Title Page

Can 3-month models or observed 3- or 6-month patient-reported outcome measures accurately predict 12-month outcomes after lumbar decompressive surgery?

Corresponding & Lead Author
Complete Name: *Jonathan H. Geere MSc
Address: Spinal Physiotherapist, Physiotherapy Department, Spire Norwich Hospital, Old Watton Road, Colney, Norwich, NR4 7TD, UK
Telephone number: UK +44 (0)1603 255587
Email address: jonathan.geere@nhs.net
ORCID: 0000-0003-2069-6845

Other Authors
Complete Name: Paul R. Hunter M.D. Prof., Norwich Medical School, University of East Anglia, Norwich, U.K.
Email: paul.hunter@uea.ac.uk

Complete Name: Andrew J. Cook FRCS [Tr&Orth], Consultant Spinal Surgeon, Norfolk & Norwich University Hospital, Norwich, U.K.
Email: Andrew.Cook14@nhs.net

Complete Name: Amarjit S. Rai FRCS [Tr&Orth], Consultant Spinal Surgeon, Spire Norwich Hospital, Norwich, U.K.
Email: am.rai@icloud.com

Acknowledgements

We thank Girish G. N. Swamy, Lennel L. N. Lutchman, and Nick Steele for data contribution. We thank present and past physiotherapy and administration staff for data collection and collation. We thank present and past Spire Norwich Hospital physiotherapy and medical record staff in assistance with data collection.

Abstract

Introduction: Most lumbar decompressive surgery patients experience symptom improvement to 3 months, after which about 13% experience a clinically relevant deterioration. Patient-reported outcome measures (PROMs) are accepted as indicators of clinical outcome, but the earliest timepoint when PROMs stabilise is unknown and few studies have assessed risk factors for symptom deterioration. This study aimed to identify risk factors for symptom deterioration and identify whether 3-month models or observed 3- or 6-month outcomes accurately predict 12-month outcomes.

Methods: The development cohort included 1096 prospectively collected discectomy or laminectomy cases with or without single-level fusion from 01/01/2008–31/12/2020 at a single centre. Three-month models were developed using baseline clinical variables and 3-month PROM. The primary 12-month outcome was the Oswestry Disability Index (ODI), and secondary outcomes were back and leg pain. Validation was on 364 cases from 01/01/2021–31/12/2022. Predictive accuracy was evaluated by interclass correlation coefficient (ICC) and by area-under-the-curve (AUC) to classify to a minimal clinically important difference (MCID). MCID concordance rates for observed 3-month and 6-month with 12-month PROM were calculated.

Results: Three-month predictors of 12-month PROM were condition duration, smoking, diabetes, rheumatic disorder, lower limb arthroplasty, mobility aided, female, scoliosis, underweight (BMI <18.5kg/m^2), 3-month PROM, and 3-month unemployment. ODI model and observed 3-month ODI had equivalent ICC and AUC values. Observed 3-month ODI ICC was 0.71 [95% confidence intervals (CI) 0.68–0.74] and AUC was 0.83 [95%CI 0.80–0.86]. Observed 6-month ODI ICC was 0.82 [95%CI 0.79–0.85] and AUC was 0.92 [95%CI 0.89–0.95]. MCID concordance for 3-month ODI was 84% and 6-month ODI was 91%.

Conclusion: Symptom deterioration after 3 months is linked to nine baseline factors. This is the first study to demonstrate that 6-month PROM accurately predict individual patient 12-month PROM after discectomy or laminectomy. Fusion surgery requires a minimum 12-month PROM follow-up.

Keywords: lumbar surgery, discectomy, stenosis, outcomes, spine register

Statements and Declarations

Authors' contribution statements: Conceptualization was performed by Jonathan H. Geere and Amarjit S. Rai; Data curation was performed by Jonathan H. Geere, Andrew J. Cook and Amarjit S. Rai; Methodology was performed by Jonathan H. Geere and Paul R. Hunter; Formal analysis and investigation was performed by Jonathan H. Geere and Paul R. Hunter; Writing - original draft preparation was performed by Jonathan H. Geere; Writing - review and editing was performed by Jonathan H. Geere, Paul R. Hunter, Andrew J. Cook, and Amarjit S. Rai; Supervision was performed by Jonathan H. Geere and Amarjit S. Rai. All authors read and approved the final manuscript.

Source of funding: No funds, grants, or other support was received

Competing interests: The authors have no competing financial or non-financial interests to declare.

Ethics approval: The study was approved by the Spire Norwich Hospital Ethics Committee and the Norfolk and Norwich University Hospital NHS Foundation Trust Audit and Governance Committee (ref: 43119).

Introduction

Pain and disability patient-reported outcome measures (PROMs) are used in national spinal registries to demonstrate value for money, monitor surgeon and hospital practice, and measure individual patient outcome [1, 2]. One- or 2-year PROM follow-up is recommended after lumbar decompressive surgery to allow outcomes to reach stability at the population and individual patient level [3, 4]. However, earlier PROM follow-up reduces administrative costs, limits attrition bias, and enables prompt appraisal of clinical practice and individual patient outcome [5-7]. The highly correlated 6- to 12-month disability scores reported in one small study suggests that outcomes reach a stable state earlier than previously considered [8].

At the population level, postoperatively the mean disability score improves until 3 months, after which the change in score is relatively small [9, 10]. However, it is unclear when individual patients reach the most stable state after surgery. Outcome improvement relevant to individual patients is frequently measured by the achievement of a minimal clinically important difference (MCID) in a PROM [11], with stability measured as the concordance in MCID status (achieved or not achieved) between timepoints. For individual patient disability, concordance in the MCID between 3 and 12 months shows slightly lower stability than that between 1 and 2 years (77–78% versus 77–87%, respectively) [3-6, 12].

Although most patients achieve improvement in disability by 3 months, 12–14% subsequently lose clinically relevant improvement by 12 months [4, 12]. However, study on clinical risk factors for symptom deterioration after 3 months is limited [13]. Identification of risk factors for recurrent symptoms will help focus on perioperative rehabilitation strategies for at-risk populations. In addition, risk factors inputted into 3-month PROM predictive models could provide the earliest accurate evaluation of outcomes.

This study aimed to identify 3-month risk factors for 12-month PROM after lumbar decompressive surgery and build 3-month clinical predictive models. The secondary aim is to identify whether 3-month models or observed 3- or 6-month PROMs accurately predict 12-month PROMs.

Materials and methods

The study observed the guidelines for transparent reporting of a multivariable prediction model for individual prognosis or diagnosis [14]. We analysed prospectively collected clinical data from consecutive patients who underwent elective lumbar decompressive surgery for lumbar disc herniation or spinal stenosis, some with single-level arthrodesis, under the care of five consultants at a single centre between 01/01/2008–31/12/2022. Baseline variables, PROM, and written informed consent were collected preoperatively. PROMs were collected at 3 months after outpatient physiotherapy or via postal questionnaire (PQ), and at 6 and 12 months via a secure web-based portal (SWBP) or PQ (6-month PROM principally via SWBP since 2014).

The primary outcome was the Oswestry Disability Index 2.1a (ODI) and secondary outcomes were back and leg pain. The ODI is a 10-item disability questionnaire with a 0–100% (least–worst) scale range [15]. Average pain over a week was measured by the visual analogue scale (VAS) with a 0–10 (none–worst imaginable) range. MCID, the smallest change in outcome to be clinically meaningful, was defined as a ≤30% reduction in preoperative PROM score [11]. Patients with zero baseline disability or pain were excluded from the respective MCID calculation. A secondary definition of individual patient level outcome used the patient acceptable symptom state (PASS), where PASS describes whether the current level of symptoms is acceptable and is defined as ≤22 for the ODI and ≤3 for pain [16, 17].

Ethics

Approval was obtained from the institution and the local NHS authority review boards (ref: 43119).

Inclusion and exclusion criteria

Eligibility criteria was 1–2 level decompressive surgery, with or without single-level spinal fusion. Exclusion criteria included nonconsent, reoperation within 1-year, spinal infection, cauda equina syndrome, extant tumour, major psychological disorder, or overlapping spinal myelopathy.

Data handling and statistical analysis

Cases with missing baseline data, 3-month PROM or 12-month ODI were removed. A subgroup of all-timepoint responders with 3-, 6- and 12-month ODI was created. Analysis of group categorical data was by Chi-squared, cross-sectional continuous data by t-test, and longitudinal continuous data by paired t-test. The development cohort included 1096 collected between 01/01/2008–31/12/2020. Models were developed via general linear regression inputted with 16 baseline variables, 3-month PROM, and 3-month work status (Supplementary Fig. i), with outputs of variable beta coefficient (ß) and *p*-value, and R^2. The validation cohort included 364 collected between 01/01/2021–31/12/2022.

Five methods were used to evaluate predictive performance. Calibration was primarily measured by score agreement via the intraclass correlation coefficient absolute agreement two-way random effects single measures model, (ICC) as it measures association and absolute differences in scores [18]. An ICC value of ≥ 0.70 is recommended for PROM used as a group measure [18]. Calibration was secondarily measured by change score and, exclusive to model validation, R^2 [18]. Discrimination was primarily measured by classification according to MCID status and secondarily measured by classification according to PASS achievement and calculated by area-under-the-curve (AUC). AUC values range from 0.5–1.0 (no discrimination–perfect accuracy), with a rule of thumb for intermediate values of <0.7 as poor, 0.7–<0.8 as acceptable, 0.8–<0.9 as excellent, and ≥ 0.9 as outstanding measures of separability [19]. In addition, concordance rates for MCID and PASS, defined as no change in the respective MCID/PASS status between timepoints, were obtained.

The predictive performances of 3-month models and observed 3-month PROM were compared in the validation cohort. Observed 3- and 6-month PROM predictive performances and concordance rates were calculated in the development cohort, with the reproducibility of results checked in the validation cohort. In the validation cohort, the probable cause of loss in 3-month MCID or PASS achievement for ODI was obtained from patient notes and unsolicited patient comments on the 12-month postal questionnaire.

Statistical significance was defined by $p \leq 0.05$, based on a two-sided hypothesis test with no adjustment for multiple comparisons. No sample size analysis was performed. Statistical analysis was performed with SPSS version 28 (SPSS Inc., Chicago, IL, USA).

Results

Development cohort

The development cohort flowchart is shown in Supplementary Fig. ii. Baseline characteristics, baseline PROM, and 3- and 12-month PROM by procedure is shown in Table 1. The overall baseline to 3-month PROM scores significantly improved (∆ODI 27.0; ∆back pain 3.63; ∆leg pain 4.76, all *p*-values < 0.001).

Three-month model predictors

Significant 3-month predictors of 12-month ODI were baseline condition duration, smoking, diabetes, rheumatic disorder, lower limb arthroplasty, mobility aided, female, and underweight (BMI <18.5kg/m^2), and 3-month ODI, back pain, and unemployment status (Table 2). Each 3-month PROM was the primary predictor of its respective 12-month outcome. Condition duration ≥2 years, 3-month ODI, and diabetes were predictors for all PROMs. Lower limb arthroplasty and smoking also predicted back pain. Rheumatic disorder, mobility aided, scoliosis (Cobb angle >30°), and underweight also predicted leg pain.

Model validation

The predictive performance of 3-month PROM models was similar to the predictive performance of observed 3-month PROM: ∆ICC -0.03–0.00; change scores all $p > 0.05$, and the ∆AUC for MCID or PASS -0.01–0.03 (Table 3).

Model R^2 for ODI was 0.60 (95% CI 0.53–0.66), back pain was 0.37 (95% CI 0.29–0.44), and leg pain was 0.40 (95% CI 0.32–0.47).

Observed 3-month outcome predictive performance in the development cohort

Observed 3-month PROM predictive performance and concordance rates are shown in Table 4.

Calibration results show that 3-month ODI had sufficient ICC for group use in discectomy or laminectomy (ICC 0.72) but insufficient ICC in fusion (ICC 0.51), and pain had insufficient ICC for group use in each procedure (ICC 0.46–0.54). ODI or pain showed a small, but statistically significant, 3-to-12-month score change in each procedure.

Three-month ODI had excellent discrimination for MCID in discectomy (AUC 0.89) and laminectomy (AUC 0.80) and acceptable discrimination for fusion (AUC 0.74). Three-month pain had acceptable discrimination in each procedure (AUC 0.73–0.78) except leg pain had poor discrimination in fusion (AUC 0.67).

MCID concordance for 3-month ODI was highest for discectomy (89%) compared to laminectomy or fusion (79–81%). MCID concordance for 3-month pain was highest for discectomy (leg pain 85%) and fusion (leg pain 83%).

Comparison of semi-responder and all-timepoint responder cohorts

In the development cohort, comparison of semi-responders (3-, 12-month) versus all-timepoint responders (3-, 6-, 12-month) showed minimal difference in baseline characteristics, baseline PROM (Supplementary Table i), 3-month predictive performance and 3-to-12-month concordance rates (Supplementary Table ii), indicating that the all-timepoint responder subgroup had minimal selection bias.

Observed 6-month outcome predictive performance in the development cohort

Observed 6-month PROM predictive performance and concordance rates are shown in Table 5.

Calibration results show that 6-month ODI had sufficient ICC for group use in each procedure (ICC 0.74–0.81). In contrast, 6-month pain had insufficient ICC for group use in each procedure (ICC 0.55–0.69), although back pain had sufficient ICC for laminectomy (ICC 0.70). ODI and pain showed no 6-to-12-month score change in discectomy or laminectomy, but ODI and leg pain significantly deteriorated after 6 months in fusion.

Six-month ODI had outstanding discrimination for MCID in discectomy and laminectomy (AUC 0.91–0.94) and excellent discrimination for fusion (AUC 0.88). Six-month pain had excellent discrimination in discectomy and laminectomy (AUC 0.82–0.84) and borderline poor/acceptable discrimination in fusion (AUC 0.69–0.71).

MCID concordance was similar across procedures for 6-month ODI (88–92%), back pain (80–86%), and leg pain (80–86%).

At 3 or 6 months, AUC and concordance values for PASS were not systematically different to the respective MCID value.

Development and validation cohort comparison

Development and validation cohorts had equal baseline PROM scores, but significant differences in baseline characteristics, and 3-, 6- and 12-month PROM scores (Supplementary Table iii), indicating that validation was performed on a different case-mix. The development and validation 3-month model predictive performances were similar, indicating minimal overfitting of data in models (Supplementary Table iv).

Observed PROM predictive performance results in the development cohort were reproduced in the validation cohort. Development and validation observed 3-month PROM predictive performances were similar: ΔICC 0.03–0.10; Δchange score 1–2% of scale; ΔAUC for MCID or PASS -0.08–0.06; and Δconcordance for MCID or PASS 0–4% of absolute value (Supplementary Table v). Development and validation observed 6-month PROM predictive performances were similar: ΔICC 0.04–0.17; Δchange score 0–4% of scale; ΔAUC for MCID or PASS -0.06–0.04; Δconcordance for MCID or PASS 0–9% of absolute value (Supplementary Table vi).

Probable cause of loss in 3-month disability success

In the validation cohort, 65 patients lost 3-month MCID or PASS achievement in the ODI. The probable causes were the lumbar spine (40%, 26/65), cervical spine (5%, 3/65), lower limb joint (14%, 9/65), other comorbidities (15%, 10/65), and undetermined (26%, 17/65).

Discussion

This is the first temporal validation of 3-month predictive models for change in symptoms between 3 and 12 months after lumbar decompressive surgery. Deterioration in outcome after 3 months is commonly attributed to the degenerative spinal condition [20] but we that found other factors contribute to worsening outcome. Consistent with other studies, population-level risk factors for worsening ODI after 3 months were smoking, and diabetes [13, 21], whereas our study found lower limb arthroplasty, condition duration > 2 years, rheumatic disorders, or mobility aided also have notable deleterious effects. Our study confirmed that 3-month models and observed 3-month outcomes are insufficient to predict 12-month outcomes at the individual patient level [4, 12, 13]. However, 6 months after non-fusion surgery outcomes stabilise, and the observed 6-month ODI and pain can be used to accurately predict individual patient 12-month outcome.

Models help predict the direction and average magnitude of change in populations and thus improve the explained variance (R^2) of the 12-month outcome. Model R^2 is ultimately limited by the random variation in PROM score in the period around follow-up, as models cannot predict random variation. In this context, our 3-month ODI model 60% R^2 is close to the limit set by random variation in the ODI score (R^2 67–74%) over a 5-to-6-week measurement period in clinically stable populations [22-24]. Our 3-month ODI model 60% R^2 is greater than the 47–51% reported in another study with fewer significant risk factors and less robust validation by bootstrap resampling [13, 14]. Model performance might, however, still be improved by the inclusion of other postoperative measures including resilience and self-efficacy [25], glycaemic control [26], and access to healthcare.

Our observed 6-month results indicate that outcomes reach the most stable state 6 months after non-fusion surgery. The observed 6-month ODI agreement with 12-month results (ICC 0.81–0.85) is close to the ODI stable state found in 5-to-6-week test-retest (ICC 0.82–0.86) [22-24] and between 1 and 2 years post-surgery (r = 0.71–0.89) [3, 5, 27] with all results displaying no difference in absolute score. After non-fusion surgery, 6-month ODI had excellent-to-outstanding discrimination values (AUC 0.88–0.94) and was higher than that reported in previous studies that

evaluated observed 3-month ODI (AUC 0.83 [12]) or 3-month ODI model (C-index 0.76 [13]).

Similarly, 6-month pain was highly predictive of 12-month pain after non-fusion surgery. Six-month pain had excellent discrimination, and the 6-month MCID concordance (82–86%) was close to the reported 1- to 2-year MCID concordance (84%) [5] and 1- to 2-year concordance in pain satisfaction (85–89%) [20]. This provides evidence that 6 months after non-fusion surgery no further PROM follow-up is needed, with resources potentially better spent on performance metrics that evaluate adjunct interventions such as revision surgery or therapeutic spinal injections [1, 21]. For fusion surgery a 1- or 2-year follow-up is recommended as outcomes do not stabilise before 1-year [3, 20].

This study allows clinicians to advise patients at postoperative review of their likely 'significant improvement' or 'acceptable level of symptoms' at 1 year. On the basis of the 6-month ODI score, 90% (±2%) of patients will maintain their improved/not improved ODI status at 12 months. Importantly this finding also indicates that a 'watch-and-wait' approach to patients without ODI improvement at 6 months will yield poor results as only 15% (12/77) in the main study group improved between 6 and 12 months.

Finally, our study showed that new-onset or exacerbation of pre-existing musculoskeletal comorbidities affected outcome after the 3-month stage and account for a minimum 34% of patients who lost 3-month ODI improvement. This finding supports the widespread adoption of the Finnish Spine Register postoperative questionnaire enquiry about new-onset comorbidities [28].

Strengths and limitations

Our single-centre study risks centre-related bias, which limits the generalizability of results. However, observed 3-month ODI predictive values were close to previously reported values which suggests that our results align with a common phenomenon. Likewise, 3-month results were reproduced in the 2021–2022 validation cohort that had different baseline characteristics and case-mix. The slightly poorer predictive performance of the observed 6-month PROM in the validation cohort is likely due to

proportionally more (23.3%, 10/43) achieving ODI improvement after 6 months in response to more patients receiving extended rehabilitation (>3 months) in 2021–2022. Six-month pain scores had borderline sufficient agreement (ICC 0.66–0.70). However, 6-month pain can be considered appropriate for group use as mean 6-month scores were stable at 12 months, and the agreement values were close to 1- to 2-year *r* values (0.71–0.78) [5] and pain 14-day test-retest ICC (0.7–0.8) [29, 30]. Six-month to 2-year outcome follow-up studies are needed to evaluate longer-term prediction.

Conclusion

This prospectively collected data study identified nine baseline factors linked to deterioration in symptoms 3 months after lumbar decompressive surgery. In the current climate of mandated data collection to justify escalating healthcare costs, this study shows that 6-month PROMs are reliable outcomes for discectomy or laminectomy. Fusion surgery requires a minimum 1-year PROM follow-up.

References

1. St-Pierre GH, Yang MH, Bourget-Murray J, Thomas KC, Hurlbert RJ, Matthes N (2018) Performance Indicators in Spine Surgery. Spine (Phila Pa 1976) 43:275-280. https://doi.org/10.1097/brs.0000000000002309

2. McGirt MJ, Parker SL, Asher AL, Norvell D, Sherry N, Devin CJ (2014) Role of prospective registries in defining the value and effectiveness of spine care. Spine (Phila Pa 1976) 39:S117-S128. https://doi.org/10.1097/BRS.0000000000000552

3. 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 (Phila Pa 1976) 42:1331-1338. https://doi.org/10.1097/BRS.0000000000002087

4. Asher AL, Chotai S, Devin CJ, Speroff T, Harrell FE, Jr., Nian H, Dittus RS, Mummaneni PV, Knightly JJ, Glassman SD, Bydon M, Archer KR, Foley KT, McGirt MJ (2016) Inadequacy of 3-month Oswestry Disability Index outcome for assessing individual longer-term patient experience after lumbar spine surgery. J Neurosurg Spine 25:170-180. https://doi.org/10.3171/2015.11.SPINE15872.

5. Staartjes VE, Siccoli A, de Wispelaere MP, Schröder ML (2019) Patient-reported outcomes unbiased by length of follow-up after lumbar degenerative spine surgery: Do we need 2 years of follow-up? Spine J 19:637-644. https://doi.org/10.1016/j.spinee.2018.10.004

6. Fekete TF, Loibl M, Jeszenszky D, Haschtmann D, Banczerowski P, Kleinstuck FS, Becker HJ, Porchet F, Mannion AF (2018) How does patient-rated outcome change over time following the surgical treatment of degenerative disorders of the thoracolumbar spine? Eur Spine J 27:700-708. https://doi.org/10.1007/s00586-017-5358-2.

7. Ruseckaite R, Mudunna C, Caruso M, Ahern S (2023) Response rates in clinical quality registries and databases that collect patient reported outcome measures: a scoping review. Health Qual Life Outcomes 21:71. https://doi.org/10.1186/s12955-023-02155-5

8. Whitmore RG, Curran JN, Ali ZS, Mummaneni PV, Shaffrey CI, Heary RF, Kaiser MG, Asher AL, Malhotra NR, Cheng JS, Hurlbert J, Smith JS, Magge SN, Steinmetz MP, Resnick DK, Ghogawala Z (2015) Predictive value of 3-month lumbar

discectomy outcomes in the NeuroPoint-SD Registry. J Neurosurg Spine 23:459-466. https://doi.org/10.3171/2015.1.SPINE14890.

9. Fritsch CG, Ferreira ML, Maher CG, Herbert RD, Pinto RZ, Koes B, Ferreira PH (2017) The clinical course of pain and disability following surgery for spinal stenosis: a systematic review and meta-analysis of cohort studies. Eur Spine J 26:324-335. https://doi.org/10.1007/s00586-016-4668-0

10. Machado GC, Witzleb AJ, Fritsch C, Maher CG, Ferreira PH, Ferreira ML (2016) Patients with sciatica still experience pain and disability 5 years after surgery: A systematic review with meta-analysis of cohort studies. Eur J Pain 20:1700-1709. https://doi.org/10.1002/ejp.893.

11. Asher AM, Oleisky ER, Pennings JS, Khan I, Sivaganesan A, Devin CJ, Bydon M, Asher AL, Archer KR (2020) Measuring clinically relevant improvement after lumbar spine surgery: is it time for something new? Spine J 20:847-856. https://doi.org/10.1016/j.spinee.2020.01.010

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

13. Rundell SD, Pennings JS, Nian H, Harrell FE, Khan I, Bydon M, Asher AL, Devin CJ, Archer KR (2020) Adding 3-month patient data improves prognostic models of 12-month disability, pain, and satisfaction after specific lumbar spine surgical procedures: development and validation of a prediction model. Spine J 20:600-613. https://doi.org/10.1016/j.spinee.2019.12.010

14. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 350:g7594. https://doi.org/10.1136/bmj.g7594

15. Fairbank JC, Pynsent PB (2000) The Oswestry Disability Index. Spine (Phila Pa 1976) 25:2940-2953. https://doi.org/10.1097/00007632-200011150-00017

16. van Hooff ML, Mannion AF, Staub LP, Ostelo RWJG, Fairbank JCT (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 16:1221-1230. https://doi.org/10.1016/j.spinee.2016.06.010

17. Fekete TF, Haschtmann D, Kleinstück FS, Porchet F, Jeszenszky D, Mannion AF (2016) What level of pain are patients happy to live with after surgery for lumbar degenerative disorders? Spine J 16:S12-S18. http://dx.doi.org/10.1016/j.spinee.2016.01.180
18. Portney LG, Watkins MP (2000) Foundations of Clinical Research: Application to Practice. Prentice Hall Upper Saddle River
19. Hosmer DW, Jr., Lemeshow S, Sturdivant RX (2013) Applied Logistic Regression. John Wiley & Sons, Inc., New York, UNITED STATES
20. Parai C, Hägg O, Lind B, Brisby H (2019) Follow-up of degenerative lumbar spine surgery—PROMs stabilize after 1 year: an equivalence study based on Swespine data. Eur Spine J 28:2187-2197. https://doi.org/10.1007/s00586-019-05989-0
21. Suri P, Pearson AM, Zhao W, Lurie JD, Scherer EA, Morgan TS, Weinstein JN (2017) Pain Recurrence After Discectomy for Symptomatic Lumbar Disc Herniation. Spine (Phila Pa 1976) 42:755-763. https://doi.org/10.1097/brs.0000000000001894
22. Davidson M, Keating JL (2002) A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther 82:8-24. https://doi.org/10.1093/ptj/82.1.8.
23. Maughan EF, Lewis JS (2010) Outcome measures in chronic low back pain. Eur Spine J 19:1484-1494. https://doi.org/10.1007/s00586-010-1353-6.
24. Cleland JA, Whitman JM, Houser JL, Wainner RS, Childs JD (2012) Psychometric properties of selected tests in patients with lumbar spinal stenosis. Spine J 12:921-931. https://doi.org/10.1016/j.spinee.2012.05.004
25. Coronado RA, Robinette PE, Henry AL, Pennings JS, Haug CM, Skolasky RL, Riley LH, Neuman BJ, Cheng JS, Aaronson OS, Devin CJ, Wegener ST, Archer KR (2021) Bouncing back after lumbar spine surgery: early postoperative resilience is associated with 12-month physical function, pain interference, social participation, and disability. Spine J 21:55-63. https://doi.org/10.1016/j.spinee.2020.07.013
26. Canseco JA, Chang M, Karamian BA, Nicholson K, Patel PD, Shenoy K, Hartman J, Kurd MF, Rihn JA, Anderson DG, Hilibrand AS, Kepler CK, Vaccaro AR, Schroeder GD (2022) Postoperative Glycemic Variability as a Predictor of Adverse Outcomes Following Lumbar Fusion. Spine (Phila Pa 1976) 47:E304-e311. https://doi.org/10.1097/brs.0000000000004214

27. 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 25:689-696. https://doi.org/10.3171/2015.8.SPINE15476.

28. Marjamaa J, Huttunen J, Kankare J, Malmivaara A, Pernaa K, Salmenkivi J, Pekkanen L (2023) The Finnish spine register (FinSpine): development, design, validation and utility. Eur Spine J 32:3731-3743. https://doi.org/10.1007/s00586-023-07874-3

29. Van Lerbeirghe J, Van Schaeybroeck P, Robijn H, Rasschaert R, Sys J, Parlevliet T, Hallaert G, Van Wambeke P, Depreitere B (2018) Cross-cultural adaptation and validation of the Dutch version of the core outcome measures index for low back pain. Eur Spine J 27:76-82. https://doi.org/10.1007/s00586-017-5255-8

30. Parai C, Hägg O, Lind B, Brisby H (2020) ISSLS prize in clinical science 2020: the reliability and interpretability of score change in lumbar spine research. Eur Spine J 29:663-669. https://doi.org/10.1007/s00586-019-06222-8

Table 1 Baseline characteristics and outcomes by procedure

Variable	Overall (*n* = 1096)	Discectomy (*n* = 540)	Laminectomy (*n* = 411)	Fusion (*n* = 145)
Baseline ODI (SD)	43.2 (16.9)	45.1 (17.3)	40.9 (16.3)	42.5 (16.6)
Baseline back pain (SD)	5.52 (2.81)	5.25 (2.80)	5.85 (2.85)	5.57 (2.64)
Baseline leg pain (SD)	6.61 (2.43)	6.58 (2.45)	6.74 (2.40)	6.35 (2.44)
Age, years (SD)	59.4 (15.2)	52.1 (14.0)	70.1 (9.8)	56.5 (14.5)
Sex (female), *n* (%)	529 (48.3%)	250 (46.3%)	202 (49.1%)	77 (53.1%)
Current smoker, *n* (%)	127 (11.6%)	72 (13.3%)	39 (9.5%)	16 (11.0%)
BMI kg/m^2 (SD)	27.4 (4.5)	27.2 (4.7)	27.7 (4.1)	27.1 (4.5)
One-level procedure, *n* (%)	946 (86.3%)	520 (96.3%)	305 (74.2%)	121 (83.4%)
Fusion, *n* (%)	145 (13.2%)	0 (0.0%)	0 (0.0%)	145 (100%)
Duration of leg pain, n (%)				
Up to 3 months	271 (24.7%)	189 (35.0%)	55 (13.4%)	27 (18.6%)
3–12 months	541 (49.4%)	263 (48.7%)	202 (49.1%)	76 (52.4%)
1–2 years	122 (11.1%)	48 (8.9%)	57 (13.9%)	17 (11.7%)
More than 2 years	162 (14.8%)	40 (7.4%)	97 (23.6%)	25 (17.2%)
Previous operation, *n* (%)				
None	906 (82.7%)	464 (85.9%)	326 (79.3%)	116 (80.0%)
One	159 (14.5%)	67 (12.4%)	70 (17.0%)	22 (15.2%)
2 or more	31 (2.8%)	9 (1.7%)	15 (3.6%)	7 (4.8%)
Scoliosis Cobb >30°, *n* (%)	27 (2.5%)	5 (0.9%)	18 (4.4%)	4 (2.8%)
Diabetes, *n* (%)	71 (6.5%)	20 (3.7%)	41 (10.0%)	10 (6.9%)
Rheumatic disorder, *n* (%)	92 (8.4%)	29 (5.4%)	52 (12.7%)	11 (7.6%)
Lower limb arthroplasty, *n* (%)	120 (10.9%)	22 (4.1%)	82 (20.0%)	16 (11.0%)
Mobility aided, *n* (%)	14 (1.3%)	2 (0.4%)	12 (2.9%)	0 (0.0%)
Depression / anxiety, *n* (%)	115 (10.5%)	63 (11.7%)	40 (9.7%)	12 (8.3%)
3-month work status, *n* (%)				
Unemployed	38 (3.5%)	23 (4.3%)	10 (2.4%)	5 (3.4%)
Other ¶ ¶¶	1058 (96.5%)	517 (95.7%)	401 (97.6%)	140 (96.6%)
Outcome				
3-month ODI (SD)	16.2 (14.9)	14.9 (14.7)	17.5 (15.4)	17.3 (13.5)
12-month ODI (SD)	17.1 (16.8)	15.0 (16.2)	19.5 (17.4)	18.0 (16.4)
3-month back pain (SD)	1.89 (2.07)	1.72 (1.94)	2.07 (2.30)	2.01 (1.82)
12-month back pain (SD)	2.24 (2.47) [1]	2.09 (2.38) [2]	2.49 (2.62) [3]	2.09 (2.30) [4]
3-month leg pain (SD)	1.85 (2.30)	1.65 (2.09)	2.30 (2.62)	1.35 (1.86)
12-month leg pain (SD)	2.14 (2.69) [5]	1.88 (2.50) [6]	2.49 (2.93) [7]	2.09 (2.54) [8]

ODI, Oswestry Disability Index; SD, standard deviation

¶ Other includes employed, sick leave, disability benefits, retired, homemaker, full-time student.

¶¶ No workers' compensation law or programme in the country of the study

Incomplete 12-month back pain data: [1] *n* = 1087, [2] *n* = 538, [3] *n* = 407, [4] *n* = 142

Incomplete 12-month leg pain data: [5] *n* = 1073, [6] *n* = 531, [7] *n* = 402, [8] *n* = 140

Table 2 Significant 3-month predictors of 12-month ODI, back pain, or leg pain

Variable	Oswestry Disability Index			Back pain			Leg pain		
	Coef.	95% CI	*p*-value	Coef.	95% CI	*p*-value	Coef.	95% CI	*p*-value
3-month ODI*	0.65	0.58–0.72	**<0.001**	0.04	0.03–0.06	**<0.001**	0.04	0.27–0.54	**<0.001**
3-month back pain*	0.77	0.03–0.12	**<0.001**	0.36	0.28–0.44	**<0.001**	0.01	-0.08–0.10	0.846
3-month leg pain*	-0.05	-0.43–0.34	0.810	0.04	-0.02–0.11	0.209	0.40	0.33–0.48	**<0.001**
Sex (female)	1.5	0.0–2.9	**0.050**	0.0	-0.3–0.2	0.770	0.1	-0.2–0.4	0.456
Current smoker	3.6	1.4–5.8	**0.001**	0.5	0.1–0.9	**0.008**	0.3	-0.1–0.8	0.131
BMI [a]			**0.033**			0.601			0.124
< 18.5 kg/m^2	11.7	4.8–18.6	**<0.001**	0.4	-0.8–1.7	0.505	1.5	0.2–2.8	**0.029**
25–29.9 kg/m^2	0.1	-1.6–1.7	0.934	-0.0	-0.3–0.3	0.983	0.2	-0.2–0.5	0.327
30–34.9 kg/m^2	0.5	-1.6–2.5	0.663	0.2	-0.1–0.6	0.185	0.2	-0.2–0.6	0.459
35–39.9 kg/m^2	1.6	-1.8–5.0	0.352	0.3	-0.3–0.9	0.342	0.5	-0.2–1.2	0.134
≥ 40.0 kg/m^2	0.1	-6.0–6.3	0.968	0.2	-0.8–1.3	0.679	-0.7	-1.9–0.5	0.253
Work status [b]			0.080			0.369			0.747
Unemployed	4.2	0.4–8.1	**0.031**	0.4	-0.3–1.0	0.284	0.1	-0.6–0.9	0.761
On disability benefits	2.8	-4.5–10.0	0.451	0.6	-0.6–1.9	0.323	0.5	-0.9–1.9	0.471
Duration of leg pain [c]			**0.008**			0.137			**<0.001**
3–12 months	1.5	-0.2–3.2	0.087	0.2	-0.1–0.5	0.206	0.3	0.0–0.7	0.051
1–2 years	2.6	0.1–5.1	**0.044**	0.3	-0.2–0.7	0.200	0.2	-0.3–0.6	0.524
2 years or more	4.0	1.7–6.4	**<0.001**	0.5	0.1–0.9	**0.022**	1.0	0.5–1.4	**<0.001**
Scoliosis Cobb angle >30°	2.5	-2.0–6.9	0.465	0.3	-5–1.1	0.439	0.9	0.0–1.8	**0.042**
Diabetes	5.2	2.4–8.1	**<0.001**	0.6	0.1–1.1	**0.025**	0.6	0.1–1.2	**0.024**
Rheumatic disorders	3.5	1.0–6.0	**0.006**	0.2	-0.2–0.7	0.308	0.7	0.2–1.2	**0.006**
Lower limb arthroplasty	3.9	1.6–6.3	**0.001**	0.5	0.1–0.9	**0.027**	0.3	-0.2–0.7	0.231
Mobility aided	8.0	1.9–14.1	**0.011**	0.5	-0.6–1.6	0.354	1.5	0.3–2.8	**0.014**

Coef.: Beta coefficient with 95% confidence intervals (CI), with *p*-values (≥0.05) in bold

* Increments in ODI (0–100 scale) are per 1 unit, and back or leg pain (0–10 scale) are per 1 unit

[a] BMI (Body Mass Index) reference category is '18.5–24.9' Kg/m^2

[b] Work status reference category is 'Other' (employed, sick leave, retired, homemaker, full-time student)

[c] Duration of leg pain reference category is < 3 months

Table 3 Comparison of 3-month model and observed 3-month outcome predictive performance

Outcome	ICC (95% CI)	*n*	Change score †	*n*	AUC for MCID (95% CI)	*n*	AUC for PASS (95% CI)	*n*
ODI								
Model	0.74 (0.68–0.78)	364	0.2 (*p* = 0.711)	364	0.86 (0.82–0.90)	364	0.87 (0.83–0.91)	364
Observed	0.74 (0.69–0.78)	364	-0.3 (*p* = 0.623)	364	0.87 (0.83–0.91)	364	0.85 (0.82–0.89)	364
Back pain								
Model	0.55 (0.48–0.62)	364	0.1 (*p* = 0.412)	364	0.76 (0.70–0.82)	350	0.79 (0.74–0.85)	364
Observed	0.55 (0.47–0.61)	364	-0.2 (*p* = 0.063)	364	0.73 (0.67–0.79)	350	0.78 (0.73–0.83)	364
Leg pain								
Model	0.58 (0.50–0.65)	361	0.2 (*p* = 0.143)	361	0.80 (0.74–0.85)	358	0.81 (0.75–0.86)	361
Observed	0.61 (0.54–0.67)	361	-0.1 (*p* = 0.341)	361	0.84 (0.79–0.89)	356	0.80 (0.75–0.85)	361

ODI, Oswestry Disability Index
ICC, interclass correlation coefficient
CI, confidence intervals
† Paired t-test, negative value indicates worsening symptoms
AUC, area under curve
MCID, minimal clinically important difference
PASS, patient acceptable symptom state

Table 4 Development cohort observed 3-month predictive performance and concordance with 12-month outcomes

Metric	Overall	*n*	Discectomy	*n*	Laminectomy	*n*	Fusion	*n*
Oswestry Disability Index								
ICC (95% CI)	0.71 (0.68–0.74)	1096	0.72 (0.69–0.76)	540	0.72 (0.67–0.77)	411	0.55 (0.43–0.66)	145
Change score †	-0.9 (***p* = 0.015**)	1096	-0.1 (*p* = 0.855)	540	-2.0 **(*p* <0.001)**	411	-0.7 (*p* = 0.572)	145
AUC for MCID (95% CI)	0.83 (0.80–0.86)	1095	0.89 (0.85–0.93)	539	0.80 (0.75–0.85)	411	0.74 (0.63–0.84)	145
AUC for PASS (95% CI)	0.86 (0.84–0.89)	1096	0.87 (0.84–0.91)	540	0.86 (0.82–0.90)	411	0.81 (0.73–0.89)	145
MCID concordance *n*, (%)	924 (84.4%)	1095	482 (89.4%)	539	325 (79.1%)	411	117 (80.7%)	145
PASS concordance *n*, (%)	888 (81.0%)	1096	446 (82.6%)	540	328 (79.8%)	411	114 (78.6%)	145
Back pain								
ICC (95% CI)	0.52 (0.47–0.57)	1087	0.54 (0.47–0.60)	538	0.50 (0.42–0.57)	407	0.52 (0.39–0.63)	142
Change score †	-0.4 (***p* < 0.001**)	1087	-0.4 **(*p* <0.001)**	538	-0.4 **(*p* <0.001)**	407	-0.1 (*p* = 0.721)	142
AUC for MCID (95% CI)	0.75 (0.72–0.79)	1027	0.77 (0.72–0.82)	514	0.73 (0.67–0.79)	378	0.78 (0.69–0.87)	135
AUC for PASS (95% CI)	0.77 (0.74–0.81)	1087	0.80 (0.76–0.85)	538	0.74 (0.69–0.79)	407	0.81 (0.73–0.89)	142
MCID concordance *n*, (%)	792 (77.1%)	1027	402 (78.2%)	514	284 (75.1%)	378	106 (78.5%)	135
PASS concordance *n*, (%)	855 (78.7%)	1087	440 (81.2%)	538	303 (74.4%)	407	112 (78.9%)	142
Leg pain								
ICC (95% CI)	0.51 (0.46–0.55)	1073	0.52 (0.45–0.58)	531	0.49 (0.42–0.57)	402	0.46 (0.31–0.59)	140
Change score †	-0.3 (***p* < 0.001**)	1073	-0.2 **(*p* = 0.013)**	531	-0.2 (*p* = 0.107)	402	-0.7 **(*p* <0.001)**	140
AUC for MCID (95% CI)	0.76 (0.72–0.80)	1056	0.78 (0.73–0.84)	525	0.76 (0.70–0.81)	393	0.67 (0.54–0.80)	138
AUC for PASS (95% CI)	0.76 (0.73–0.80)	1073	0.77 (0.72–0.82)	531	0.76 (0.71–0.81)	402	0.70 (0.59–0.82)	140
MCID concordance *n*, (%)	861 (81.5%)	1056	444 (84.6%)	525	302 (76.8%)	393	115 (83.3%)	138
PASS concordance *n*, (%)	838 (78.1%)	1073	433 (81.5%)	531	292 (72.6%)	402	113 (80.7%)	140

ICC, interclass correlation coefficient
CI, confidence interval
† Paired t-test, negative value indicates worsening symptoms, ($p \leq 0.05$ in bold)
AUC, area under curve
PASS, patient acceptable symptom state
MCID, minimal clinically important difference

Table 5 Development cohort observed 6-month predictive performance and concordance with 12-month outcomes

Metric	Overall	*n*	Discectomy	*n*	Laminectomy	*n*	Fusion	*n*
Oswestry Disability Index								
ICC (95% CI)	0.82 (0.79–0.85)	481	0.81 (0.76–0.85)	238	0.85 (0.80–0.89)	160	0.74 (0.60–0.83)	83
Change score †	-1.3 (***p* = 0.004**)	481	-0.3 (*p* = 0.558)	238	-1.3 (*p* = 0.090)	160	-4.0 (***p* = 0.002**)	83
AUC for MCID (95% CI)	0.92 (0.89–0.95)	480	0.91 (0.87–0.96)	237	0.94 (0.90–0.99)	160	0.88 (0.78–0.97)	83
AUC for PASS (95% CI)	0.90 (0.87–0.93)	481	0.88 (0.81–0.94)	238	0.92 (0.87–0.97)	160	0.89 (0.82–0.97)	83
MCID concordance *n*, (%)	435 (90.6%)	480	217 (91.6%)	237	145 (90.6%)	160	73 (88.0%)	83
PASS concordance *n*, (%)	420 (87.3%)	481	213 (89.5%)	238	140 (87.5%)	160	67 (80.7%)	83
Back pain								
ICC (95% CI)	0.68 (0.63–0.73)	461	0.69 (0.62–0.76)	229	0.70 (0.61–0.78)	154	0.57 (0.41–0.71)	78
Change score †	-0.1 (*p* = 0.190)	461	-0.1 (*p* = 0.292)	229	-0.1 (*p* = 0.639)	154	-0.2 (*p* = 0.472)	78
AUC for MCID (95% CI)	0.82 (0.77–0.87)	432	0.84 (0.78–0.90)	218	0.84 (0.75–0.92)	139	0.71 (0.57–0.86)	75
AUC for PASS (95% CI)	0.85 (0.80–0.89)	461	0.88 (0.82–0.93)	229	0.83 (0.75–0.90)	154	0.79 (0.66–0.92)	78
MCID concordance *n*, (%)	361 (83.6%)	432	182 (83.5%)	218	119 (85.6%)	139	60 (80.0%)	75
PASS concordance *n*, (%)	389 (84.2%)	461	200 (87.3%)	229	124 (80.5%)	154	65 (83.3%)	78
Leg pain								
ICC (95% CI)	0.65 (0.60–0.70)	450	0.66 (0.58–0.73)	226	0.68 (0.58–0.76)	148	0.55 (0.34–0.70)	76
Change score †	-0.2 (*p* = 0.094)	450	0.0 (*p* = 0.745)	226	-0.1 (*p* = 0.545)	148	-0.9 (***p* < 0.001**)	76
AUC for MCID (95% CI)	0.79 (0.73–0.85)	446	0.82 (0.72–0.91)	223	0.82 (0.73–0.91)	148	0.69 (0.54–0.85)	75
AUC for PASS (95% CI)	0.82 (0.77–0.87)	450	0.85 (0.82–0.90)	226	0.85 (0.78–0.92)	148	0.77 (0.62–0.89)	76
MCID concordance *n*, (%)	374 (83.9%)	446	192 (86.1%)	223	122 (82.4%)	148	60 (80.0%)	75
PASS concordance *n*, (%)	376 (83.6%)	450	196 (86.7%)	226	120 (81.1%)	148	60 (78.9%)	76

ICC, interclass correlation coefficient
CI, confidence interval
† Paired t-test, negative value indicates worsening symptoms, ($p \leq 0.05$ in bold)
AUC, area under curve
PASS, patient acceptable symptom state
MCID, minimal clinically important difference