



Follow-up of degenerative lumbar spine surgery—PROMs stabilize after 1 year: an equivalence study based on Swespine data

C. Parai^{1,2} · O. Hägg² · B. Lind^{1,2} · H. Brisby^{1,3}

Received: 2 July 2018 / Revised: 20 March 2019 / Accepted: 21 April 2019 / Published online: 30 April 2019
© The Author(s) 2019

Abstract

Purpose To evaluate the outcome of degenerative lumbar spine surgery in a credible way, patient-reported outcome measures (PROMs) should be assessed after an adequate follow-up period. Most authors/journals consider a follow-up period of less than two years to be too short. The purpose of this study was to explore the possibility of restricting follow-up to one year.

Methods Adult patients operated between 1998 and 2017 were retrieved from Swespine (Lumbar Disc Herniation $n = 31,314$, Lumbar Spinal Stenosis $n = 53,043$ and Degenerative Disc Disease $n = 14,375$). The proportion reaching the minimal important change (MIC) in Visual Analogue Scale for pain (VAS_{BACK/LEG}), Oswestry Disability Index (ODI) and the quality-of-life measure EQ-5D_{INDEX} at 1 and 2 years, respectively, was calculated. The single-item questions such as Global Assessment (GA_{BACK/LEG}) and Satisfaction were analysed by the McNemar test. Threshold values for a successful outcome based on the final scores of each PROM at 1 and 2 years post-surgery were also defined.

Results For all the three diagnostic groups, the differences in proportions reaching MIC of each PROM at 1 and 2 years were below 2%. Global Assessment and Satisfaction with outcome at one year remained at 2 years. There were no important differences of threshold values of treatment success based on final scores

Conclusion No clinically important changes in PROMs appeared between 1 and 2 years after surgery for degenerative lumbar conditions, demonstrating that a follow-up period of 1 year as opposed to 2 years is sufficient in effectiveness studies if PROMs are to be used as outcome variables.

Graphical abstract

These slides can be retrieved under Electronic Supplementary Material.

The graphical abstract consists of three slides. The first slide, titled 'Key points', lists: 1. National register study, 2. Outcome Measurement, and 3. One year of follow-up is sufficient. The second slide, titled 'Fig. 3', contains four histograms showing the distribution of score changes in ODI and VAS between baseline and FU1 (1 year) and between FU1 and FU2 (2 years) for three diagnostic groups: LDD, LSS, and DDD. The histograms show that the distributions are very similar at 1 and 2 years. The third slide, titled 'Take Home Message', states: 'Follow-up with PROMs at both 1 year and 2 years is unnecessary for outcome measurement of degenerative lumbar spine surgery - 1 year is sufficient.'

Keywords Degenerative lumbar surgery · Patient-reported outcome measures · Follow-up · Equivalence

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00586-019-05989-0>) contains supplementary material, which is available to authorized users.

✉ C. Parai
catharina.parai@spinecenter.se

Extended author information available on the last page of the article

Introduction

The efficacy of a new surgical procedure should be tested in an RCT if possible or at least against a control group. Depending on the research question, the follow-up period may vary from months to decades. In effectiveness studies,

data from a high-quality registry offer a unique opportunity to study the outcome of interventions—whose efficacy is previously documented—on a large scale and in clinical praxis. To make conclusions about the outcome of surgery for degenerative spine conditions, most scientific journals and authors consider a follow-up period of at least two years to be necessary. Recent studies indicate, however, that a shorter evaluation time may be sufficient since the improvement curve for patient-reported outcome measures (PROMs) has been reported to level out and patients reach a threshold of change somewhere between 3 and 12 months after surgery [1, 2].

The Swedish Spine Register, Swespine, is controlling the quality of lumbar spine surgery by sending follow-up questionnaires at 1, 2, 5 and 10 years post-operatively. It offers a unique possibility to evaluate any differences in outcome between one- and two-year follow-up in a large real-life database [3]. The response rates are approximately 75% at follow-up 1 at one year post-op (FU1) and 65% at follow-up 2 at two years post-op (FU2).

Our aim was to study the outcome of lumbar surgery, measured with PROMs, at one and two years after procedure, with the specific question: are there clinically important differences in outcome between one- and two-year follow-up?

Patients and methods

Characteristics of the population

Analyses were performed on prospectively collected data from 98,732 degenerative lumbar spine surgical procedures registered in the Swespine spine register [4] between 1998 and 2017. To facilitate comparisons with previous trials and to find potential sub-population differences, the population was divided into the three diagnostic groups: Lumbar Disc Herniation (LDH), Lumbar Spinal Stenosis (LSS) and Degenerative Disc Disease (DDD).

The LDH group consisted of patients with disc herniation treated with most often solely a discectomy but for some supplemented with a fusion. The LSS group consisted of an older and more heterogeneous population with greater degenerative changes causing spinal stenosis. The number of levels, the severity of the degeneration and the location of the stenosis (in the central canal, in the lateral recess or in the neural foramen) generate a variety of treatment options from minimally invasive nerve root decompression where midline structures are preserved, to multilevel decompression and fusion. The DDD group conveyed patients with severe chronic low back pain as the main symptom and fusion or total disc replacement as the single treatment.

Inclusion and exclusion criteria are shown in the study flowchart in Fig. 1. Details of the registry have been previously described [5].

Outcome variables (patient-reported outcome measures—PROMs)

The Visual Analogue Scale for back (VAS_{BACK}) and leg (VAS_{LEG}) pain is measured on a scale from 0 (no pain) to 100 (worst imaginable pain) [6]; the disease-specific measure such as Oswestry Disability Index (ODI) [7] aims to capture important aspects of low back pain influencing function in everyday life activities with the index score ranging from 1 (best) to 100 (worst); the quality-of-life measure EuroQoL-5-Dimensions [8] is translated into a preference-based single index score based on a scale from 0.000 (death) to 1.000 (perfect health); the single-item transition questions such as Satisfaction ('How would you describe your satisfaction with the surgical outcome?' 1/Satisfied, 2/Uncertain, 3/dissatisfied) and Global Assessment for back pain (GA_{BACK}) and leg pain (GA_{LEG}) ('How is your back/leg pain today as compared to before the surgery?' where 0 represents no back/leg pain before the surgery, 1 completely pain free, 2 much better, 3 somewhat better, 4 unchanged, 5 worse) [9] are used at follow-up.

Study design

Our choice of an equivalence study design was based on the expectation that the result after a follow-up period of one year is not different from the result after a follow-up of two years by more than an acceptable amount—given that the endpoint variable is a PROM. Therefore, confirmation of similarity in this trial involved the pre-specification of the clinically relevant minimal important change (MIC) of improvement for each PROM.

Statistical methods

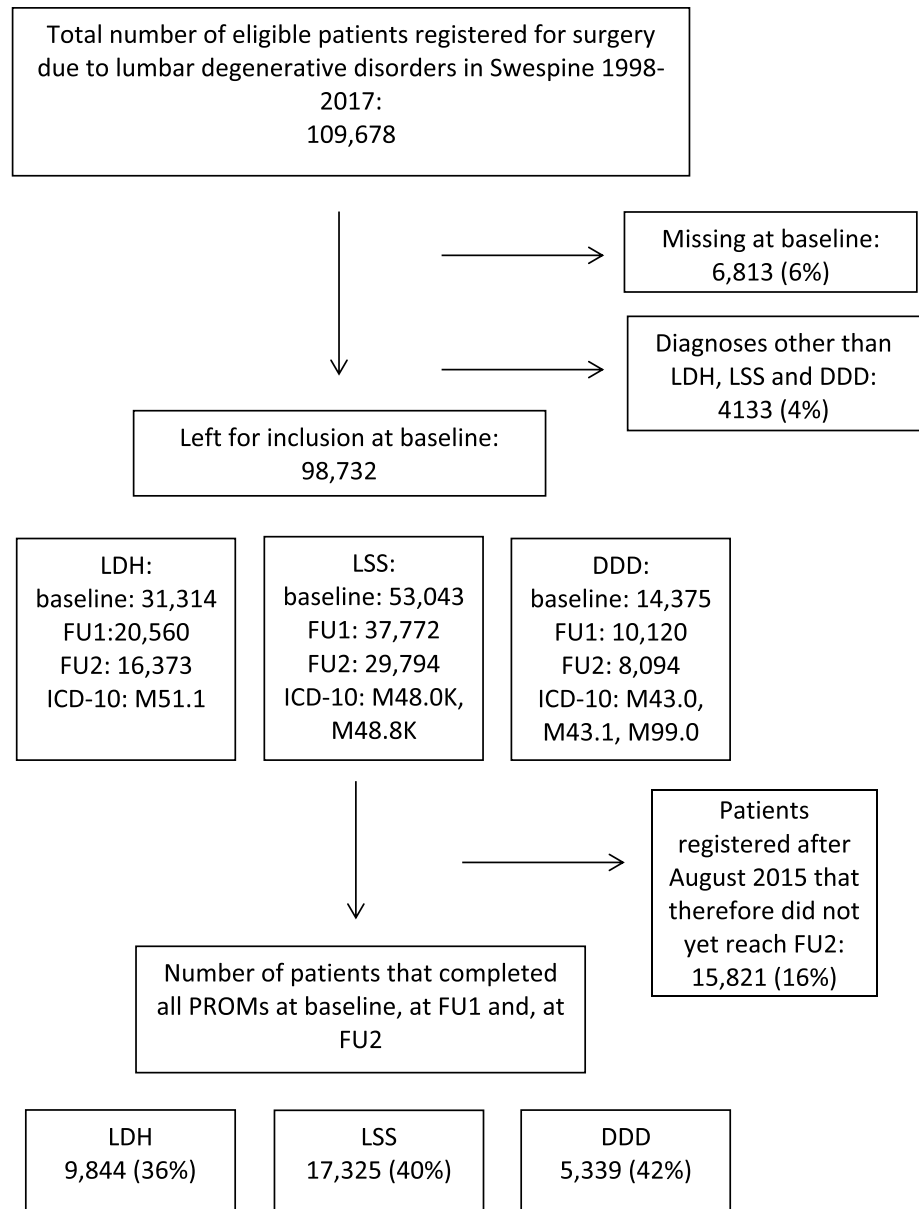
Descriptive data

Descriptive data are presented as means \pm SD or numbers (%).

Minimal important change (MIC)

The MIC represents the smallest amount of change in an outcome that is considered important to the patient. It is expressed in the units used in the measurement. MIC values were calculated for each PROM to set appropriate margins of clinical relevance in the analyses of possible

Fig. 1 Study flowchart. *LDH* Lumbar Disc Herniation, *LSS* Lumbar Spinal Stenosis, *DDD* Degenerative Disc Disorders, *FU1/2* follow-up at 1 and 2 years post-operatively



differences between outcomes at FU1 and FU2. In a ROC curve analysis, the MIC corresponds to the optimal ROC cut-off point, which is the value for which the sum of the percentages of false positive and false negative classifications is the smallest [10]. As reference criteria, the transition questions such as Global Assessment for back or leg pain ($GA_{BACK/LEG}$) were used. $GA_{BACK/LEG}$ was dichotomized into ‘success’, corresponding to a considerable improvement (i.e. those who responded as ‘completely pain free’ or ‘much better’) and ‘not success’ (i.e. those who responded ‘somewhat better’, ‘unchanged’ or ‘worse’). The choice of using an anchor-based approach was based on the recommendation by the COSMIN expert panel [11].

Clinically relevant differences between FU1 and FU2

For ODI, $VAS_{BACK/LEG}$ and EQ-5D, differences on a group level were explored by the calculation of mean differences between baseline and follow-up at one year and between the first year and second year. *p* values were not presented since even very small differences in large study populations tend to show a statistical significance. Instead, the differences were related to the MIC value as well as to a threshold of treatment success at one year post-operatively of each PROM. Thresholds of treatment success are an alternative to the MIC. The statistical method here was the same as the one for computing MIC values described above, but instead of score changes, final scores were used. The proportion of

patients reaching these cut-off points at FU1 and at FU2 were calculated.

For the retrospective single-item questions such as $GA_{BACK/LEG}$ and Satisfaction, the test for two paired proportions (McNemar's test) [12] was used to explore to what extent patients moved from the successful outcome group to the unfavourable outcome group or vice versa between 1 and 2 years of follow-up. $GA_{BACK/LEG}$ was dichotomized as described above. Satisfaction was dichotomized into 'success' (i.e. 'satisfied') and 'not success' (i.e. 'uncertain' or 'dissatisfied').

The SPSS version 24 was used for the statistical analyses except for the computation of the MIC values, where SAS JMP was used.

Ethical considerations

Informed consent was obtained from all participants in Swespine. This research project was approved by the regional ethical review board.

Results

ODI, $VAS_{BACK/LEG}$ and EQ-5D

Baseline PROM data

Demographic and baseline PROM data are presented in Table 1 and follow-up PROM data in Table 2. The LSS group had the lowest mean score in ODI, 43 (± 16). The DDD group had the highest score in VAS_{BACK} , 67 (± 22). The LDH group had the highest VAS_{LEG} score, 67 (± 26), as well as the lowest quality of life as measured by the EQ-5D, 0.26 (± 0.34).

Mean differences between baseline and FU1 and between FU1 and FU2

The mean differences between baseline and FU1 and between FU1 and FU2 are visualized in Fig. 2. In the LDH group, the mean differences between baseline and FU1 were $-29(\pm 21)$ for ODI, $-24(\pm 32)$ for VAS_{BACK} , $-46(\pm 35)$ for VAS_{LEG} and $0.47(\pm 0.40)$ for EQ-5D. In the LSS group, the corresponding numbers were $-17(\pm 18)$ for ODI, $-25(\pm 32)$ for VAS_{BACK} , $-31(\pm 35)$ for VAS_{LEG} and $0.29(\pm 0.35)$ for EQ-5D, and in the DDD group, $-20(\pm 18)$ for ODI, $-36(\pm 31)$ for VAS_{BACK} , $-23(\pm 30)$ for VAS_{LEG} and $0.32(\pm 0.37)$ for EQ-5D. The mean differences between FU1 and FU2 (Table 2) were negligible in comparison with the changes seen between baseline and FU1.

Table 1 Baseline data

	Diagnostic group		
	LDH N=31,314	LSS N=53,043	DDD N=14,375
Age mean (SD)	45 (14)	67 (11)	47 (13)
Female %	45	54	53
Smoker %	19	12	14
Previous spine surgery %	13	20	26
Unemployed %	11	10	13
Employed %	81	37	75
ODI	48 (18)	43 (16)	44 (15)
VAS_{BACK}	48 (40)	57 (53)	67 (22)
VAS_{LEG}	67 (26)	64 (26)	47 (30)
EQ-5D	0.26 (0.34)	0.35 (0.32)	0.33 (0.32)
Back pain > 1 year, %	36	75	89
Leg pain > 1 year, %	29	69	70

Pre-operative data of three populations operated for degenerative lumbar spine disorders. *LDH* Lumbar Disc Herniation, *LSS* Lumbar Spinal Stenosis, *DDD* Degenerative Disc Disorder, *VASBACK/LEG* Visual Analogue Scale for back and leg pain, *ODI* Oswestry Disability Index, *EQ-5D* EuroQol-5-Dimensions

In Fig. 3a–d, score changes on the individual level are illustrated by histograms. These figures highlight the large spread in score change occurring up to one year post-op in contrast to the harmonious and narrow bell-shaped histograms depicting the changes between FU1 and FU2, where the median equals '0'.

Proportions reaching MIC and thresholds of treatment success at FU1 and FU2

To detect clinically relevant differences in outcome between FU1 and FU2, the MIC—corresponding to a considerable improvement—was computed for each PROM and diagnostic group (Table 3). The difference in proportion reaching MIC at FU2 as compared to FU1 in any of the PROMs in the LDH and the DDD groups was at the most 2% smaller. In the LSS group, 2–4% fewer patients reached MIC at FU2 than at FU1. The percentage reaching thresholds of treatment success at the two follow-up occasions (Table 4) displayed a similar pattern. The largest deterioration (5%) was seen in VAS_{BACK} in the LSS group.

$GA_{BACK/LEG}$ and Satisfaction

Test of two paired proportions of Satisfaction and $GA_{BACK/LEG}$

As shown in Table 5, 85–89% of the patients, depending on the diagnosis, made the same self-assessment at FU1 as at

Table 2 Mean values at the two follow-up occasions and their mean differences in patient-reported outcome measures stratified by three different diagnostic groups. Paired data

PROM	<i>N</i>	1-Year follow-up	2-Year follow-up	Mean diff	SD
LDH					
ODI	8359	18.7 ± 17	18.8 ± 18	−0.15	10
VASBACK	9193	23.7 ± 25	25.0 ± 26	−1.31	21
VASLEG	9778	21.0 ± 26	22.6 ± 28	−1.60	22
EQ-5D	9214	0.73 ± 0.3	0.72 ± 0.3	0.004	0.2
LSS					
ODI	17,549	26.0 ± 19	27.4 ± 20	−1.4	12
VASBACK	16,038	32.4 ± 29	35.5 ± 30	−2.8	23
VASLEG	16,083	32.9 ± 31	35.5 ± 31	−2.6	26
EQ-5D	19,252	0.64 ± 0.3	0.62 ± 0.3	0.02	0.26
DDD					
ODI	5493	24.1 ± 19	24.1 ± 20	0.05	11
VASBACK	5510	30.5 ± 28	31.8 ± 29	−1.3	21
VASLEG	4081	23.9 ± 28	25.9 ± 29	−2.0	23
EQ-5D	5930	0.65 ± 0.3	0.65 ± 0.3	0.005	0.2

LDH Lumbar Disc Herniation, *LSS* Lumbar Spinal Stenosis, *DDD* Degenerative Disc Disorder, *SD* Standard Deviation of mean difference, *VASBACK/LEG* Visual Analogue Scale for back and leg pain, *ODI* Oswestry Disability Index, *EQ-5D* EuroQol-5-Dimensions. The range of the *VASBACK/LEG* and *ODI* scales is from 0–100 and the *EQ-5D* from −0.59 to 1.00

FU2. In the LDH group, 6% switched from being satisfied ('success') at FU1 to being uncertain or unsatisfied ('not success') at FU2 and 5.5% shifted from being uncertain or unsatisfied to being satisfied. The corresponding numbers in the LSS and DDD groups were somewhat higher but showed a similar pattern. Despite the relatively small differences, a statistical significance was reached in the latter groups implying a deterioration in outcome between FU1 and FU2.

Regarding patient self-assessment of changes in back pain and leg pain (Table 6), 82–85% of patients displayed the same outcome at the two follow-up occasions. Apart from responses on GA_{BACK} in the DDD group, there was a statistically significant deterioration in outcome between FU1 and FU2, even though the difference between patients shifting from a better outcome to a worse outcome was no more than 2%.

Discussion

This study confirms that potential improvements of clinical importance occur during the first year after lumbar surgery, irrespective of diagnosis and which PROM is being used. The mean differences between FU1 and FU2 in ODI, $VAS_{BACK/LEG}$ and EQ-5D as well as the proportions reaching MIC indicated a minor deterioration in outcome between the first year and second year of follow-up, in all of the three diagnosis groups. However, a deterioration of this size could be expected in populations with degenerative conditions. Our data confirm the results by Adogwa et al. [2] who concluded that ODI and $VAS_{BACK/LEG}$ obtained at 12 months

adequately predict the outcomes at 24 months in patients operated with lumbar nerve root decompression and fusion. Glassman et al. [13] found no differences between FU1 and FU2 in the ODI and the $NRS_{BACK/LEG}$ in patients with adult spinal deformity. In a recent study, Kim et al. concluded that 1-year outcomes reliably can predict 2-year outcomes for discectomy, but it was not clear for laminectomy or fusion procedures [14]. The discordant results may be caused by the use of different methods compared to our study. In the report by Kim et al., the conclusion was based on a model where a change in ODI score of just one point could turn a meaningful outcome into a non-meaningful outcome, which was also pointed out by the authors. Small differences between two timepoints in a change score are likely to be seen and are possibly caused by normal fluctuations.

The proportion shifting from 'success' on GA_{LEG} and GA_{BACK} on FU1 to 'not success' on FU2 was 8.5% and 8% in the LDH group, 11% and 10% in the LSS group and 10% and 8% in the DDD group. Some of these individuals may represent well-known causes for reoperation such as recurring disc herniation, incomplete decompression, adjacent level stenosis and pseudarthrosis. The movements from 'success' to 'not success' and vice versa between the two follow-ups may also be manifestations of normal symptom fluctuations, recall bias or response shift and other measurement errors in PROMs.

Also, worth noticing is that the 'not success' allotment does not entirely consist of patients experiencing a deterioration. Included in this category are also individuals responding as 'somewhat better' and 'unchanged'. Indeed, when investigating the number of patients shifting from

Fig. 2 Line charts of pre-operative and post-operative mean values of the VAS back and leg pain, ODI and EQ-5D index highlighting a large change during the first post-operative year and only small variations thereafter. The LDH group corresponds to the continuous lines, the LSS group to the dotted lines, and the DDD group to the broken lines. The error bars are 95% CI. LDH Lumbar Disc Herniation, LSS Lumbar Spinal Stenosis, DDD Degenerative Disc Disorders, VAS_{BACK/LEG} Visual Analogue Scale for back and leg pain, ODI Oswestry Disability Index, EQ-5D EuroQol-5-Dimensions index score

‘pain free’ or ‘much better’ to ‘worse’, it turns out that no more than 2.5% of the whole study population shift from ‘success’ at FU1 to ‘worse’ at FU2 on GA_{LEG}, the corresponding percentage for GA_{BACK} being 1.8%. Although there was a statistically significant deterioration in outcome between FU1 and FU2, the relevance in clinical practice can certainly be questioned.

These findings suggest that a follow-up at both one and two years post-surgery in effectiveness studies is unnecessary.

The difficulties in reaching a consensus regarding the definition of a minimal important change relevant to patients (MIC) and the most appropriate method to compute it [15, 16] have led to the search for new strategies to define a clinically relevant outcome. Such an alternative is the definition of cut points of treatment success based on absolute scores as opposed to change in scores [17]. Such a threshold is the value above (or below depending on the scale of the PROM) which a patient considers the magnitude of his or her encumbrance to be acceptable following the surgery. A final score may be less affected by response shift and recall bias than a change score would, although that possible advantage is yet to be proven. Thresholds of success based on absolute post-operative scores at FU1 and FU2, respectively, were estimated as previously described by Tubach et al. [18] and recently by van Hooff et al. in a lumbar pain population [17]. No relevant differences could be found.

The number of patients reaching the thresholds of success was rather low in the LSS group—approximately 47–60% depending on the PROM. The cause might be the stringent definition of a successful outcome, leaving those that responded as ‘somewhat better’ and ‘unchanged’ as unsuccessful. Other contributing factors to this finding may be that (i) the LSS population is relatively older, with a larger prevalence of comorbidity and probably other painful degenerative conditions such as hip arthritis, that may confound the outcome of the surgery if measured by PROMs, (ii) the indication for surgery is too wide and (iii) the degeneration of the spine is often a multisegmental process that opens for future pain and disability. In this study, we chose not to make a distinction between LSS patients undergoing decompression surgery only and those treated with decompression and fusion surgery. The decision was based on the studies by

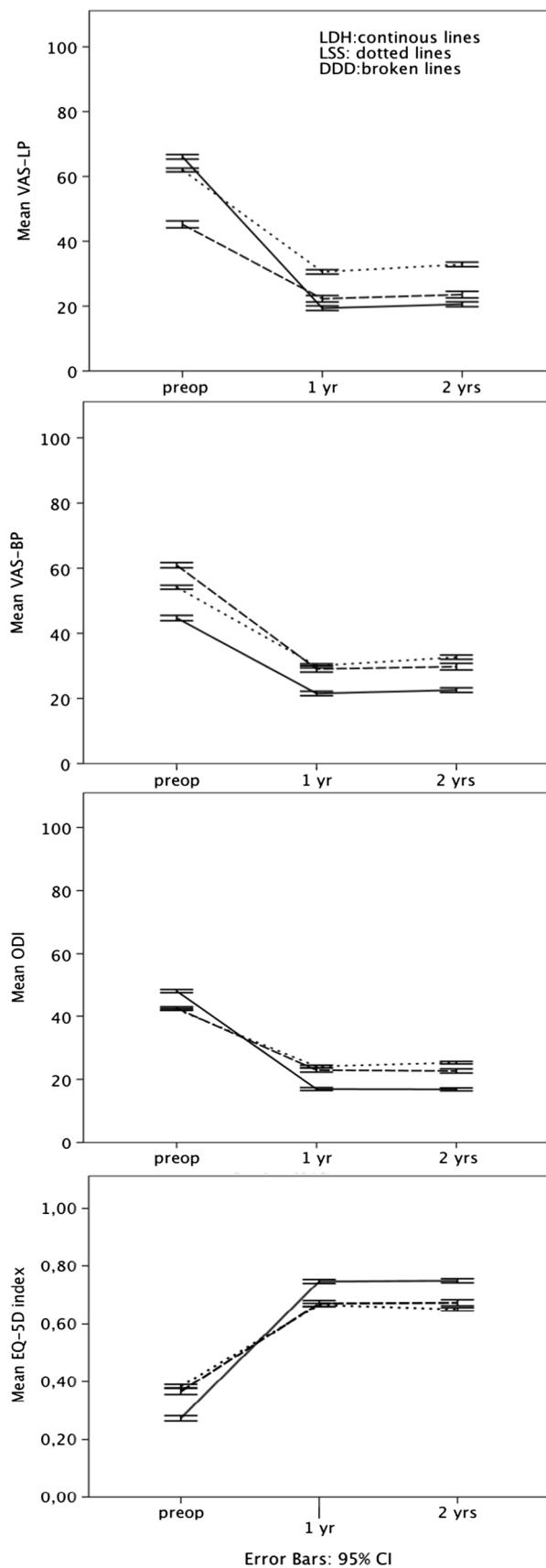


Table 3 MIC for treatment success based on PROM score changes between baseline and FU1 and FU2, respectively. The analyses are based on paired data

PROM	Diagnostic group	N	MIC for success	Reaching MIC at FU1 (%)	Reaching MIC at FU2 (%)
ODI	LDH	8359	−22	64	64
	LSS	17,549	−14	57	54
	DDD	5493	−16	58	59
VASBACK	LDH	9193	−20	53	51
	LSS	16,038	−28	45	41
	DDD	5510	−29	55	52
VASLEG	LDH	9778	−39	62	60
	LSS	16,083	−27	54	51
	DDD	4081	−23	50	49
EQ-5D	LDH	9214	0.18	71	69
	LSS	19,252	0.10	61	59
	DDD	5930	0.10	66	66

ROC curve analyses. GALEG or GABACK was used as criterion standard depending on which of the two have the best correlation with each PROM. The MIC corresponds to the minimal change in score needed for patients to perceive their improvement as ‘completely pain free’ or ‘much better’ at 1 year post-surgery as compared to before surgery. Score changes not reaching MIC correspond to a self-assessment of ‘somewhat better’, ‘unchanged’ or ‘worse’. *LDH* Lumbar Disc Herniation, *LSS* Lumbar Spinal Stenosis, *DDD* Degenerative Disc Disorder, *VASBACK/LEG* Visual Analogue Scale for back and leg pain, *ODI* Oswestry Disability Index, *EQ-5D* EuroQol-5-Dimensions

Försth et al., who concluded that the two groups had similar clinical outcomes at 2 years [19, 20].

There appears to be no need for a 2-year follow-up of PROMs in this population and resources could be better spent on increasing the response rates at one year for registries, thereby improving the quality of the data and reducing the risks of assessment bias. Efforts could also be made to define certain groups of patients or procedures where a longer follow-up period such as 2 years may be needed—or shorter. The value of PROM assessment based on Swespine data collected at 5 and 10 years post-surgery is unclear and is yet to be studied. The more time has passed since the operation, the higher the risk of other health-related events aggravating the interpretation of the PROMs. More ‘objective’ endpoints, such as reoperation rates, cause of reoperation, time from index surgery to a new surgical event or in relation to implant survival, might be of greater importance than PROMs in the very long-term assessments.

Limitations

Although the strength of this study is the large real-life database, a selection bias may be present because of the proportion of non-respondents. In this study, the response rates were 75% at FU1 and 65% at FU2 and adjustments for risk factors associated with responding were not made. Solberg and colleagues concluded, however, that there were no differences in outcome between respondents and non-respondents in a population with degenerative lumbar

disorders retrieved from the Norwegian spine registry and that the non-respondents could be treated as missing at random [21]. The same conclusion was later drawn in a similar study based on data from the Danish spine registry [22]. The high similarity between these countries should make the results applicable to the current study. The EQ-5D index is a measure designed for cost-effectiveness analyses and not for a similarity study such as this one. Therefore, the interpretation of the results involving EQ-5D should be made with caution. As in all studies where means are calculated and compared—longitudinally or cross-sectionally—there is a risk that patients worsening and improving neutralize each other to some extent. As illustrated in Fig. 3, this risk appears to be low. This study investigated whether there are any clinically important differences in PROMs between 1 and 2 years of follow-up or not. However, possible differences in outcome between 1 and 5 or 10 years are yet to be shown.

Conclusion

All PROMs used in Swespine in follow-up of degenerative lumbar spine surgery demonstrate minor differences between one and two years that can be expected in populations symptomatically treated for degenerative conditions. This suggests that when a PROM is the outcome variable, a one-year follow-up is sufficient in clinical effectiveness studies.

Fig. 3 Histograms visualizing the distribution of changes in scores between baseline and FU1 and between FU1 and FU2 for ODI, VAS back and leg pain and EQ-5D in the three diagnostic groups: LDH, LSS and DDD. The number of patients is presented in Tables 3 and 4. *LDH* Lumbar Disc Herniation, *LSS* Lumbar Spinal Stenosis, *DDD* Degenerative Disc Disorders, *VAS_{BACK/LEG}* Visual Analogue Scale for back and leg pain, *ODI* Oswestry Disability Index, *EQ-5D* EuroQol-5-Dimensions index score

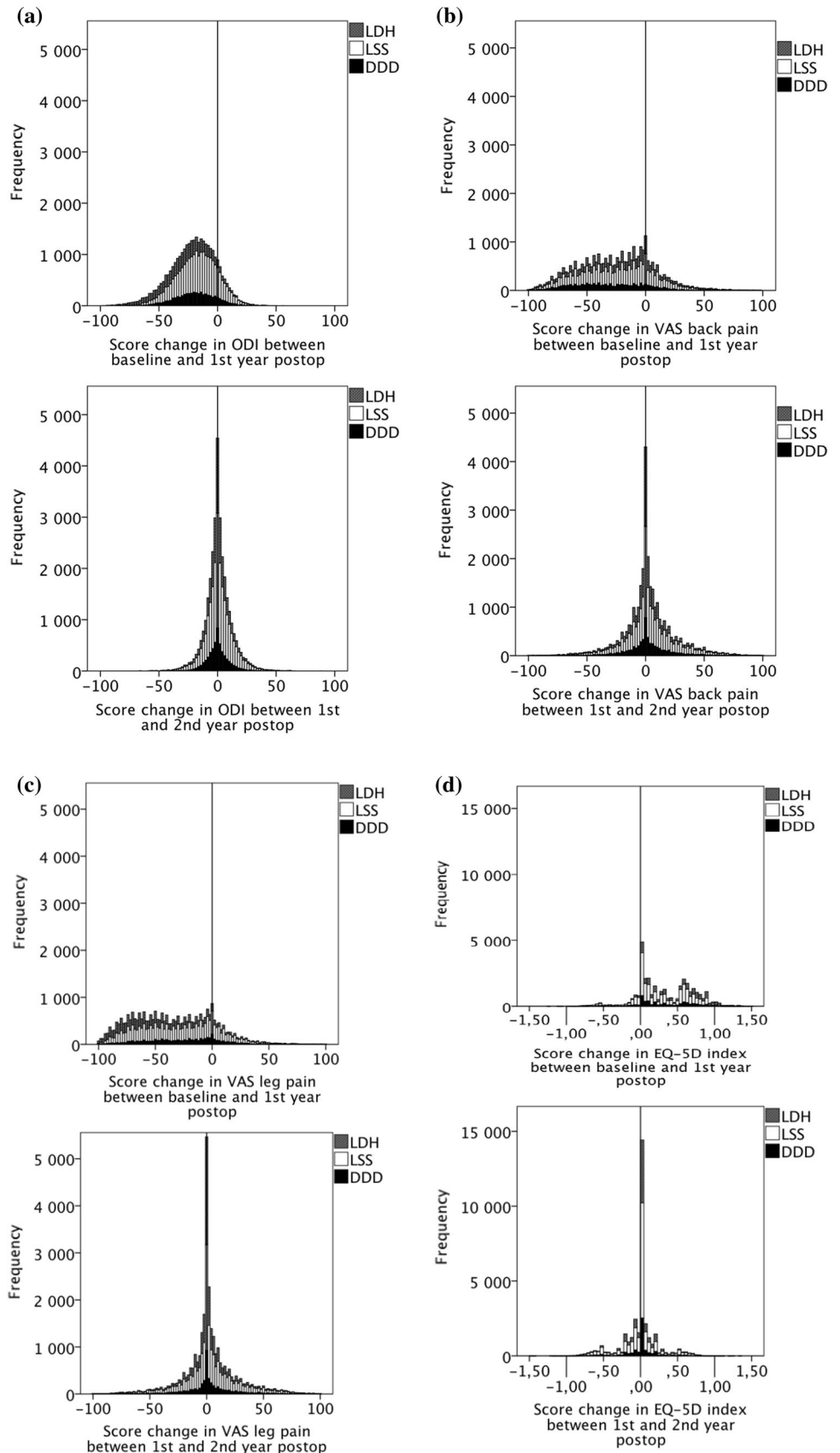


Table 4 Thresholds of treatment success based on absolute PROM scores at one and two years of follow-up. Paired data

PROM	Diagnostic group	N	Threshold of treatment success	Reaching threshold at FU1 (%)	Reaching threshold at FU2 (%)
ODI	LDH	8359	<23	67	67
	LSS	17,549	<27	53	52
	DDD	5493	<25	58	59
VASBACK	LDH	9193	<29	66	65
	LSS	16,038	<32	55	50
	DDD	5510	<33	61	59
VASLEG	LDH	9778	<27	70	70
	LSS	16,083	<31	55	51
	DDD	4081	<21	67	64
EQ-5D	LDH	9214	>0.70	71	71
	LSS	19,252	>0.68	63	61
	DDD	5930	>0.68	70	69

ROC curve analyses. GALEG or GABACK was used as criterion standard depending on which of the two have the best correlation with each PROM. Values below (ODI, VAS BACK/LEG pain) or above (EQ-5D) the absolute threshold correspond to a patient’s self-assessment as ‘completely pain free’ or ‘much better’ at 1 year post-surgery, and values above (ODI, VAS BACK/LEG pain) or below (EQ-5D) correspond to a self-assessment as ‘somewhat better’, ‘unchanged’ or ‘worse’. LDH Lumbar Disc Herniation, LSS Lumbar Spinal Stenosis, DDD Degenerative Disc Disorder, VASBACK/LEG Visual Analogue Scale for back and leg pain, ODI Oswestry Disability Index, EQ-5D EuroQol-5-Dimensions

Table 5 Differences in patient responses on Satisfaction between FU1 and FU2

Diagnostic group	Success at FU2			p value	Comment	
	Yes	No	Total			
PROM: Satisfaction						
LDH	Success at FU1					
	Yes	10,183	810	10,993	0.422	FU1 = FU2: 89% FU1 = Yes and FU2 = No: 6% FU1 = No and FU2 = Yes: 5.5%
	No	777	2243	3020		
Total	10,960	3053	14,013			
LSS	Success at FU1					
	Yes	14,723	2222	14,723	0.000	FU1 = FU2: 85% FU1 = Yes and FU2 = No: 8.5% FU1 = No and FU2 = Yes: 6.5%
	No	1675	7187	8862		
Total	16,398	9409	25,807			
DDD	Success at FU1					
	Yes	4619	498	5117	0.003	FU1 = FU2: 87% FU1 = Yes and FU2 = No: 7% FU1 = No and FU2 = Yes: 6%
	No	407	1467	1874		
Total	5026	1965	6991			

Success = Yes: patients that self-assessed as ‘satisfied’. No: patients that self-assessed as ‘uncertain’ or ‘dissatisfied’. Test statistic: McNemar at 95% confidence level. FU1 = FU2 (exact agreement): sum of respondents giving the same response at both FU1 and FU2 divided by the total sum of respondents

Table 6 Differences in patient responses on $GA_{BACK/LEG}$ between FU1 and FU2

Diagnostic group		Success at FU2			p value	Comment
		Yes	No	Total		
PROM: GA_{LEG}						
LDH	Success at FU1					
	Yes	9427	1183	10,610		FU1 = FU2: 85%
	No	882	2431	3313		FU1 = Yes and FU2 = No: 8.5%
	Total	10,309	3614	13,923	0.000	FU1 = No and FU2 = Yes: 6.5%
LSS	Success at FU1					
	Yes	11,100	2778	13,878		FU1 = FU2: 81%
	No	1909	8150	10,059		FU1 = Yes and FU2 = No: 11%
	Total	13,009	10,928	23,937	0.000	FU1 = No and FU2 = Yes: 8%
DDD	Success at FU1					
	Yes	3288	602	3890		FU1 = FU2: 82%
	No	482	1719	2201		FU1 = Yes and FU2 = No: 10%
	Total	3770	2321	6091	0.000	FU1 = No and FU2 = Yes: 8%
PROM: GA_{BACK}						
LDH	Success at FU1					
	Yes	8556	1045	9601		FU1 = FU2: 85%
	No	933	2594	3527		FU1 = Yes and FU2 = No: 8%
	Total	9489	3639	13,128	0.013	FU1 = No and FU2 = Yes: 7%
LSS	Success at FU1					
	Yes	11,345	2478	13,823		FU1 = FU2: 82%
	No	1822	8302	10,124		FU1 = Yes and FU2 = No: 10%
	Total	13,167	10,780	23,947	0.000	FU1 = No and FU2 = Yes: 8%
DDD	Success at FU1					
	Yes	3941	528	4469		FU1 = FU2: 85%
	No	493	1980	2473		FU1 = Yes and FU2 = No: 8%
	Total	4434	2508	6942	0.287	FU1 = No and FU2 = Yes: 7%

Success = Yes: patients that self-assessed as ‘pain free’ or ‘much better’. No = patients that self-assessed as ‘somewhat better’, ‘unchanged’ or ‘worse’. Test statistic: McNemar on the 95% confidence level. FU1 = FU2 (exact agreement): the sum of respondents giving the same response at both FU1 and FU2 divided by the total sum of respondents

Acknowledgements The authors would like to acknowledge The Neubergh foundation, GHP Spine Center Göteborg and the Regional Agreement on Medical Training and Clinical Research (ALF) for their financial support.

Compliance with ethical standards

Conflict of interest None of the authors has any potential conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.


References

1. Fekete T, Loibl M, Jeszenszky D, Haschtmann D, Banczerowski P, Kleinstück F, Becker H, Porchet F, Mannion A (2018) How does patient-rated outcome change over time following the surgical treatment of degenerative disorders of the thoracolumbar spine? *Eur Spine J* 27(3):700–708. <https://doi.org/10.1007/s00586-017-5358-2>
2. Adogwa O, Elsamadicy AA, Han JL, Cheng J, Karikari I, Bagley CA (2016) Do measures of surgical effectiveness at 1 year after lumbar spine surgery accurately predict 2-year outcomes? *J Neurosurg Spine* 25(6):689–696. <https://doi.org/10.3171/2015.8.Spine.15476>
3. <Report_2014_Swespine_Engl_ver_141204.pdf>
4. Stromqvist B, Fritzell P, Hagg O, Jonsson B (2009) The Swedish Spine Register: development, design and utility. *Eur Spine J* 18(Suppl 3):294–304. <https://doi.org/10.1007/s00586-009-1043-4>
5. Parai C, Hagg O, Lind B, Brisby H (2018) The value of patient global assessment in lumbar spine surgery: an evaluation based

- on more than 90,000 patients. *Eur Spine J* 27(3):554–563. <https://doi.org/10.1007/s00586-017-5331-0>
6. Zanolì G, Stromqvist B, Jonsson B (2001) Visual analog scales for interpretation of back and leg pain intensity in patients operated for degenerative lumbar spine disorders. *Spine* 26(21):2375–2380
 7. Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry low back pain disability questionnaire. *Physiotherapy* 66(8):271–273
 8. EuroQol (1990) A new facility for the measurement of health-related quality of life. *Health Policy* 16(3):199–208
 9. Hagg O, Fritzell P, Oden A, Nordwall A (2002) Simplifying outcome measurement: evaluation of instruments for measuring outcome after fusion surgery for chronic low back pain. *Spine* 27(11):1213–1222
 10. Vet HCWd (2011) Measurement in medicine: a practical guide. Practical guides to biostatistics and epidemiology. Cambridge University Press, Cambridge
 11. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 19(4):539–549. <https://doi.org/10.1007/s11136-010-9606-8>
 12. Altman DG (1991) Practical statistics for medical research. Chapman and Hall, London
 13. Glassman SD, Schwab F, Bridwell KH, Shaffrey C, Horton W, Hu S (2009) Do 1-year outcomes predict 2-year outcomes for adult deformity surgery? *Spine J* 9(4):317–322. <https://doi.org/10.1016/j.spinee.2008.06.450>
 14. Kim EJ, Chotai S, Archer KR, Bydon M, Asher AL, Devin CJ (2017) Need for two-year patient-reported outcomes score for lumbar spine surgery is procedure-specific: analysis from a prospective longitudinal spine registry. *Spine* 42(17):1331–1338. <https://doi.org/10.1097/brs.0000000000002087>
 15. King MT (2011) A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 11(2):171–184. <https://doi.org/10.1586/erp.11.9>
 16. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, Croft P, de Vet HC (2010) Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 63(5):524–534. <https://doi.org/10.1016/j.jclinepi.2009.08.010>
 17. van Hooff ML, Mannion AF, Staub LP, Ostelo RW, Fairbank JC (2016) Determination of the Oswestry Disability Index score equivalent to a "satisfactory symptom state" in patients undergoing surgery for degenerative disorders of the lumbar spine—a Spine Tango registry-based study. *Spine J* 16(10):1221–1230. <https://doi.org/10.1016/j.spinee.2016.06.010>
 18. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, Bombardier C, Felson D, Hochberg M, van der Heijde D, Dougados M (2005) Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheumatic Dis* 64(1):34–37. <https://doi.org/10.1136/ard.2004.023028>
 19. Forsth P, Michaelsson K, Sanden B (2013) Does fusion improve the outcome after decompressive surgery for lumbar spinal stenosis? A two-year follow-up study involving 5390 patients. *Bone Joint J* 95-b(7):960–965. <https://doi.org/10.1302/0301-620x.95b7.30776>
 20. Forsth P, Olafsson G, Carlsson T, Frost A, Borgstrom F, Fritzell P, Ohagen P, Michaelsson K, Sanden B (2016) A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. *N Engl J Med* 374(15):1413–1423. <https://doi.org/10.1056/NEJMoal1513721>
 21. Solberg TK, Sorlie A, Sjaavik K, Nygaard OP, Ingebrigtsen T (2011) Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? *Acta Orthop* 82(1):56–63. <https://doi.org/10.3109/17453674.2010.548024>
 22. Hojmark K, Stotttrup C, Carreon L, Andersen MO (2016) Patient-reported outcome measures unbiased by loss of follow-up. Single-center study based on DaneSpine, the Danish spine surgery registry. *Eur Spine J* 25(1):282–286. <https://doi.org/10.1007/s00586-015-4127-3>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

C. Parai^{1,2}  · O. Hägg² · B. Lind^{1,2} · H. Brisby^{1,3}

¹ Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

² Spine Center Göteborg, Gothenburg, Sweden

³ Department of Orthopaedics, Sahlgrenska University Hospital, Gothenburg, Sweden