


# Can a Successful Outcome After Surgery for Lumbar Disc Herniation Be Defined by the Oswestry Disability Index Raw Score?

Global Spine Journal  
2020, Vol. 10(1) 47-54  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2192568219851480  
journals.sagepub.com/home/gsj



David A. T. Werner, MD<sup>1,2</sup> , Margreth Grotle, MD, PhD<sup>3,4</sup>,  
Sasha Gulati, MD, PhD<sup>5,6</sup>, Ivar M. Austevoll, MD<sup>7</sup>,  
Mattis A. Madsbu, MD<sup>5</sup>, Greger Lønne, MD, PhD<sup>6,8,9</sup>,  
and Tore K. Solberg, MD, PhD<sup>1,2,9</sup>

## Abstract

**Study Design:** Prospective multicenter cohort study.

**Objective:** To investigate (1) the discriminative ability and cutoff estimates for success 12 months after surgery for lumbar disc herniation on the Oswestry Disability Index (ODI) raw score compared with a change and a percentage change score and (2) to what extent these clinical outcomes depend on the baseline disability.

**Methods:** A total of 6840 patients operated for lumbar disc herniation from the Norwegian Registry for Spine Surgery (NORspine) were included. In receiver operating characteristic (ROC) curve analyses, a global perceived effect (GPE) scale (1-7) was used as an external anchor. Success was defined as categories 1-2, “completely recovered” and “much better.” Cutoffs for success for subgroups with different preoperative disability were also estimated.

**Results:** When defining success after surgery for lumbar disc herniation, the accuracy (sensitivity, specificity, area under the curve, 95% CI) for the ODI raw score (0.83, 0.87, 0.930, 0.924-0.937) was comparable to the ODI percentage change score (0.85, 0.85, 0.925, 0.918-0.931), and higher than the ODI change score (0.79, 0.73, 0.838, 0.830-0.852). The cutoff for success was highly dependent on the amount of baseline disability (low-high), with cutoffs ranging from 13 to 28 for the ODI raw score and 39% to 66% for ODI percentage change. The ODI change score (points) was not as accurate.

**Conclusion:** The 12-month ODI raw score, like the ODI percentage change score, can define a successful outcome with excellent accuracy. Adjustment for the baseline ODI score should be performed when comparing outcomes across groups, and one should consider using cutoffs according to preoperative disability (low, medium, high ODI scores).

## Keywords

PASS, ODI, success criteria, lumbar disc surgery, PROM, lumbar disc herniation

<sup>1</sup> University Hospital of Northern Norway, Tromsø, Norway

<sup>2</sup> University of Tromsø, Institute of Clinical Medicine, Tromsø, Norway

<sup>3</sup> Oslo Metropolitan University, Oslo, Norway

<sup>4</sup> Oslo University Hospital, Oslo, Norway

<sup>5</sup> St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

<sup>6</sup> Norwegian University of Science and Technology, Trondheim, Norway

<sup>7</sup> Kysthospitalet in Hagevik, Haukeland University Hospital, Bergen, Norway

<sup>8</sup> Innlandet Hospital Trust, Lillehammer, Norway

<sup>9</sup> Norwegian Registry for Spine Surgery, Northern Norway Regional Health Authority, Bodø, Norway

## Corresponding Author:

David Werner, Department of Neurosurgery, University Hospital of Northern Norway, Sykehusveien 38, 9019 Tromsø, Norway.

Email: mail@david-werner.com



Creative Commons Non Commercial No Derivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

## Introduction

In Norway, operative treatment of lumbar disc herniation is the most frequently performed spine surgery procedure in patients younger than 50 years.<sup>1</sup> The indication for surgery is most often relative, that is, reducing pain-related disability.<sup>2</sup> To compare treatment effects across interventions and institutions, changes in patient-reported outcome measures (PROMs) are frequently used, but their interpretation is complex. Previous studies have used score changes of the Oswestry Disability Index (ODI) to calculate clinically meaningful improvements, such as cutoffs for a “successful outcome.”<sup>3-6</sup> However, the amount of change needed for success is highly dependent on the baseline scores of the PROMs.<sup>7</sup>

Studies from other medical fields, such as rheumatology, have used a Patient Acceptable Symptom State (PASS)<sup>8-10</sup> in order to define a cutoff for a successful outcome on a PROM. The PASS could be viewed as a separate entity to the underlying change score.<sup>8,9,11</sup> We have previously defined cutoffs for success based on PROM change scores for patients operated for lumbar disc herniation, by either open- or micro-discectomy.<sup>6</sup> In a recent study, we found cutoffs on the 12-month ODI raw score that had the highest accuracy for identifying cases that could be classified as failed and worsened after lumbar disc surgery,<sup>11</sup> indicating that patients could be more focused on their current disability than on health changes when reporting clinical outcomes.

In the present study, we sought (1) to define the discriminative ability and cutoff estimates of success for a 12-month ODI raw score (current disability), an ODI change and ODI percentage change score and (2) to investigate if these clinical outcomes depend on the baseline disability, that is, the preoperative ODI score. We defined success by the patient’s ratings of a substantial effect of surgery (Global Perceived Effect scale, GPE), that is, when the patient is feeling “completely recovered” (GPE = 1), or “much better” (GPE = 2) 12 months after the operation. Such information would aid in the classification and understanding of successful outcome, facilitating reporting and comparisons of treatment results.

## Materials and Methods

### Patient Population and Data Collection

A total of 6840 patients operated for lumbar disc herniation at 38 different surgical units between January 1, 2007 and February 28, 2014 were followed for 12 months, according to the standard protocol of the Norwegian registry for spine surgery (NORspine). The NORspine is a comprehensive clinical registry for quality control and research. During the study period, the NORspine comprised 95% (38 of 40) Norwegian public and private centers performing lumbar disc surgery. Completeness, the proportion of patients operated on for lumbar disc herniation reported to the NORspine, was 65%.<sup>1</sup> The registry excluded patients unable to consent, children aged <16 years, patients with documented drug abuse or severe psychiatric disorders, and patients with traumatic, infectious or malignant

conditions in the spine. In this study, we included all elective and emergency cases operated for lumbar disc herniation. Fusion procedures and/or procedures including laminectomy were not included.

Informed consent was obtained from all patients and participation was neither mandatory, nor required to gain access to healthcare. The registry protocol has been approved by the Data Inspectorate of Norway. The study protocol had been submitted to the regional ethical committee for medical research which categorized it as a clinical audit study, not in need of their formal approval.<sup>12</sup>

At admission for surgery, the patients completed a baseline questionnaire on demographics, lifestyle issues and PROMs (Figures 10x-13x, appendix). During the hospital stay, the surgeon recorded data concerning diagnosis, treatment, and comorbidity on a standard registration form (Figures 14x-15x, appendix). Twelve months after surgery a questionnaire was distributed by regular post, completed at home by the patients, and returned in prestamped envelopes to the central registry unit without involvement of the treating hospitals. One reminder with a new copy of the questionnaire was sent to those who did not respond.

### Patient-Reported Outcome Measures

This study is based on the cohort used in a previous study by the authors.<sup>11</sup> The ODI version 2.1a was used to assess pain-related disability. It contains 10 questions on limitations of activities of daily living. Each item is rated 0 to 5 and then transferred into a percentage score ranging from 0 (none) to 100 (maximum pain-related disability).<sup>13</sup>

The patient-rated benefit of the operation was rated on the GPE at follow-up.<sup>11,14</sup> The response alternatives were as follows: 1 = “completely recovered,” 2 = “much better,” 3 = “somewhat better,” 4 = “no change,” 5 = “somewhat worse,” 6 = “much worse,” and 7 = “worse than ever.”

### Statistics

All statistical analyses were performed with the Statistical Package for the Social Sciences (SPSS, IBM Version 23.0).

We excluded all patients who did not respond at 12 months. This strategy was based on a study from the NORspine on a comparable patient population, and a recent and similar Danish registry study, both indicating that patients lost to follow-up could be handled as missing at random in the analyses.<sup>15,16</sup>

We assessed the mean 12-month ODI raw score, as well as the mean ODI percentage change score and the mean ODI change score after 12 months against the GPE by one-way analyses of variance (ANOVA) with post hoc analysis (Tukey,  $\alpha = .05$ ) and by analyses of covariance (ANCOVA, generalized linear model) with adjustment for baseline scores. Correlation analyses between the different ODI tools and the GPE were done by Spearman rank correlation.

Cutoffs for all scores were estimated by receiver operating characteristic (ROC) curves. We calculated cutoffs for a

**Table 1.** Baseline Characteristics of the Study Population.

Characteristic	Mean (SD)
Age (years)	48.7 (13.6)
BMI (kg/m <sup>2</sup> )	26.6 (4.2)
ASA	1.5 (0.6)
ODI	46 (18.9)
Backpain, NRS	6.2 (2.5)
Legpain, NRS	6.9 (2.2)
Characteristic	n (%)
Female	3952 (58)
Smoker	1936 (27)
Married	3827 (56)
Emergency surgery	653 (10)
Lower education	4279 (63)
Comorbidity	1664 (28)
Previously operated	1417 (21)
Sickness benefits	4180 (61)

Abbreviations: SD, standard deviation; BMI, body mass index; NRS, numeric rating scale; ASA, American Society of Anesthesiologists score; ODI, Oswestry Disability Index score.

substantial improvement from baseline (GPE 1-2 vs 3-7), termed success.

To determine the cutoff with the highest sensitivity and specificity, the closest point to the upper left corner of the ROC curve was calculated from the coordinates of the curve. The area under the curve (AUC) determined how well the instruments differentiated between the outcome groups. An AUC value of >0.70 was considered acceptable, >0.80 good, and >0.9 excellent. The overall accuracy for each cutoff was calculated with a confusion matrix.<sup>17</sup>

To be able to study the impact of low and high baseline disability on the outcome cutoffs (success criteria), we split the patient sample based on the baseline ODI score into low (<25th percentile), medium (25th-75th percentile) and high disability (>75th percentile) and calculated cutoffs for the 12-months ODI raw score, ODI percentage change and ODI change after 12 months, for each of these percentiles.

Floor and ceiling effects were assessed by calculating the frequency of the highest and lowest possible scores at baseline. If 15% of patients had a minimal or maximal score value at baseline, these were considered as floor or ceiling effects.<sup>18,19</sup>

**Results**

Baseline characteristics of both respondents and nonrespondents of this patient population have been shown and discussed in a previous study.<sup>11</sup> Characteristics of the study population are listed in Table 1. Follow-up data after 12 months were available for 6840 (69%) out of 9930 of patients. The sample was divided into low ODI baseline (n = 1617), medium ODI baseline (n = 3718), and high ODI baseline (n = 1505). Only 13 data points (0.2%) were missing for the baseline ODI. At 12-month follow-up, 40 values (0.6%) were missing for the GPE and 11 (0.2%) on the ODI. As shown in a previous article, the lost to

**Table 2.** Baseline Adjusted Mean Scores of the 12-Month Oswestry Disability Index (ODI) Raw Score, ODI Change Score, and ODI Percentage Change Score Analysis of Covariance (ANCOVA) by the Global Perceived Effectiveness (GPE) Scale.<sup>a</sup>

	GPE						
	Completely Recovered (1)	Much Better (2)	Somewhat Better (3)	No Change (4)	Somewhat Worse (5)	Much Worse (6)	Worse Than Ever (7)
n (%)	1659 (24)	3265 (48)	1093 (16)	358 (5)	216 (3)	153 (2)	66 (1)
Mean baseline ODI (SD)	45.4 (19.6)	46.1 (19.1)	47.3 (18.0)	42.0 (17.1)	44.7 (16.9)	50.7 (16.5)	50.9 (17.1)
12-month ODI raw score (95% CI)	2.9 (2.4-3.3)	13.6 (13.3-14.0)	30.9 (30.3-31.5)	39.4 (38.4-40.4)	42.3 (41-43.6)	51.6 (50-53.1)	58.9 (56.5-61.2)
ODI change (95% CI)	43.1 (42.7-43.6)	32.4 (32.1-32.7)	15.1 (14.5-15.7)	6.6 (5.6-7.6)	3.7 (2.5-5.0)	-5.6 (-7.1 to -4.1)	-12.9 (-15.2 to -1.5)
ODI % change (95% CI)	93.4 (91.9-94.9)	66.7 (65.6-67.8)	26.7 (24.9-28.6)	1.3 (-1.9 to 4.5)	-6.7 (-10.8 to -2.5)	-13.7 (-18.6 to -8.7)	-30.7 (-38.3 to -23.0)

Abbreviation: CI, confidence interval.

<sup>a</sup>Negative prefix indicates a worsening of the ODI from baseline. The mean ODI score for the entire study population prior to surgery was 46.

**Table 3.** Cutoff for the 12-Month Oswestry Disability Index (ODI) Raw Score, the 12-Month ODI Change Score, and the 12-Month ODI Percentage Change Score, Classifying Success in the Whole Study Population (Receiver Operating Curve [ROC] Analyses) and Accuracy (Confusion Matrix).

	AUC	95% CI	Cutoff	Sens/ Spec	Accuracy (%)
12-month ODI raw score	0.93	0.92-0.94	19	0.83/0.87	84
ODI change score	0.84	0.83-0.85	19	0.79/0.73	78
ODI percentage change score	0.93	0.92-0.93	52	0.85/0.85	85

Abbreviations: AUC, area under the curve; 95% CI, 95% confidence interval; Sens, sensitivity; Spec, specificity.

**Table 4.** Cutoffs for the 12-Month Oswestry Disability Index (ODI) Raw Score, the 12-Month ODI Change Score, and the 12-Month ODI Percentage Change Score When Classifying Success in Each of the 3 ODI Baseline Subgroups.<sup>a</sup>

Baseline Subgroup	AUC	95% CI	Cutoff	Sens/ Spec	Accuracy (%)
12-month ODI raw score					
<25th perc	0.92	0.90-0.93	13	0.81/0.88	83
25th-75th perc	0.95	0.94-0.95	21	0.85/0.89	86
>75th perc	0.94	0.93-0.96	28	0.89/0.85	88
ODI change					
<25th perc	0.89	0.88-0.91	9	0.77/0.84	79
25th-75th perc	0.92	0.91-0.93	24	0.83/0.84	83
>75th perc	0.92	0.91-0.94	48	0.85/0.84	85
ODI % change					
<25th perc	0.91	0.90-0.93	39	0.82/0.84	83
25th-75th perc	0.94	0.94-0.95	53	0.86/0.88	86
>75th perc	0.94	0.93-0.96	66	0.85/0.88	88

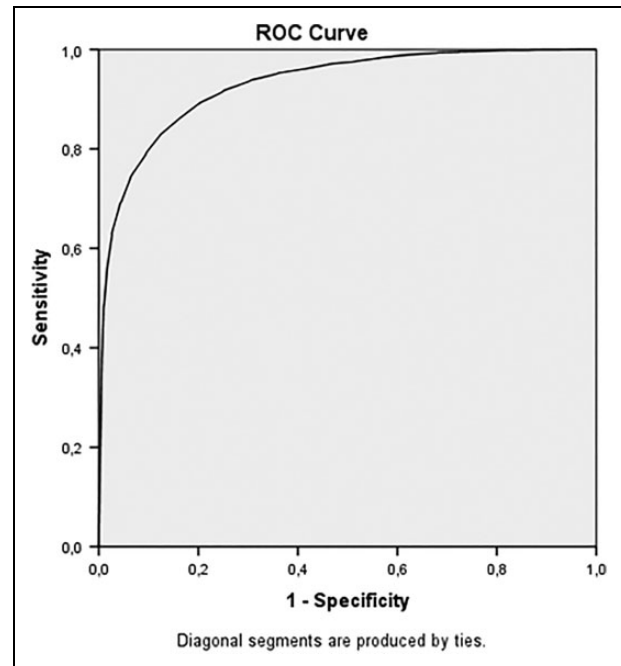
Abbreviations: AUC, area under the curve; 95% CI, 95% confidence interval; Sens, sensitivity; Spec, specificity; perc, percentile.

<sup>a</sup>Analyses were done by receiver operating curve (ROC) analyses. Overall accuracy was determined by a confusion matrix.

follow-up group contained more smokers, fewer with higher education, more sickness benefits recipients, more previously operated patients and fewer cases operated for paresis.<sup>11</sup> The Spearman correlation coefficients were 0.6 for the 12-month ODI change ( $P < .001$ ), 0.8 for the 12-month ODI percentage change ( $P < .001$ ), and 0.8 for the 12-month ODI raw score ( $P < .001$ ). ANOVA with post hoc analysis indicated that the 12-month ODI raw scores of all estimates were significantly different between GPE categories. For each outcome, baseline adjusted mean ODI scores (ANCOVA), are shown in Table 2.

**Cutoffs for Success**

The discriminative ability for success was significantly higher for the ODI percentage change and the 12-month ODI raw score in comparison with the ODI change score (Table 3). In the subgroup analyses, we found that the cutoffs for success



**Figure 1.** Receiver operating curve for the 12-month Oswestry Disability Index (ODI) score cutoff for “success”. AUC (area under the curve) = 0.93 (0.92-0.94).

were dependent on the baseline ODI score. Patients with a low baseline ODI (<25th percentile, ODI score <32) had a cutoff on the 12-month ODI raw score (ODI % change) of 13 points (39%), those with medium baseline ODI (25th-75th percentile, ODI score 32-60) a cutoff of 21 points (55%), and those with high baseline ODI (>75th percentile, ODI score >60) a cutoff of 28 points (66%). The cutoffs for all ODI scores for all the different ODI baseline groups are listed in Table 4. Figure 4 shows that for the subgroups, the change cutoff (downward arrow) reaches the ODI raw score cutoff for success (horizontal line).

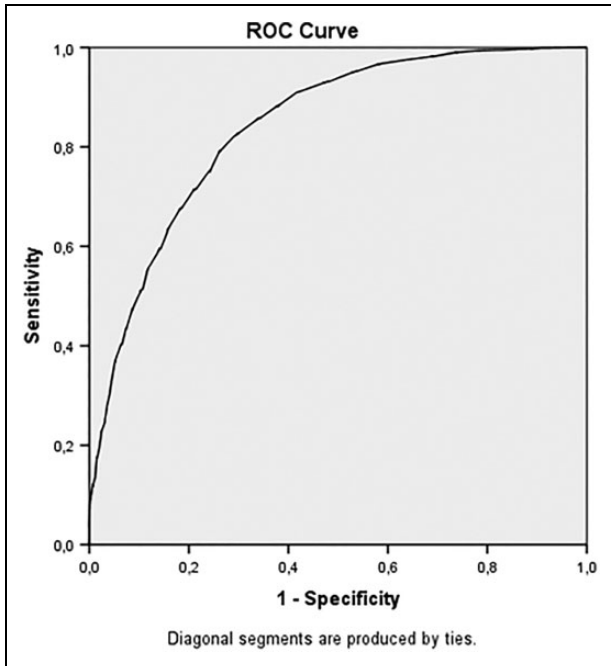
For the entire population, the cutoffs were 19 (ODI raw score), 19 (ODI change score), and 52% (ODI percentage change score) (Table 3). AUCs were high for all curves, ranging from 0.84 (ODI change score) to 0.93 (ODI raw score, ODI percentage score) (Figures 1–3).

**Proportion of Success at 12 Months**

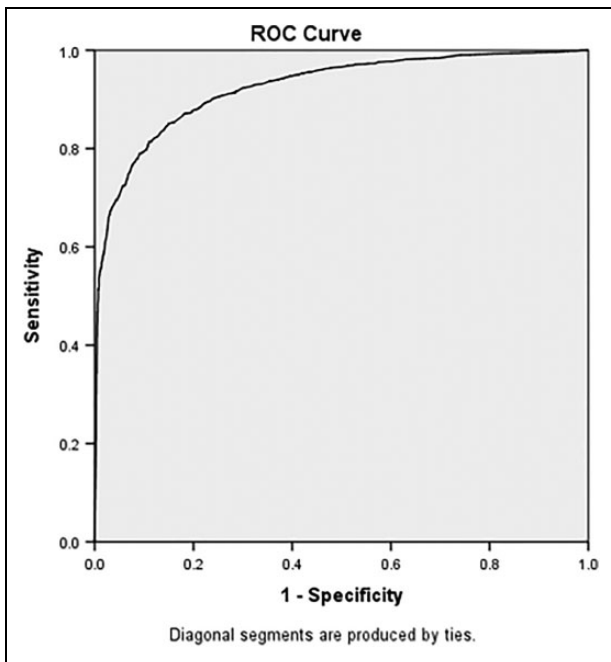
For the entire population, the ODI percentage change score and the ODI raw scores corresponded better to a successful outcome (groups 1 and 2 on the GPE-scale) than the ODI change score (Table 3). Table 5 shows the proportion of cases classified as success 12 months after surgery. Table 1x (appendix) shows these proportions using separate cutoffs based on the different baseline ODI levels (percentiles). The ODI percentage change classified the highest proportions of success for the whole sample.

**Floor and Ceiling Effects**

No floor or ceiling effects were detected.



**Figure 2.** Receiver operating curve for the 12-month Oswestry Disability Index (ODI) change cutoff for “success”. AUC (area under the curve) = 0.84 (0.83-0.85).



**Figure 3.** Receiver operating curve for the 12-month Oswestry Disability Index (ODI) percentage change cutoff for “success”. AUC (area under the curve) = 0.93 (0.92-0.93).

**Discussion**

We found that success after surgery for lumbar disc herniation could as accurately be defined by the 12-month ODI raw score,

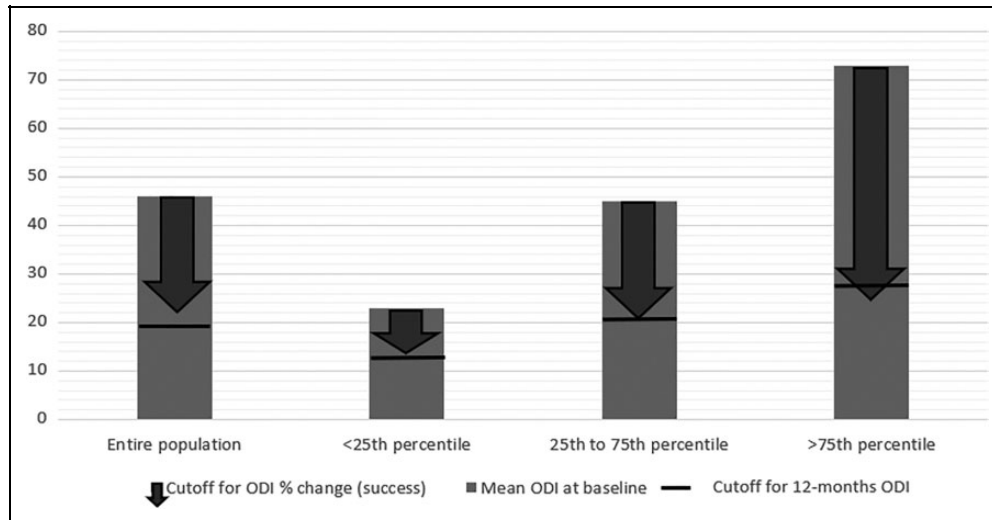
**Table 5.** Total Number (N) of Cases Classified as Success by Each Oswestry Disability Index (ODI) Outcome Tool, for the Entire Population.

	Total population, n (%)
12-month ODI score	
Success	4322 (63)
No success	2507 (37)
ODI change score	
Success	4391 (64)
No success	2425 (36)
ODI percentage change score	
Success	4446 (65)
No success	2361 (35)

as by the ODI percentage change score, and more accurately than by the ODI change scores from baseline. In a previous study we also found that the 12-month ODI raw score was more robust than the change scores for defining failure and worsening.<sup>11</sup> In the subgroup analyses we found that the cutoffs for success were dependent on the baseline ODI score. For those with low baseline disability the amount of improvement from baseline was considerably lower than for those with high baseline disability (Table 4). This dependency on the baseline score illustrates that patients perceive their postoperative improvements based on the amount of disability they experienced prior to surgery. Thus, in a patient sample with a low mean ODI other criteria for a positive outcome need to be applied, than in a patient sample with medium or high baseline ODI scores. This also implies that the previous recommendation to use a 30% change score cutoff for minimal clinical change<sup>3</sup> must be reconsidered for patients with medium and high baseline ODI scores. Our results confirm the importance of adjusting for baseline scores when comparing success rates between groups, for example, hospitals and surgical interventions.<sup>11,20</sup> When evaluating outcomes for individual patients or groups, one should consider using cutoffs according to baseline disability (low, medium, or high ODI scores). Moreover, statistical studies aimed at predicting outcome after surgery for lumbar disc herniation should be modeled with adjustment for preoperative ODI score, for example, by stratification.

The ODI change score had the lowest accuracy for defining success, especially among patients with high and low baseline disability. Therefore, we only recommend using the 12-month ODI raw score and the ODI percentage change score cutoffs.

Interestingly, the success rates among patients with low and high baseline scores were the same. This indicates that patients with low baseline disability may have higher demands for physical performance, and they may be more sensitive to smaller improvements which they would consider meaningful compared to those with high baseline disability.<sup>21</sup> Prior to surgery, these issues should be discussed with the patient. Differences in symptom tolerance before and after the operation may also reflect variation in patient expectations and coping strategies.



**Figure 4.** Mean baseline Oswestry Disability Index (ODI) (bar) with the 12-month ODI percentage change cutoff for success (arrow) and the 12-month ODI score cutoff for success (line).

Studies in rheumatology suggest that a treatment needs to reduce symptom intensity below a certain threshold to be perceived as successful by the patient. This threshold has been termed a “patient acceptable symptom state” (PASS).<sup>8,9,22</sup> As illustrated in Figure 4, it does not matter if a patient experiences, for instance, a 30% or 50% improvement of the baseline score, as long as he or she achieves the cutoff for the 12-month ODI raw score. Moreover, for the study population as a whole, the 12-month ODI raw score cutoff for success ( $\leq 19$ ) corresponds to what van Hooff et al<sup>23</sup> defined a cutoff for a patient acceptable symptom state (PASS = ODI  $\leq 22$  at follow-up).

### Methodological Challenges

By collecting data from “real-world” clinical practice, studies from clinical registries not only have advantages such as large sample sizes and high external validity but also limitations such as lower follow-up rates compared with closely monitored clinical trials. Still, there is increasing evidence in the literature that observational studies conducted according to the STROBE check list report corresponding results similar to those found in randomized controlled trials.<sup>24</sup>

Loss to follow-up was 31%. In three previous studies from the Scandinavian spine registries (NORspine, SWEspine, and DANEspine), dropout cases (rates of 12%-38%) were traced and interviewed. These studies found the same differences in baseline characteristics that we found between patients who responded and those who did not, yet the same clinical outcomes at 1 and 2 years of follow-up.<sup>15,16,25</sup> Thus, we do not expect that loss to follow-up would bias our success rate estimates. Furthermore, the aim of the study was not a clinical effectiveness evaluation, but rather to define cutoffs for success over the wide range of different outcomes found in this large cohort. Generalizability of our findings beyond the Norwegian population is supported by previous comparative

studies in Scandinavian countries and the United States, who report conceding results on baseline data and clinical outcomes (effect sizes).<sup>26-28</sup>

Using the GPE as an external anchor has been criticized since recall bias may exist. Moreover, the patients tend to be more focused on their current health state than health change when responding on a GPE scale, indicating a weakness of its construct validity.<sup>11,29</sup> The ideal anchor should objectively measure the patient’s status before and after surgery with high reliability and validity. It should be easy to use, and universally applicable in different clinical settings. However, to the best of our knowledge no such anchor exists. In the search for such a tool, other variables have been evaluated by different research groups, such as return to work, use of painkillers, or surgeon-reported outcome. However, such measures also have limitations, namely, bias due to selection of certain subgroups, and subjective information based on surgeon’s assessment of the clinical outcome.<sup>30-33</sup> Acknowledging these limitations, both the FDA (Food and Drug Administration) and the IMMPACT consensus group recommend a 7-point Likert-type scale, like the GPE, to be used as an external anchor.<sup>34,35</sup>

### Conclusion

The ODI raw score can be used to define a successful outcome 12 months after surgery for lumbar disc herniation with high accuracy, similar that of the ODI percentage change. The ODI change score in points was not as accurate. Since these cutoffs are point estimates and vary depending on the baseline disability, adjustment for the baseline ODI should be performed when comparing success rates between hospitals or interventions. We recommend using ODI raw score or ODI percentage change (value in parentheses) cutoffs for success, according to their level of baseline disability, low= 13 points (39%), medium= 21 points (53%), or high= 28 points (66%).


## Declaration of Conflicting Interests

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

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by The Norwegian Medical Association–Foundation for Quality Improvement and Patient Safety, and The Regional Health Authority of Northern Norway.

## ORCID iD

David A. T. Werner  <https://orcid.org/0000-0002-0865-7395>

## Supplemental Material

The supplemental material is available in the online version of the article.

## References

- Solberg TK, Olsen LR. Nasjonalt kvalitetsregister for ryggkirurgi (NKR): Årsrapport for 2015 med plan for forbedringstiltak 2016. In: Solberg TK, ed. *NORspine Annual Report*. Bodø, Norway: NORspine; 2016:46.
- Deyo RA, Mirza SK. Clinical practice. Herniated lumbar intervertebral disk. *N Engl J Med*. 2016;374:1763-1772. doi:10.1056/NEJMcpl512658
- Ostelo RWJG, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)*. 2008;33:90-94. doi:10.1097/BRS.0b013e31815e3a10
- Glassman SD, Copay AG, Berven SH, Polly DW, Subach BR, Carreon LY. Defining substantial clinical benefit following lumbar spine arthrodesis. *J Bone Joint Surg Am*. 2008;90:1839-1847.
- Copay AG, Martin MM, Subach BR, et al. Assessment of spine surgery outcomes: inconsistency of change amongst outcome measurements. *Spine J*. 2010;10:291-296.
- Solberg T, Johnsen LG, Nygaard ØP, Grotle M. Can we define success criteria for lumbar disc surgery? Estimates for a substantial amount of improvement in core outcome measures. *Acta Orthop*. 2013;84:196-201.
- de Vet HCW, Foumani M, Scholten MA, et al. Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. *J Clin Epidemiol*. 2015;68:518-524. doi:10.1016/j.jclinepi.2014.07.008
- Escobar A, Riddle DL. Concordance between important change and acceptable symptom state following knee arthroplasty: the role of baseline scores. *Osteoarthritis Cartilage*. 2014;22:1107-1110. doi:10.1016/j.joca.2014.06.006
- Tubach F, Ravaud P, Baron G, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis*. 2005;64:29-33. doi:10.1136/ard.2004.022905
- Myles PS, Myles DB, Galagher W, et al. Measuring acute postoperative pain using the visual analog scale: the minimal clinically important difference and patient acceptable symptom state. *Br J Anaesth*. 2017;118:424-429. doi:10.1093/bja/aew466
- Werner DAT, Grotle M, Gulati S, et al. Criteria for failure and worsening after surgery for lumbar disc herniation: a multicenter observational study based on data from the Norwegian Registry for Spine Surgery. *Eur Spine J*. 2017;26:2650-2659. doi:10.1007/s00586-017-5185-5
- Ruyter KW. The Regional Ethical Committee for Medical Research, South-Eastern Norway Regional Health Trust: 22845518 24.06.2015. 2015.
- Fairbank JCT. Why are there different versions of the Oswestry Disability Index? *J Neurosurg Spine*. 2014;20:83-86. doi:10.3171/2013.9.SPINE13344
- Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol*. 2010;63:760-766.e1.
- Hojmark 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*. 2015;25:282-286. doi:10.1007/s00586-015-4127-3
- Solberg TK, Sørli A, Sjaavik K, Nygaard ØP, Ingebrigtsen T. Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? *Acta Orthop*. 2011;82:56-63. doi:10.3109/17453674.2010.548024
- Fawcett T. An introduction to ROC analysis. *Pattern Recogn Lett*. 2006;27:861-874. doi:10.1016/j.patrec.2005.10.010
- de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res*. 2007;16:131-142.
- Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34-42.
- Fekete TF, Haschtmann D, Kleinstück FS, Porchet F, Jeszenszky D, Mannion AF. What level of pain are patients happy to live with after surgery for lumbar degenerative disorders? *Spine J*. 2016;16(4 suppl):S12-S18. doi:10.1016/j.spinee.2016.01.180
- Mancuso CA, Duculan R, Stal M, Girardi FP. Patients expectations of lumbar spine surgery. *Eur Spine J*. 2014;24:2362-2369. doi:10.1007/s00586-014-3597-z
- Tubach F, Dougados M, Falissard B, Baron G, Logeart I, Ravaud P. Feeling good rather than feeling better matters more to patients. *Arthritis Rheum*. 2006;55:526-530. doi:10.1002/art.22110
- van Hooff ML, Mannion AF, Staub LP, Ostelo RWJG, Fairbank JCT. 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*. 2016;16:1221-1230. doi:10.1016/j.spinee.2016.06.010
- Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342:1878-1886. doi:10.1056/NEJM200006223422506

25. Elkan P, Lagerbäck T, Möller TLH, Gerdhem P. Response rate does not affect patient-reported outcome after lumbar discectomy. *Eur Spine J.* 2018;27:1538-1546. doi:10.1007/s00586-018-5541-0
26. Weinstein J, Tosteson TD, Lurie JD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. 2006;296:2441-2450. doi:10.1001/jama.296.20.2441
27. Lønne G, Schoenfeld AJ, Cha TD, Nygaard ØP, Zwart JAH, Solberg T. Variation in selection criteria and approaches to surgery for lumbar spinal stenosis among patients treated in Boston and Norway. *Clin Neurol Neurosurg.* 2017;156:77-82. doi:10.1016/j.clineuro.2017.03.008
28. Lagerbäck T, Fritzell P, Hägg O, Nordvall D, Lønne G. Effectiveness of surgery for sciatica with disc herniation is not substantially affected by differences in surgical incidences among three countries: results from the Danish, Swedish and Norwegian spine registries [published online September 29, 2018]. *Eur Spine J.* doi:10.1007/s00586-018-5768-9
29. Grøvle L, Haugen AJ, Hasvik E, Natvig B, Brox JI, Grotle M. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. *J Clin Epidemiol.* 2014;67:508-515. doi:10.1016/j.jclinepi.2013.12.001
30. Gatchel RJ, Mayer TG. Testing minimal clinically important difference: additional comments and scientific reality testing. *Spine J.* 2010;10:330-332. doi:10.1016/j.spinee.2010.01.019
31. Glassman SD, Carreon LY. Thresholds for health-related quality of life measures: reality testing. *Spine J.* 2010;10:328-329. doi:10.1016/j.spinee.2009.12.026
32. Ferreira ML, Herbert RD, Ferreira PH, et al. A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain. *J Clin Epidemiol.* 2012;65:253-261.
33. Pochon L, Kleinstück FS, Porchet F, Mannion AF. Influence of gender on patient-oriented outcomes in spine surgery. *Eur Spine J.* 2016;25:235-246. doi:10.1007/s00586-015-4062-3
34. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11:163-169. doi:10.1586/erp.11.12
35. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain.* 2008;9:105-121. doi:10.1016/j.jpain.2007.09.005