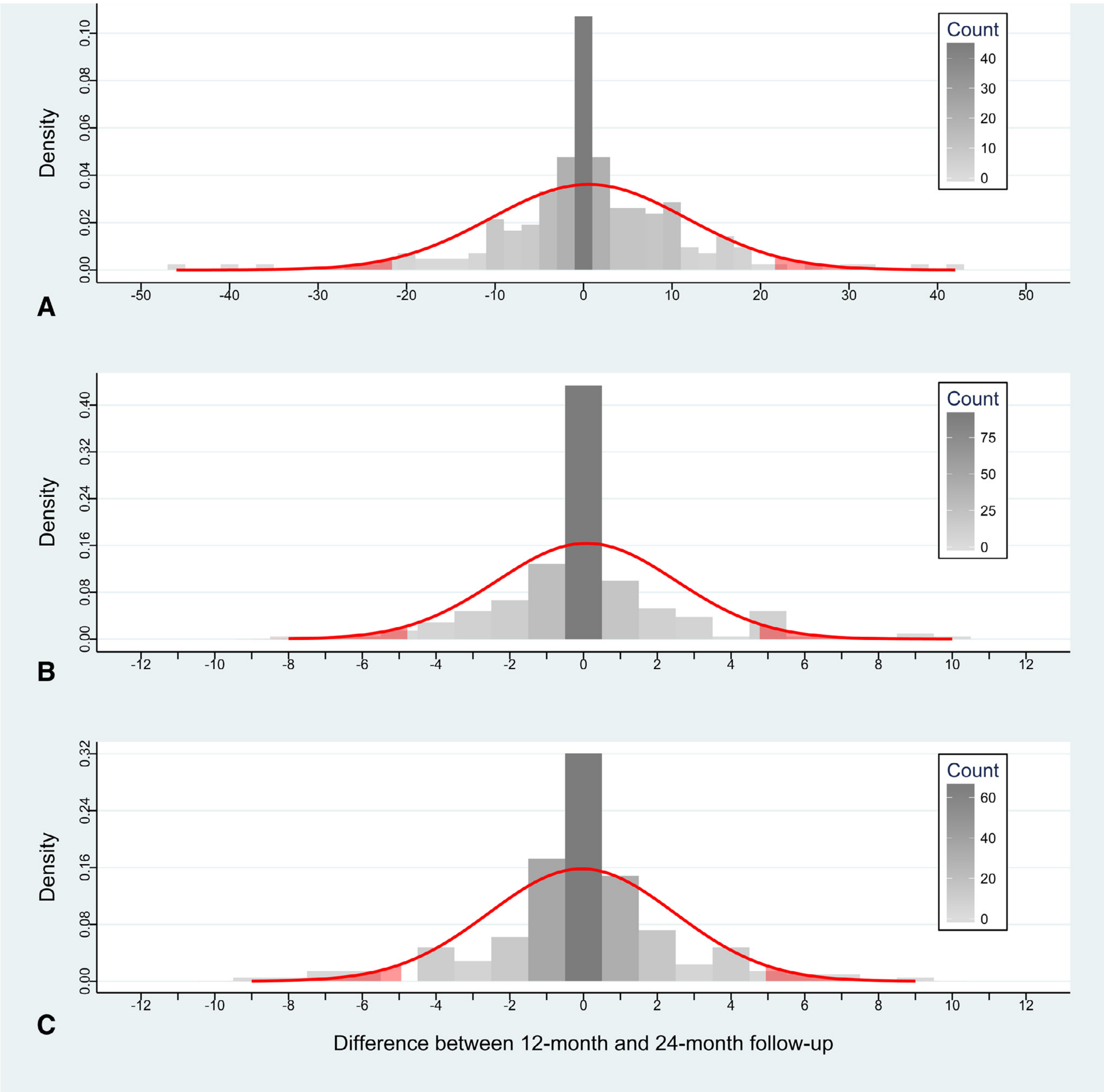

Fig. 2. Density plots illustrating the differences in patient-reported outcomes between the 12-month and 24-month follow-up for the Oswestry Disability Index (A), as well as for numeric rating scales (NRS) for leg pain (B) and back pain (C) severity.

Convergent validity

We observed a direct correlation of 12-month and 24-month PROMs as measured by ODI ($r = 0.88$, 95% CI: 0.85−0.91), NRS for leg pain severity ($r = 0.76$, 95% CI: 0.70−0.81), and NRS back pain severity ($r = 0.72$, 95% CI: 0.65−0.78, all $p < .001$). Bland-Altman plots demonstrate a mean bias and 95% CI limits for the difference between the two measurements of 0.57 (−21.1 to 22.2, Fig. 3) for ODI, 0.08 (−4.7 to 4.9, Fig. 4) for NRS leg pain severity, and −0.05 (−5.0 to 4.9, Fig. 5) for NRS back pain severity. When performing the same analysis stratified by type of surgical procedure, equal results were observed for discectomy, decompression, and spinal fusion (Table 3). All of these Pearson correlation coefficients can be classified as "large" in effect size as defined by Cohen [19].

Minimum clinically important difference

Concordance rates for ODI (87.2%), NRS leg pain (83.8%), and NRS back pain (84.2%) were high (Table 4). Logistic regression analysis revealed that patients achieving the MCID at 12-month follow-up were also more likely to

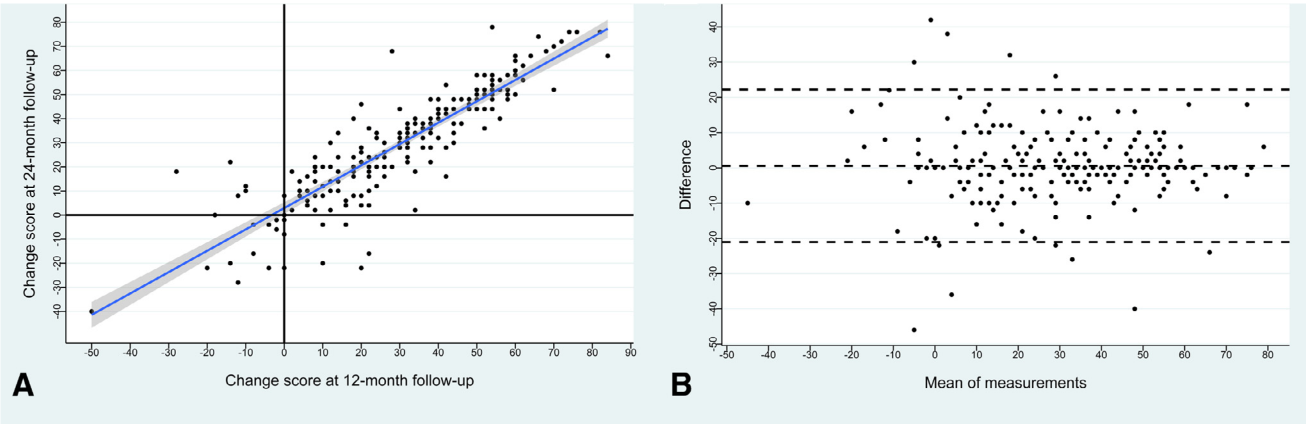

Fig. 3. Scatterplot (A) demonstrating the correlation of 12-month and 24-month Oswestry Disability Index (ODI) values. We observed a correlation coefficient of $r = 0.88$. The Bland-Altman plot (B) demonstrates the mean bias of 0.57 (−21.1 to 22.2) between ODI measurements.

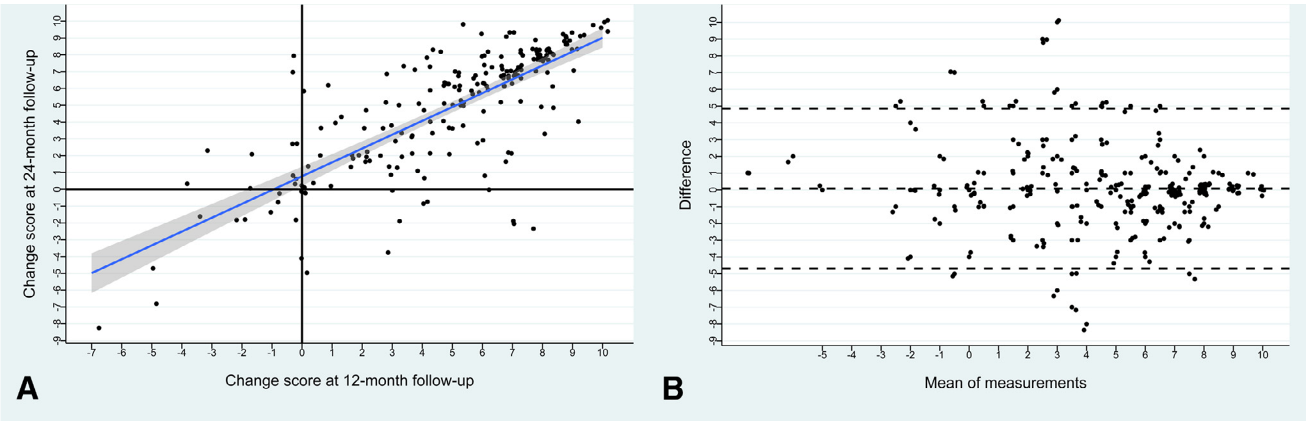

Fig. 4. Scatterplot (A) demonstrating the correlation of 12-month and 24-month leg pain severity as measured by the numeric rating scale (NRS-LP). We observed a correlation coefficient of $r = 0.76$. The Bland-Altman plot (B) demonstrates the mean bias of 0.08 (−4.7 to 4.6) between NRS-LP measurements. Plots have been jittered to illustrate density of multiple coinciding datapoints.

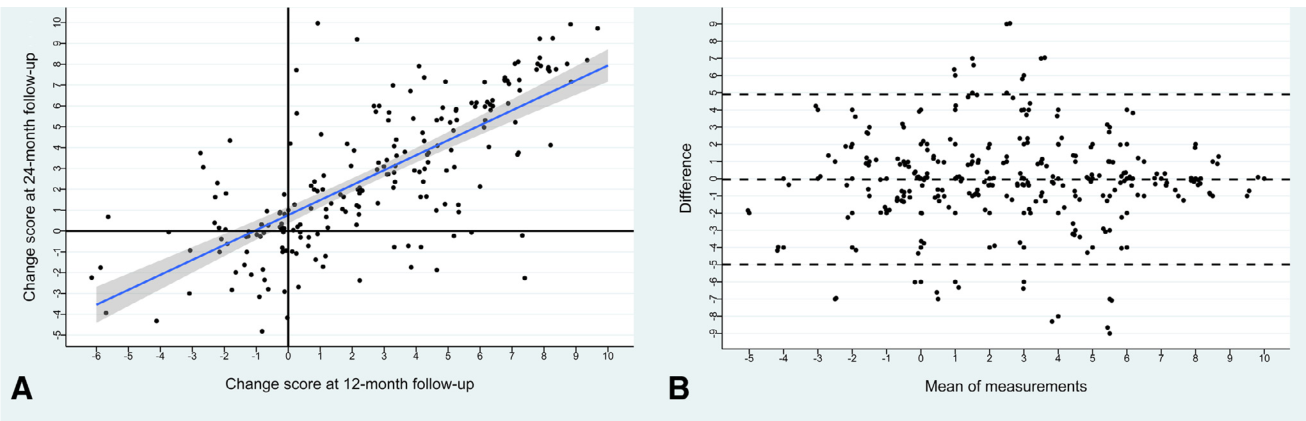

Fig. 5. Scatterplot (A) demonstrating the correlation of 12-month and 24-month back pain severity as measured by the numeric rating scale (NRS-BP). We observed a correlation coefficient of $r = 0.72$. The Bland-Altman plot (B) demonstrates the mean bias of −0.05 (−5.0 to 4.9) between NRS-BP measurements. Plots have been jittered to illustrate density of multiple coinciding datapoints.

reach MCID at 24-month follow-up. For ODI, we measured an OR = 3.25 (2.44−4.14, $p < .001$), while we measured an OR = 2.99 (2.22−3.84, $p < .001$) for NRS leg pain, and an OR = 3.39 (2.67−4.19, $p < .001$) for NRS back pain.

Discussion

In an analysis of 210 patients from a prospective registry of patients that underwent surgery for lumbar degenerative

Table 3
Association of 12-month and 24-month change scores in PROM as assessed by Pearson's product-moment correlation. Pearson correlation coefficients, as well as their respective 95% confidence intervals are given for the overall study population, and for each type of surgical procedures specifically

Measure	Overall			Discectomy			Decompression			Fusion		
	r	95% CI	*p*	*r*	95% CI	*p*	*r*	95% CI	*p*	*r*	95% CI	*p*
ODI	0.88	0.85−0.91	<.001*	0.88	0.84−0.91	<.001*	0.91	0.83−0.95	<.001*	0.80	0.55−0.92	<.001*
NRS leg pain	0.76	0.70−0.81	<.001*	0.76	0.68−0.82	<.001*	0.78	0.61−0.88	<.001*	0.73	0.43−0.89	<.001*
NRS back pain	0.72	0.65−0.78	<.001*	0.72	0.63−0.79	<.001*	0.71	0.51−0.83	<.001*	0.72	0.40−0.88	<.001*

CI, confidence interval; ODI, Oswestry Disability Index; NRS, numeric rating scale.
* p ≤0.05

Table 4
Logistic regression models of 12-month MCID predicting 24-month MCID. Odds ratios, as well as their 95% confidence intervals, and the area under the curve of the models are provided. The concordance rate between MCID at both follow-ups is reported

Measure	Odds ratio	95% CI	*p*	AUC	Concordance
ODI	3.3	2.4−4.1	<.001*	0.81	87.2%
NRS leg pain	3.0	2.2−3.8	<.001*	0.77	83.8%
NRS back pain	3.4	2.7−‘‘4.2	<.001*	0.84	84.2%

CI, confidence interval; AUC, area under the receiver operating characteristics curve; ODI, Oswestry Disability Index; NRS, numeric rating scale.
* p ≤.05

pathologies, the agreement of 12-month and 24-month PROMs was quantified. We found excellent correlations in measures of functional disability as well as in back and leg pain scales. These correlations were independent of the type of surgical procedure. Furthermore, clinical success at 12 months exhibited a strong association and high concordance with clinical success at 24 months postoperatively. Our data suggest that 12 months of follow-up after degenerative spine surgery may be adequate for identification of effective versus ineffective surgical treatment.

Our study is among the first to methodologically assess the association of 1- and 2-year PROMs, and the first to do so for lumbar decompression [7,11,14,15]. Data were taken from a prospective registry that included patients who were surgically treated for a broad spectrum of degenerative spinal pathologies. This allowed us to make conclusions about not only the individual techniques, but also about the general patient population seen by most spine surgeons dealing with degenerative spinal pathologies on a day-to-day basis. Every analysis performed found significant agreement between the two timepoints, with large effect sizes. This is best demonstrated by the AUC values for the logistic regression models, which range between 0.77 and 0.84 and thus indicate a very strong relationship between clinical success at 1 year and at 2 years. Subgroup analyses stratified into discectomy, decompression, and fusion procedures revealed effect sizes ranging around those found for the overall cohort.

Adogwa et al. found Pearson correlation coefficients of 0.82, 0.85, and 0.90 for ODI, leg pain, and back pain after lumbar fusion, respectively [11]. Furthermore, MCID at 1 year was highly predictive of MCID at 2 years. Glassman et al. also found that there was no difference in 1- and 2-year ODI or Short Form-36 after lumbar fusion [14] as well as after deformity surgery [7]. However, they reported that 6-month outcomes did not agree with long-term results. Asher et al. found that 3-month ODI cannot reliably serve as a proxy for long-term patient experience, and they concluded that prospective longitudinal registries should span at least 12 months to determine effectiveness of spine care for individual patients [20]. Parker et al. corroborated these findings in a prospective study [21]. Whitmore et al. analyzed the agreement of 3-month and 12-month outcomes in discectomy patients from a large prospective registry, and determined that early follow-up is sufficient for patients who do not have to undergo reoperation in this patient population [15]. They also calculated that shortening of the follow up period leads to a relevant reduction in related study costs. These previous analyses make up a solid evidence base for this topic. Our results corroborate these findings in a mixed patient population representative of clinical practice, and confirm that these principles are also true for lumbar stenosis. Symptom improvement seems to happen primarily during the first postoperative year, after which outcomes remain relatively stable [9,11,15]. Based on our data, in combination with the peer-reviewed literature, it is likely that PROMs at 12 months are an accurate representation of those at 24 months.

Patient-reported measures of health status, and particularly ODI, Short For-36, and EuroQOL-5D, have become pivotal in judging the quality of surgical intervention in the lumbar spine. However, these measures of treatment effectiveness alone may not always correspond to clinically important improvement [5,11,22]. The concept of MCID has been introduced to distinguish findings that are merely statistically significant in high-powered trials from those that actually represent a clinically meaningful improvement for the patient. The MCID is a quantitative indicator of clinical relevance specific to a certain outcome measure. Various methods to determine MCID, such as the minimum detectable change, AUC-derived, or Delphi methods, have been used to determine sensible MCID thresholds [10,18,22]. We opted to use the values suggested by a consensus statement that applied a combination of evidence synthesis from the literature, a survey, and an expert panel [18]. The choice of MCID estimation methods evidently has a crucial influence on methodological studies, like our present one, and on clinical trials. Thus, it is conceivable that a different choice of MCID values could introduce a bias in studies looking at the correlation among 1- and 2-year results, and limits the comparability between those studies [7,11,14,22]. Generally, there is a striking lack of uniformity in methods and reporting in terms of MCID choice in the lumbar spine community.

PROM assessment in a large number of patients can be hard to achieve. It is crucial that cohort studies with long-term follow-up are carried out to estimate the incidence of disease recurrence, reoperation, instrumentation failure, and mortality, among other critical parameters. However, a significant part of the clinical studies in spine surgery primarily focus on PROMs instead of the aforementioned measures, which are only measurable with long-term follow-up.

Complications and reoperations heavily influence PROMs after the primary surgical procedure [23]. Even when including patients who experienced complications and who have undergone reoperation, as we did, 1-year outcomes were highly correlated to those at 2-years. Both owing to the effort which long-term follow-up requires, as well as to significant dropout of up to 79%, it is becoming increasingly difficult to manage 2 years of follow-up for every clinical study [1]. Methods for missing data simulation keep improving with the introduction of data science techniques like multiple imputation [24]. Although these techniques have become highly effective for small proportions of missing values, caution should be taken when imputing larger amounts of data to salvage statistical power.

Research methods are constantly evolving [5,13]. A prime example is the shift from radiological assessments such as bony fusion rate to outcome-based assessments of treatment effect [5,7]. As we continue improving the standards and tools for quality assessment, insisting on a minimum of 2 years of follow-up may be redundant. In an analysis of over 4,000 patients, Fekete et al. showed that patients who required renewed intervention due to poor outcomes are already identifiable after 12 months, and they conclude that insistence on a 2-year follow-up could thus result in a failure to intervene early enough [9]. Although long-term follow-up certainly adds value for research specific to outcomes such as reoperations and adjacent level degeneration, the vast majority of comparative trials and registries that specify PROMs as their primary endpoint will be well-served by 1-year outcomes. The greatest improvement in PROMs after surgery is usually observed at around 1 year postoperatively [9]. Thus, it is conceivable that this is a reliable timepoint for identification of between-group differences in comparative trials, and probably allows for the most accurate determination of treatment effect size in cohort studies. Even in cases where long-term outcomes are of interest, the 1-year measures could potentially serve as a proxy for 2-year measures, which is only given for discectomy at 3 months, but not for 3- or 6-month outcomes in other patient populations [7,11,14,15,20,21].

Limitations

Although all data were taken from a registry of prospectively collected data, the performed analysis was not preplanned, and may thus be hampered by its retrospective nature. Furthermore, all data stem from a single center possibly creating center-related bias and thereby limiting the generalizability of our findings. Only 6% of patients in our registry fit the inclusion criteria, since a majority of patients did not report complete 12- or 24-month follow-up requests. This reflects the difficulty of obtaining long-term outcomes, especially when patient incentive reimbursements are not used [15]. Although our primary analyses were more than adequately powered, those stratified by subgroups of surgical procedures may not have achieved sufficient statistical power. Such subgroup analyses, in general, must be taken with a grain of salt. We used accepted MCID thresholds. However, a different choice of MCID thresholds could have produced diverging results, limiting the generalizability of our findings. PROM values were not adjusted for perioperative complications and reoperations. This could be seen as strength since we were able to demonstrate that PROMs at the two timepoints correlated even when including these heterogeneities. Based on our study, we can only make claims on the agreement on the analyzed PROMs, but not on other outcome measures such as measures of health-related quality of life or objective functional tests. Lastly, we can make no statement on the generalizability of our findings to patients receiving conservative treatment.

Conclusions

In support of previous literature, our data indicate that PROMs at 12 months probably are an accurate representation of those at 24 months. This is true for a broad range of surgical interventions in the degenerated spine. It is likely that a 1-year follow-up is sufficient to identify effective patient care, and may even have distinct logistical and possibly economic added value for long-term registries. Prospective validation in other cohorts will be necessary to establish these findings with higher level of evidence.

Acknowledgment

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] McGirt MJ, Parker SL, Asher AL, Norvell D, Sherry N, Devin CJ. Role of prospective registries in defining the value and effectiveness of spine care. Spine 2014;39:S117. doi: 10.1097/BRS.0000000000000552.

[2] Deyo RAM, Battie M, Beurskens AJHM, Bombardier CM, Croft PM, Koes B, et al. Outcome measures for low back pain research: a proposal for standardized use. Spine 1998;23:2003–13.

[3] Van Hooff ML, Spruit M, Fairbank JCT, van Limbeek J, Jacobs WCH. The Oswestry Disability Index (version 2.1a): validation of a Dutch language version. Spine 2015;40:E83–90. doi: 10.1097/BRS.0000000000000683.

[4] Schroeder GD, Coric D, Kim HJ, Albert TJ, Radcliff KE. Are patient-reported outcomes predictive of patient satisfaction 5 years after anterior cervical spine surgery? Spine J 2017;17:943–52. doi: 10.1016/j.spinee.2017.02.008.

[5] Scheer JK, Keefe M, Lafage V, Kelly MP, Bess S, Burton DC, et al. Importance of patient-reported individualized goals when assessing outcomes for adult spinal deformity (ASD): initial experience with a Patient Generated Index (PGI). Spine J 2017;17:1397–405. doi: 10.1016/j.spinee.2017.04.013.

[6] Falavigna A, Dozza DC, Teles AR, Wong CC, Barbagallo G, Brodke D, et al. Current status of worldwide use of patient-reported outcome measures (PROMs) in spine care. World Neurosurg 2017;108:328–35. doi: 10.1016/j.wneu.2017.09.002.

[7] Glassman SD, Schwab F, Bridwell KH, Shaffrey C, Horton W, Hu S. Do 1-year outcomes predict 2-year outcomes for adult deformity surgery? Spine J 2009;9:317–22. doi: 10.1016/j.spinee.2008.06.450.

[8] Birkmeyer NJO, Weinstein JN, Tosteson ANA, Tosteson TD, Skinner JS, Lurie JD, et al. Design of the Spine Patient outcomes Research Trial (SPORT). Spine 2002;27:1361–72.

[9] Fekete TF, Loibl M, Jeszenszky D, Haschtmann D, Banczerowski P, Kleinstück FS, et al. How does patient-rated outcome change over time following the surgical treatment of degenerative disorders of the thoracolumbar spine? Eur Spine J 2018;27:700–8. doi: 10.1007/s00586-017-5358-2.

[10] Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. 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;8:968–74. doi: 10.1016/j.spinee.2007.11.006.

[11] Adogwa O, Elsamadicy AA, Han JL, Cheng J, Karikari I, Bagley CA. Do measures of surgical effectiveness at 1 year after lumbar spine surgery accurately predict 2-year outcomes? J Neurosurg Spine 2016;25:689–96. doi: 10.3171/2015.8.SPINE15476.

[12] Højmark K, Støttrup C, Carreon L, Andersen MO. Patient-reported outcome measures unbiased by loss of follow-up. Single-center study based on DaneSpine, the Danish spine surgery registry. Eur Spine J 2016;25:282–6. doi: 10.1007/s00586-015-4127-3.

[13] Shultz BN, Bovonratwet P, Ondeck NT, Ottesen TD, McLynn RP, Grauer JN. Evaluating the effect of growing patient numbers and changing data elements in the national surgical quality improvement program (NSQIP) database over the years: a study of posterior lumbar fusion outcomes. Spine J 2018;0. doi: 10.1016/j.spinee.2018.03.016.

[14] Glassman S, Gornet MF, Branch C, Polly D, Peloza J, Schwender JD, et al. MOS Short Form 36 and Oswestry Disability Index outcomes in lumbar fusion: a multicenter experience. Spine J 2006;6:21–6. doi: 10.1016/j.spinee.2005.09.004.

[15] Whitmore RG, Curran JN, Ali ZS, Mummaneni PV, Shaffrey CI, Heary RF, et al. Predictive value of 3-month lumbar discectomy outcomes in the NeuroPoint-SD Registry. J Neurosurg Spine 2015;23:459–66. doi: 10.3171/2015.1.SPINE14890.

[16] Staartjes VE, Vergroesen P-PA, Zeilstra DJ, Schröder ML. Identifying subsets of patients with single-level degenerative disc disease for lumbar fusion: the value of prognostic tests in surgical decision making. Spine J 2017. doi: 10.1016/j.spinee.2017.08.242.

[17] Staartjes VE, de Wispelaere MP, Miedema J, Schröder ML. Recurrent lumbar disc herniation after tubular microdiscectomy: analysis of learning curve progression. World Neurosurg 2017. doi: 10.1016/j.wneu.2017.07.121.

[18] Ostelo RWJG, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 2008;33:90–4. doi: 10.1097/BRS.0b013e31815e3a10.

[19] Cohen J. Chapter 4 - Differences between Correlation Coefficients. Stat. Power Anal. Behav. Sci. Revis. Ed. 1977: 109–43 Academic Press. doi: 10.1016/B978-0-12-179060-8.50009-8.

[20] Asher AL, Chotai S, Devin CJ, Speroff T, Harrell FE, Nian H, et al. Inadequacy of 3-month Oswestry Disability Index outcome for assessing individual longer-term patient experience after lumbar spine surgery. J Neurosurg Spine 2016;25:170–80. doi: 10.3171/2015.11.SPINE15872.

[21] Parker SL, Asher AL, Godil SS, Devin CJ, McGirt MJ. Patient-reported outcomes 3 months after spine surgery: is it an accurate predictor of 12-month outcome in real-world registry platforms? Neurosurg Focus 2015;39:E17. doi: 10.3171/2015.9.FOCUS15356.

[22] Parker SL, Adogwa O, Paul AR, Anderson WN, Aaronson O, Cheng JS, 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. doi: 10.3171/2010.12.SPINE10472.

[23] Chotai S, Parker SL, Sivaganesan A, Sielatycki JA, Asher AL, McGirt MJ, et al. Effect of complications within 90 days on patient-reported outcomes 3 months and 12 months following elective surgery for lumbar degenerative disease. Neurosurg Focus 2015;39:E8. doi: 10.3171/2015.8.FOCUS15302.

[24] Ondeck NT, Fu MC, Skrip LA, McLynn RP, Cui JJ, Basques BA, et al. Missing data treatments matter: an analysis of multiple imputation for anterior cervical discectomy and fusion procedures. Spine J n.d. doi:10.1016/j.spinee.2018.04.001.