Eur Spine J (2018) 27:554–563
https://doi.org/10.1007/s00586-017-5331-0

ORIGINAL ARTICLE

The value of patient global assessment in lumbar spine surgery: an evaluation based on more than 90,000 patients

C. Parai[1,2] · O. Hägg[2] · B. Lind[1,2] · H. Brisby[1,3]

Received: 15 June 2017 / Revised: 21 September 2017 / Accepted: 3 October 2017 / Published online: 20 October 2017
© The Author(s) 2017. This article is an open access publication

Abstract
Purpose There are two, principally different ways to obtain patient opinions regarding the outcome of spine surgery: using prospective multi-item questionnaires preoperatively and at follow-up, and using a retrospective single-item question at follow-up—both methods have distinct advantages and limitations. The purpose of the study was to explore the utility of using the simple transition question global assessment, GA, ("How is your back/leg pain today as compared to before the surgery?") as an overall patient-reported outcome measure (PROM) based on the large real-life database in the Swedish spine registry (Swespine).
Methods The correlation between GA and the score-changes and the final scores at 1 year of follow-up for the PROMs VAS, ODI, and EQ-5D was examined. The correlations between GA and item-specific domains within the ODI, EQ-5D and SF-36 as well as the discriminative ability of PROMs with GA as reference criterion were also analysed. The cohort consisted of 94,132 patients registered in Swespine who were surgically treated for disc herniation, spinal stenosis or degenerative disc disease.
Results The correlation coefficients for GA vs. the score-changes were lower than for GA vs final scores. For VAS they ranged for the different diagnosis groups from 0.33 to 0.61 and from 0.50 to 0.79, respectively. For ODI, the corresponding values ranged from 0.43 to 0.65 and 0.63 to 0.76; for the EQ-5D from 0.32 to 0.45 and 0.54 to 0.71. Further, GA showed a somewhat stronger correlation to pain-specific PROMs than to quality-of-life PROMs.
Conclusions The single-item outcome measure global assessment (GA) appears to be a feasible overall patient-reported outcome measure (PROM) and a useful reference for interpreting the scores of patient-reported outcome measures.

Keywords Lumbar surgery · Patient-reported outcome measures · Patient outcome assessment · Global rating scales · Transition question · Validation

Introduction

Currently, there is no objective way to measure the effectiveness of surgery for a degenerative lumbar spine. The dominant indications for this surgery are pain and back-specific dysfunction in relation to pain [1].

Although prospective patient-reported outcome measures (PROMs), completed both preoperatively and at follow-up, are currently accepted as outcome measures in both clinical studies and in clinical practice [2–4], there are limitations and questions regarding the interpretation of the results. Many studies testify to the difficulty associated with determining how large a change in a certain score needs to be before it can be considered relevant for the patient and/or the caregiver [5, 6]. Sometimes it is even difficult to tell whether the score change reflects a true change or if the change

Electronic supplementary material The online version of this article (doi:10.1007/s00586-017-5331-0) contains supplementary material, which is available to authorized users.

✉ C. Parai
catharina.parai@spinecenter.se

[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

simply reflects a measurement error [7]. These extensive questionnaires are also administratively cumbersome for the patient, as well as for the health care system [8]. PROMs are further burdened by bias in the shape of response shift, as described by Schwarz et al. [9]. A response shift is described as a subtle change that concerns the different ways in which patients perceive their health status, and how they interpret and respond to questions, based on cognitive psychological mechanisms. It has also been pointed out that the patients may comprehend the questions listed in the form in different ways, or even consider them irrelevant, which increases the difficulty in interpreting the responses [10].

The frequently used PROMs are validated against some kind of basic reference, a so-called transition question (TQ), where the patient is asked to evaluate the overall result of the surgical intervention [11, 12]. The meaning of this type of single-item PROM (i.e., face validity) instantly becomes obvious to the patient, who then can state directly if the surgery was worthwhile or not. However, over the last decades the performance of single-item PROMs has been called into question mainly because of recall bias [11, 13]. Currently, TQs are recommended as independent measures of construct validation of PROMs designed to measure change over time, and as a strategy to interpret such questionnaires.

The extensive use of retrospective single-item measures indicate their important roles in interpreting the outcome of spinal surgery. Ward et al. [14] demonstrated higher correlations with changes of clinical measures for domain-specific TQs than for general health TQs in a population with rheumatoid arthritis. Since the domain of interest for this type of surgery is pain, it is reasonable to evaluate how a pain-specific TQ works as an overall outcome measure in a population with degenerative spine conditions.

In the Swedish Spine Registry (Swespine) several PROMs, including the pain-specific TQ global assessment (GA) of outcome are reported. The majority of the patients undergoing elective spine surgery in Sweden are registered in this national register and today more than 90,000 individuals operated for degenerative lumbar spine conditions are included in the database.

GA is based on a single question about treatment effect ("How is your back/leg pain today as compared to before the surgery?") with six response options in a Likert type format. It should, however, be considered a five-point asymmetric score, since the first response option—"I had no pain prior to surgery"—is excluded in the analyses. GA has previously been evaluated within a Swedish RCT [15]. It was within the context of that efficacy study found to be a feasible substitute for prospective multi-item PROMs. In the Hägg et al. study the population ($n = 289$) consisted of patients undergoing surgery for degenerative disc disease (DDD). GA was compared with two of the PROMs used in Swespine (VAS and ODI), but not against the remaining EQ-5D and SF-36. Also in that setting GA had four response options as opposed to five in the current study.

The aim of the present study was to explore the utility of using the simple transition question GA as an overall PROM based on the large real-life database in the Swedish spine registry (Swespine). The three most common diagnoses for surgical treatment in the degenerative lumbar spine—spinal stenosis (LSS), disc herniation (LDH), and degenerative disc disease (DDD)—were included in the analyses. Further, we specifically studied the DDD group by ROC curve analysis to be able to compare the results to the previous study by Hägg et al. [15]. In this study, we propose how GA can be used to assess the outcome of spine surgery.

Patients and methods

Characteristics of the population

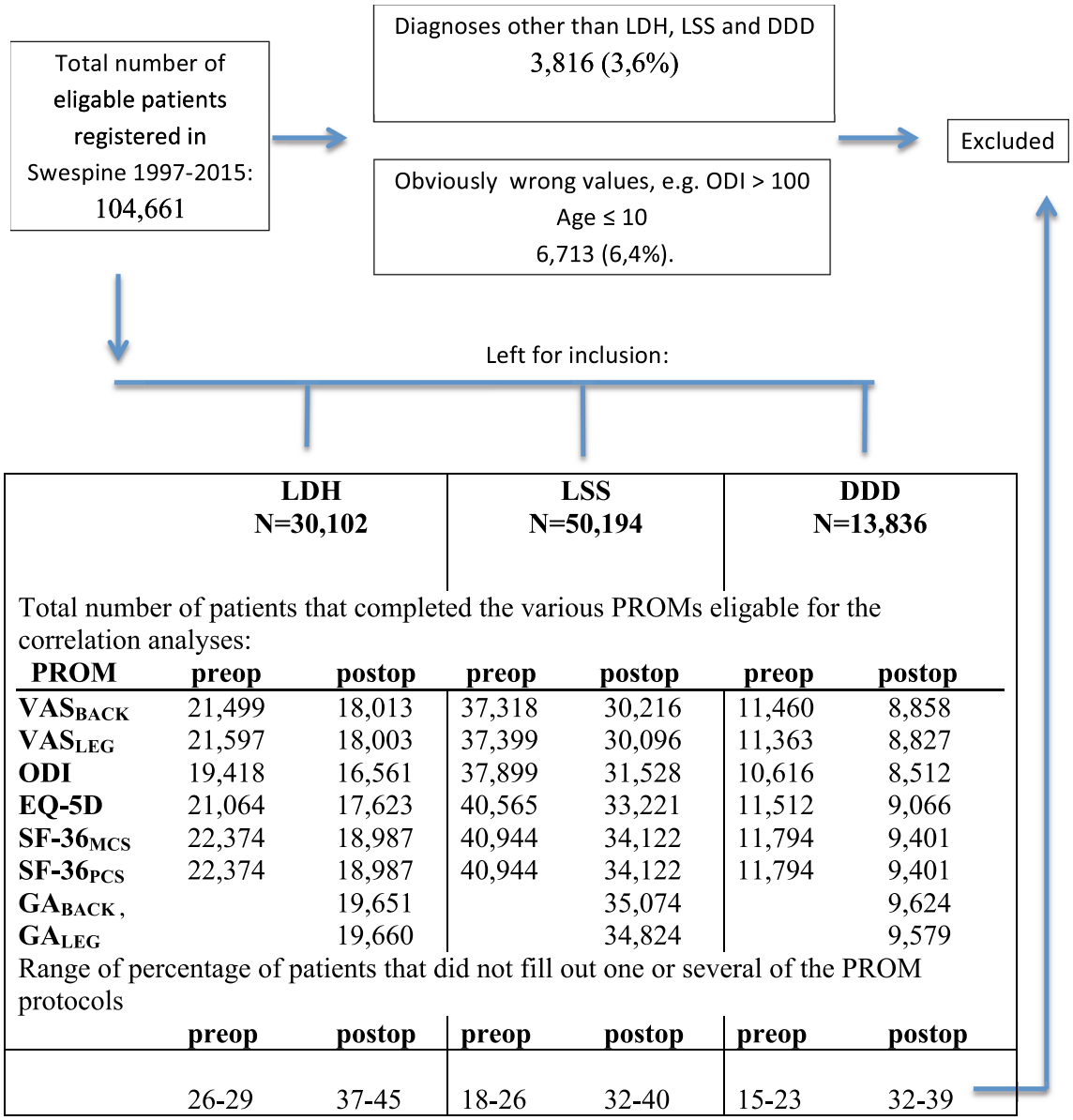
Retrospective analyses were performed on prospectively collected data from 94,132 patients registered in Swespine, between 1997 and 2015, who underwent surgery for degenerative lumbar spinal disorders. In Table 1, we describe the baseline characteristics of the three most common diagnostic groups in Swespine: Lumbar Spinal Stenosis (LSS, $N = 50{,}194$), Lumbar Disc Herniation (LDH, $N = 30{,}102$) and Degenerative Disc Disease (DDD, $N = 13{,}836$). The response rates for the various questionnaires at baseline, was in the range of 71–85%. At 1 year post-surgery the response rates had fallen to 55–68% (Fig. 1).

Swespine is considered to have nationwide coverage. The coverage, i.e., the proportion of units conducting spine surgery that report to Swespine, is approximately 90% [16]. Questionnaires with pre-, as well as post-operative data were completed by the patients after informed consent. The protocol included anthropometrics, co-morbidities, smoking status, usage of analgesics, sick-leave, and PROMs. Follow-up was conducted at 1, 2, 5 and 10 years after surgery and sent to the patients by mail, including one reminder. Reporting of surgical and perioperative data was performed by the surgeon.

Table 1 Baseline characteristics of three common diagnosis groups in Swespine from 1997 to 2015

Patient group	*N*	Mean age ± SD	Females (%)
LDH	30 102	45 ± 14	44
LSS	50 194	67 ± 12	54
DDD	13 836	47 ± 13	53

LDH lumbar disc herniation, *LSS* lumbar spinal stenosis, *DDD* degenerative disc disease



	LDH N=30,102		LSS N=50,194		DDD N=13,836	
Total number of patients that completed the various PROMs eligable for the correlation analyses:						
PROM	**preop**	**postop**	**preop**	**postop**	**preop**	**postop**
VAS_{BACK}	21,499	18,013	37,318	30,216	11,460	8,858
VAS_{LEG}	21,597	18,003	37,399	30,096	11,363	8,827
ODI	19,418	16,561	37,899	31,528	10,616	8,512
EQ-5D	21,064	17,623	40,565	33,221	11,512	9,066
$SF\text{-}36_{MCS}$	22,374	18,987	40,944	34,122	11,794	9,401
$SF\text{-}36_{PCS}$	22,374	18,987	40,944	34,122	11,794	9,401
GA_{BACK},		19,651		35,074		9,624
GA_{LEG}		19,660		34,824		9,579
Range of percentage of patients that did not fill out one or several of the PROM protocols						
	preop	**postop**	**preop**	**postop**	**preop**	**postop**
	26-29	37-45	18-26	32-40	15-23	32-39

Fig. 1 Study flow chart

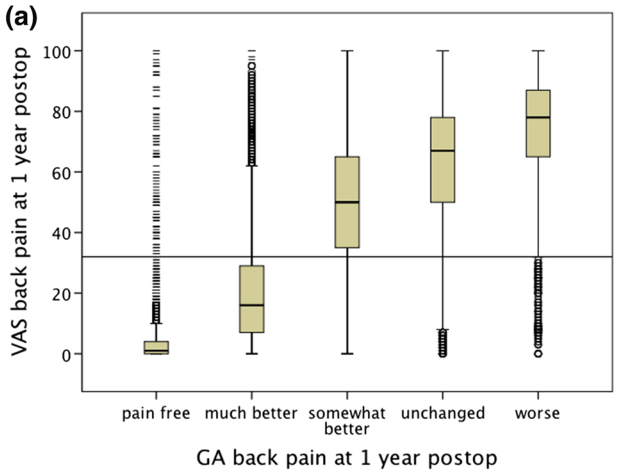
Outcome variables

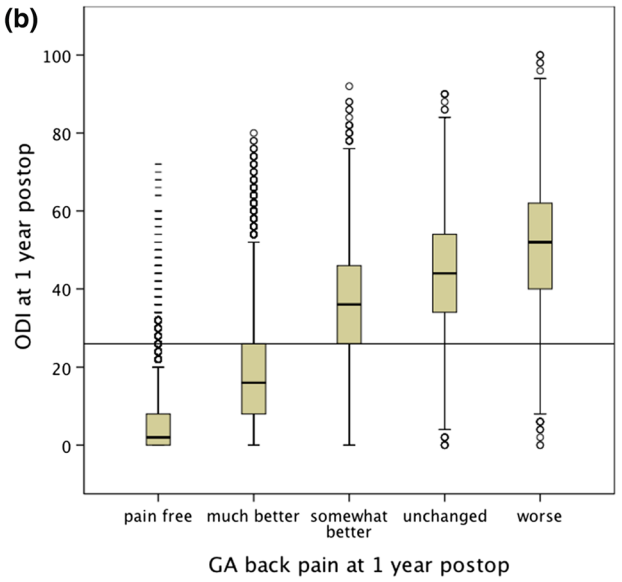
Global assessment (GA)

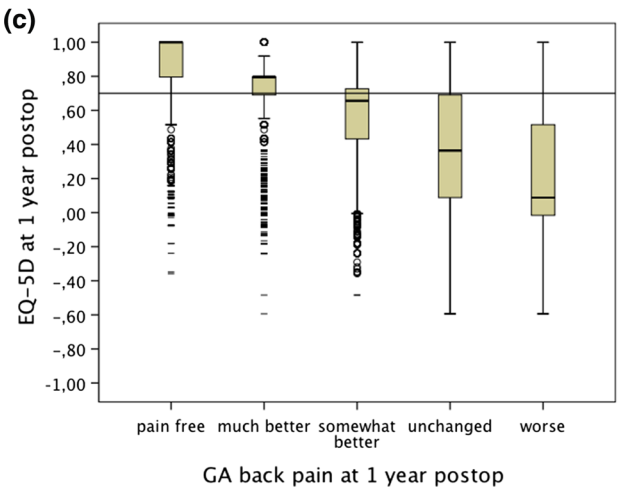
The single-item transition question global assessment for back pain (GA_{BACK}) and leg pain (GA_{LEG}) was used at follow-up: "How is your back/leg pain today as compared to before the surgery?" where 0 represents no back/leg pain before the surgery, 1 completely pain free, 2 much better, 3 somewhat better, 4 unchanged, 5 worse [15].

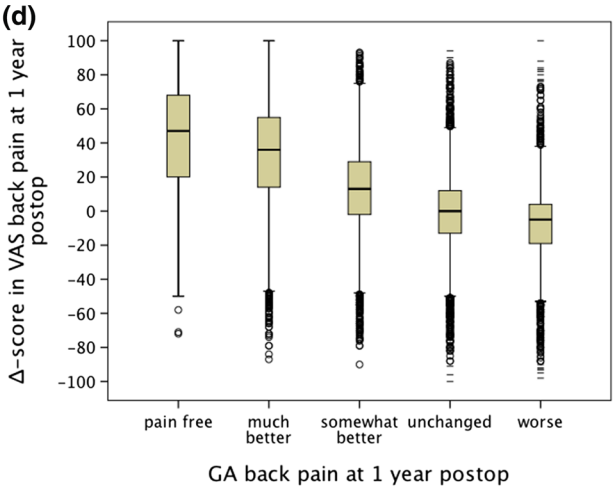
Patient-reported outcome measures (PROMs)

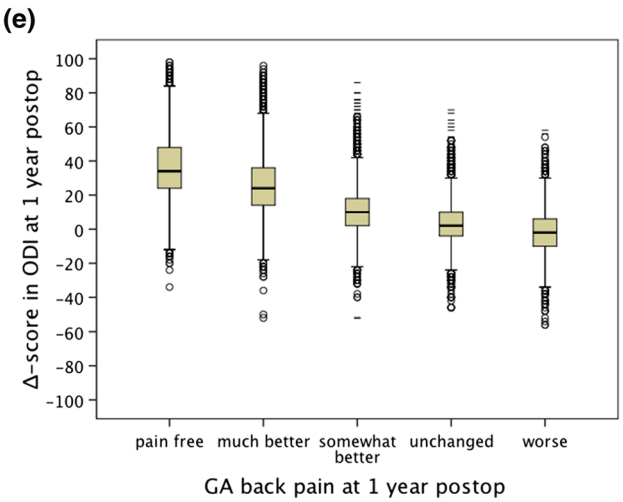
The following PROMs were used: the Visual Analogue Scale for back (VAS_{BACK}) and leg (VAS_{LEG}) pain, measured on a scale from 0 (no pain) to 100 (worst imaginable pain); the disease specific measure Oswestry Disability Index (ODI) [17] which aims to capture important aspects of low back pain that can influence quality of life with the index score ranging from 1 (best) to 100 (worst); the quality-of-life measures EuroqoL-5-dimensions [18] translated into

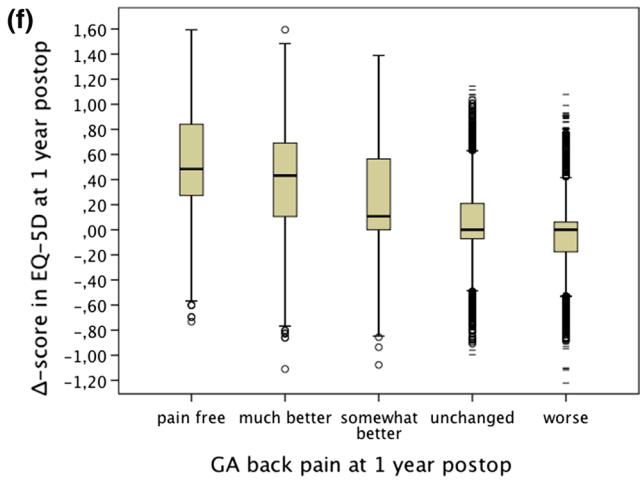












a preference based single index score based on a scale from 0.000 (death) to 1.000 (perfect health) and the Short Form 36 (SF-36) [19] that is subdivided into 8 subscales and 2 composite domains.

Statistics

Descriptive data are presented as mean ± SD, or number (%). The final scores as well as changes in score (Δ-scores)

◂**Fig. 2** Box-plots of the final scores at one year post-operatively (**a**–**c**) and the Δ-scores, i.e. score-changes (**d**–**f**) of PROMs according to the response options in GA_{BACK}. The boxes represent 50% of the study population and the line inside each box is the median. The whiskers represent 1.5-times the interquartile range (IQ). Outliers with more than 1.5-times the IQ are marked as circles and those with 3-times the IQ as lines. The horizontal lines inside the figures **a**–**c** highlights a possible cut-off point between success ("pain free" and "much better") and failure ("somewhat better", "unchanged", "worse") in outcome. Only patients that completed all the PROMs at baseline and at follow-up are included. *N* total = 37,781. N "pain free" = 7061; "much better" = 16,115; "somewhat better" = 7618; "unchanged" = 4078; "worse" = 2909. GA_{BACK} Global Assessment of back pain, VAS_{BACK} Visual Analogue Scale of back pain, *ODI* Oswestry Disability Index, *EQ-5D* Euroqol-5-Dimensions index score

of the prospective PROMs in relation to the different response options in GA are visualised by box-plots.

A non-responder analysis was performed where mean differences in gender, age, and PROMs at baseline were compared between patients undergoing surgery who filled in the PROM questionnaires sent out by the Swespine register and those who did not fill in these forms using the independent samples *t* test.

The behavior of GA in relation to the prospective PROMs was studied by calculations of several Spearman rank correlation coefficients. The correlations between GA and the Δ-scores and the postoperative scores, respectively, for the four PROMs were calculated. The three diagnosis groups LDH, LSS and DDD were analyzed. The correlation coefficient assesses in a general way whether the variables tend to rise (or fall) together. According to Di Fabio [20], the descriptive significance of the magnitude of the absolute value of the coefficient is judged arbitrarily. The recommended interpretation is that a correlation coefficient from 0.00 to 0.24 is of little to none significance. Coefficients from 0.25 to 0.49 are considered weak, from 0.50 to 0.74 moderate and from 0.75 to 1.00 strong. The direction of the correlation between $GA_{BACK/LEG}$ and the absolute score at 1 year follow-up for ODI and $VAS_{BACK/LEG}$ are expected to be positive and for SF-36 and EQ-5D negative. For example, the lower value in postop ODI, the lower value in GA. As for the change scores, negative correlations for all PROMs are expected. The 95% confidence intervals for the correlation coefficient point estimates were previously estimated in the same diagnoses registered until 2014. The widest interval was 0.09 (for the change score of EQ-5D vs GA_{BACK}). The Swespine registry expanded with more than 15,000 patients within the three examined diagnostic groups during 2014–2015. Therefore, an even narrower confidence interval can be expected in the current population. Hence, this analysis is left out. *P* values will not be presented, since they, in these correlation analyses, lack clinical relevance.

ROC curve analyses were used to generate indexes (the area under the curves, AUC) to see how the discriminative ability of GA varied according to different PROMs. GA was dichotomized into 'success' (i.e. the patients that responded as "pain free" or "much better") and 'failure' (i.e. "somewhat better", "unchanged" and "worse"). An AUC below 0.70 was interpreted as low discriminative ability, between 0.70 and 0.90 as good and above 0.90 as excellent [21].

The ROC analyses included only the DDD group. In the ROC analyses the Δ-scores as well as the absolute scores at 1 year post-operatively of VAS, ODI and EQ-5D were analyzed.

The SPSS software package version 24 was used for the statistical analysis.

Ethical considerations

Informed consent was obtained from all participants in Swespine. The use of anonymised data, that cannot be traced to any individual patient, has been approved by the holders of the registry for this research project. This research project was approved by the Regional Ethics Committee.

Results

Non-responder analysis

At the 1-year follow-up differences between responders and non-responders in gender, age, and mean PROM baseline values were significant at the 0.001 level although the differences in values were very small and the standard error of difference did not exceed 0.2 for any of the examined variables. In the non-responder group 49% were female as opposed to 51% in the responder group. The mean difference in age was 2.9 years, and in VAS_{BACK} and VAS_{LEG} 1 point, in ODI 1.5, in EQ-5D 0.043, and in $SF\text{-}36_{PCS}$ and $SF\text{-}36_{MCS}$ 1 point, respectively (additional data are given in Online Resource 1).

Correlations between GA and PROM scores

In Fig. 2a–f, box-plots of the Δ-scores and the final scores for patients that completed all of the PROM-forms at baseline and follow-up are presented according to the response options in GA_{BACK}. In this analysis complete data were available for 37,781 patients. The boxes, representing 50% of the cases, for the response options of "pain free" and "much better" appear to be separated from the other boxes, in the diagrams of the final scores of VAS and ODI, as

demonstrated by the cut-off lines. Separation of the EQ-5D and the Δ-scores is not as evident. The analyses were also made for GA_{LEG}, showing similar results and were, therefore, not presented.

As shown in Table 2, GA presented stronger correlations with the final scores than with the Δ-scores in all three diagnosis groups. The number of responders varied among the PROMs within the range of 12,710–19,660 in the LDH group, 25,507–34,824 in the LSS group, and 7581–9624 in the DDD group. The lower numbers correspond to the Δ-scores where completion of the pre-operative as well as the post-operative PROM questionnaire is needed. GA correlated somewhat stronger with domain-specific PROMs (VAS, ODI) than with the quality-of-life measure EQ-5D. In the LDH group, GA_{LEG} vs. VAS_{LEG} presented the strongest correlations. In both the LSS and the DDD groups, GA_{BACK} vs. VAS_{BACK} yielded the highest correlation coefficients.

Correlations between GA and item-specific domains within PROMs

Since GA seemed to correlate well with the pain-specific VAS, we examined if that finding was consistent with the bodily pain (BP) domain in SF-36 as well as the pain-specific items in EQ-5D and ODI as opposed to the other, less similar domains/items. The analyses were performed on the whole study population. The number of respondents to each domain or item varied within the range of 50,212–65,823.

The SF-36-domain (indicated in bold in Table 3) that had the strongest correlation with GA was Bodily Pain. Similarly in Tables 4 and 5, GA correlated best to the pain-specific items in both the EQ-5D (item 4) and ODI (item 1). The correlations to dissimilar domains or items were weaker.

Table 2 Spearman rank correlations between GA, Δ-scores, and 1 year post-operative scores, for four PROMs in patients who underwent surgery for LDH, LSS and DDD, respectively

PROM	Diagnosis groups					
	LDH $N = 30,102$		LSS $N = 50,194$		DDD $N = 13,836$	
	Δ-score	Post-op	Δ-score	Post-op	Δ-score	Post-op
GA_{BACK} vs						
VAS_{BACK}	−0.33	0.75	−0.45	0.75	−0.61	0.79
VAS_{LEG}	−0.40	0.53	−0.42	0.50	−0.36	0.52
ODI	−0.49	0.70	−0.53	0.67	−0.65	0.76
EQ-5D	−0.39	−0.64	−0.36	−0.60	−0.45	−0.71
GA_{LEG} vs						
VAS_{BACK}	−0.28	0.56	−0.33	0.51	−0.38	0.51
VAS_{LEG}	−0.55	0.79	−0.57	0.75	−0.35	0.79
ODI	−0.47	0.63	−0.47	0.59	−0.43	0.57
EQ-5D	−0.38	−0.60	−0.35	−0.56	−0.32	−0.54

The Spearman rank correlation analysis shows that the global assessment correlates stronger to the postoperative scores of the outcome measures than to the score-changes (Δ)

PROM patient-reported outcome measure, *Δ-score* change in score between pre-operation and 1 year post-operation, *post-op* post-operatively, *LDH* lumbar disc herniation, *LSS* lumbar spinal stenosis, *DDD* degenerative disc disease, $GA_{BACK/LEG}$ Global Assessment for back/leg pain, $VAS_{BACK/LEG}$ Visual Analogue Scale for back/leg pain, *ODI* Oswestry Disability Index, *EQ-5D* Euroqol-5-Dimensions Index score

Table 3 Spearman rank correlations between GA to domains within the SF-36 at 1 year post-operatively

	PF	RP	RE	SF	BP	MH	VT	GH
GA_{BACK}	−0.56	−0.51	−0.40	−0.50	**−0.69**	−0.43	−0.54	−0.49
GA_{LEG}	−0.52	−0.46	−0.37	−0.45	**−0.60**	−0.38	−0.46	−0.44

Analysis on the whole study population, $N = 94,132$

The number of respondents to each domain varied within the range of 63,597–65,823

The Spearman rank correlation analysis shows that GA correlates stronger with Bodily Pain than with the other domains as indicated in bold

$GA_{BACK/LEG}$ Global Assessment for back/leg pain, *SF-36* Short Form-36, *PF* physical functioning, *RP* role limitations due to physical health, *RE* role limitations due to emotional problems, *SF* social functioning, *BP* bodily pain, *MH* general mental health, *VT* vitality, *GH* general health perceptions

Table 4 Spearman rank correlations between GA to specific items of EQ-5D at 1 year post-operatively

	EQ-5D				
	Equation 1	Equation 2	Equation 3	Equation 4	Equation 5
GA_{BACK}	−0.46	−0.27	−0.48	**−0.61**	−0.40
GA_{LEG}	−0.47	−0.24	−0.42	**−0.53**	−0.35

Analysis on the whole study population, $N = 94{,}132$

The Spearman rank correlation analysis shows that GA correlates stronger with the pain-specific item 4, indicated in bold, than with the other items

$GA_{BACK/LEG}$ Global Assessment for back/leg pain, *EQ-5D* Euroqol-5-dimensions. The index of EQ−5D is based on 5 single-item questions that cover the domains mobility (EQ 1), self-care (EQ 2), usual activities (EQ 3), pain/discomfort (EQ 4) and anxiety/depression (EQ 5). The number of respondents to each item varied within the range of 61,299–61,680

Discriminative ability of PROMs with GA_{BACK} as reference criterion

In Fig. 3 the AUC for DDD patients for the VAS_{BACK}, ODI, and EQ-5D are visualized with GA_{BACK} as the reference criterion defining success and failure. Only DDD patients that completed all the PROMs at baseline and at follow-up are included ($n = 7372$). At 1 year post-operatively (Fig. 3a) the AUC for VAS_{BACK} was 0.924 (95% CI 0.918–0.931), for ODI 0.912 (95% CI 0.906–0.919) and for EQ-5D 0.882 (95% CI 0.874–0.890). The 95% confidence intervals indicate that the EQ-5D has a statistically significant lower AUC than the other two outcome measures. When using the Δ-scores, (Fig. 3b) the AUC for VAS_{BACK} was 0.864 (95% CI 0.855–0.873), for ODI 0.877 (95% CI 0.868–0.885) and for the EQ-5D 0.752 (95% CI 0.739–0.764).

Discussion

The results suggest that GA mainly measures and reflects the patient´s pain status, as it correlates well with the condition-specific VAS and ODI, as well as with pain-specific items within the quality-of-life PROMs. This was expected given that GA specifically asks about the change in perceived pain. This finding is coherent with the results of a recent study [14]. As also expected, the degree of association to EQ-5D and domains other than "Bodily Pain" within SF-36, was lower. The directions of the correlations were as postulated.

GA further appears to be an appropriate discriminator between a successful outcome and failure, since there were clear cut-off points in VAS and ODI between patients assessing themselves as "pain free or much better" and "somewhat better, unchanged or worse". This was not seen with EQ-5D. Our results indicate that GA, used as a reference criterion (for instance, in the calculation of minimal important change thresholds) can increase the interpretability of pain-specific PROMs. It is debatable whether GA is applicable as an anchor for generic quality-of-life measures, since it specifically asks about pain. However, GA_{BACK} demonstrated similar correlation coefficients with EQ-5D to those for ODI for all three diagnosis groups. This illustrates the importance of pain as a determinant of quality of life in individuals with degenerative spinal disorders. To increase the accuracy and decrease the bias of a transition question, it is recommended that it should address a specific construct that is anchored to a specific point in time. We believe, that these criteria are fulfilled for GA [22].

Our study further suggests that the respondents—irrespective of diagnosis—base their answers primarily on their present symptom state, and to a lesser degree, reflect on their pain in a longitudinal way, which should be the case if GA worked as an ideal transition question. During the last decades, several reports have been published that support this finding [11, 13, 23]. Two theories that are frequently used to explain this phenomenon are recall bias, as described by Ross in his paper on implicit theory of change [24]; and response shift, defined by Sprangers and Schwartz [9]. Consequently, we may question the validity of this type of question. The latter form of bias, i.e., response shift, however, applies to all PROMs and one can argue that neither a PROM that

Table 5 Spearman rank correlations between GA and specific items in ODI at 1 year postoperatively

	ODI, items									
	1	2	3	4	5	6	7	8	9	10
GA_{BACK}	**0.73**	0.55	0.52	0.48	0.51	0.56	0.53	0.53	0.58	0.58
GA_{LEG}	**0.58**	0.46	0.43	0.47	0.43	0.48	0.49	0.47	0.51	0.51

Analysis on the whole study population, $N = 94{,}132$

Spearman rank correlation analysis showing that GA has the strongest correlation with item 1 (numbers in bold) which is the question on pain severity

$GA_{BACK/LEG}$ Global Assessment for back/leg pain, *ODI* Oswestry Disability Index, which is composed of ten sections: 1/pain intensity, 2/personal care, 3/lifting, 4/walking, 5/sitting, 6/standing 7/sleeping, 8/sex life, 9/social life, 10/travelling. The number of respondents to each item varied within the range of 50,212–58,879

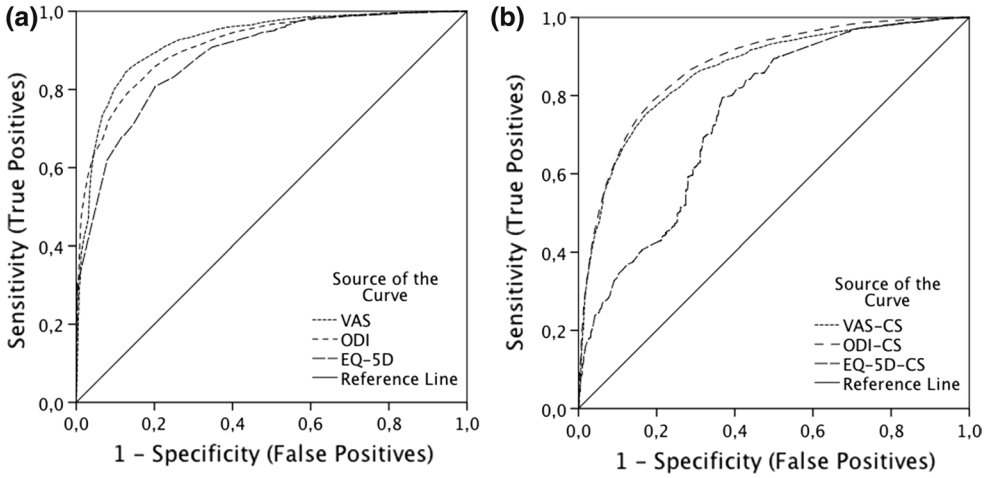


Fig. 3 Receiver operating characteristic (ROC) curve plots for **a** the absolute scores at 1 year post-operation and **b** the score changes of VAS_{BACK}, ODI and EQ-5D with $GA_{SUCCESS/FAILURE}$ as the reference criterion. The analysis was performed on patients diagnosed with degenerative disc disease. *N* in Success group = 4591. *N* in Failure group = 2781. Missing = 6464. *GA* Global Assessment for back pain, *VAS_{BACK}* Visual Analogue Scale for back pain, *ODI* Oswestry Disability Index, *EQ-5D* Euroqol-5-Dimensions Index Score

measures two points in time with an interval of ≥ 1 year, nor a retrospective PROM is an appropriate outcome measure. Due to these limitations the concept of a patient acceptable symptom state, PASS [25] was introduced suggesting that a satisfactory outcome of surgery should not be expressed in terms of change but rather as an absolute score at follow-up above which the patients express an acceptable symptom state. Van Hooff et al. [26] proposed the threshold for such a state for ODI as being 22 in a population of lumbar spine surgery patients. GA could serve as the reference criterion for calculations of acceptable symptom states for VAS and ODI and other condition-specific PROMs. The corresponding thresholds for ODI and VAS_{BACK} using $GA_{BACK/SUCCESS}$ as the reference for the PASS was 27 and 32.5, respectively, in the DDD group. For the whole population the PASS for ODI was also 27 and for VAS_{BACK} 31.5.

The discriminative ability of ODI and VAS_{BACK} with GA_{BACK} as the reference criterion for DDD, tested in ROC curve analyses, was above 90% (interpreted as "excellent") for the post-operative scores and above 86% ("good") for the Δ-scores. Compared to the Hägg et al. study [15] which demonstrated a sensitivity and specificity for the Δ-scores of 75%, the current real-life database analysis strengthened the usefulness of GA as a reference criterion.

There are several limitations to the present study. The percentage of responders 1 year postoperatively in Swespine, and, therefore, in our study, was 55–68%, which would be unacceptable in a randomized controlled study. In registry-based studies numerous non-responders is often the case. In the analyses where comparisons of PROMs are made, completion of all PROMs before as well as after surgery are required. As a consequence, there is a substantial decrease in eligible patients for these analyses and there is, therefore, a risk of selection bias. The results in the non-responder analysis, however, suggest this risk to be low. There is evidence that a high number of missing values do not distort the outcome data in a registry [27]. However, the implications of non-response bias need further investigation.

Were there a Gold standard for the assessment of effectiveness in spine surgery, then of course it would be better to correlate GA to this. The fact that outcome instruments, such as GA, are used as proxies for a Gold standard prevents the use of common psychometric methods for validity testing.

Overall the results indicate that GA, used as a reference criterion, can increase the interpretability of pain-specific PROMs. However, its applicability as an anchor for the quality-of-life PROMs is less obvious. The results on the one hand support the use of GA as an outcome measure of the present state of pain and physical function, while on the other hand demonstrate its limitations as a measure of change.

The findings in this study are exploratory and need to be confirmed in a study that examines potential confounders. Further studies that explore how different variables, such as gender, age, smoking habits, and social status, affect the response pattern of GA are needed. It has also not been explored how GA works in a non-surgically treated population. Since GA appears to have the same function as a PASS, a preliminary threshold for ODI of 27 was suggested, which is rather high in comparison with 22, as demonstrated in the van Hoof et al. study [26]. This difference needs further investigation.

Conclusion

The single-item outcome measure global assessment (GA) shows a stronger correlation to postoperative scores in multi-item patient-reported outcome measures (PROMs) than to

their corresponding change scores and a somewhat stronger correlation to pain-specific PROMs than to quality-of-life PROMs.GA can be an appropriate tool in the measurement of effectiveness of degenerative lumbar spine surgery and may work as a discriminator of success and failure in the interpretation and validation of pain-specific PROMs.

Acknowledgements The authors would like to thank Christer Johansson, independent statistician and fellow at The Royal Statistical Society and the Cochrane Patient Reported Outcome Methods Group. The Neubergh foundation, the Centre for Clinical research, County Council of Varmland, Sweden (FoU), and Region and the Regional Agreement on Medical Training and Clinical Research (ALF) are acknowledged 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. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, de Vet HC (2008) Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 33:90–94. doi:10.1097/BRS.0b013e31815e3a10
2. Chapman JR, Norvell DC, Hermsmeyer JT, Bransford RJ, DeVine J, McGirt MJ, Lee MJ (2011) Evaluating common outcomes for measuring treatment success for chronic low back pain. Spine (Phila Pa 1976) 36:S54–S68. doi:10.1097/BRS.0b013e31822ef74d
3. Werner DAT, Grotle M, Gulati S, Austevoll IM, Lonne G, Nygaard OP, Solberg TK (2017) 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. doi:10.1007/s00586-017-5185-5
4. 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. doi:10.3171/2015.8.spine15476
5. Parker SL, Mendenhall SK, Shau DN, Adogwa O, Anderson WN, Devin CJ, McGirt MJ (2012) Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. J Neurosurg Spine 16:471–478. doi:10.3171/2012.1.spine11842
6. Schwind J, Learman K, O'Halloran B, Showalter C, Cook C (2013) Different minimally important clinical difference (MCID) scores lead to different clinical prediction rules for the Oswestry disability index for the same sample of patients. J Man Manip Ther 21:71–78. doi:10.1179/2042618613y.0000000028
7. Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HC (2009) Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol 62:1062–1067. doi:10.1016/j.jclinepi.2008.10.011
8. Nilsson E, Orwelius L, Kristenson M (2016) Patient-reported outcomes in the Swedish National Quality Registers. J Intern Med 279:141–153. doi:10.1111/joim.12409
9. Schwartz CE, Sprangers MAG (1999) Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med 48:1531–1548. doi:10.1016/S0277-9536(99)00047-7
10. Froud R, Ellard D, Patel S, Eldridge S, Underwood M (2015) Primary outcome measure use in back pain trials may need radical reassessment. BMC Musculoskelet Disord 16:88. doi:10.1186/s12891-015-0534-1
11. Guyatt GH, Norman GR, Juniper EF, Griffith LE (2002) A critical look at transition ratings. J Clin Epidemiol 55:900–908
12. Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H (1999) Capturing the patient's view of change as a clinical outcome measure. JAMA 282:1157–1162
13. Grovle L, Haugen AJ, Hasvik E, Natvig B, Brox JI, Grotle M (2014) Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. J Clin Epidemiol 67:508–515. doi:10.1016/j.jclinepi.2013.12.001
14. Ward MM, Guthrie LC, Alba M (2015) Domain-specific transition questions demonstrated higher validity than global transition questions as anchors for clinically important improvement. J Clin Epidemiol 68:655–661. doi:10.1016/j.jclinepi.2015.01.028
15. 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 (Phila Pa 1976) 27:1213–1222
16. Stromqvist B, Fritzell P, Hagg O, Jonsson B, Sanden B (2013) Swespine: the Swedish spine register: the 2012 report. Eur Spine J 22:953–974. doi:10.1007/s00586-013-2758-9
17. Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry low back pain disability questionnaire. Physiotherapy 66:271–273
18. EuroQol Group (1990) EuroQol–a new facility for the measurement of health-related quality of life. Health Policy 16:199–208
19. Ware JE Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30:473–483
20. Di Fabio RP (2013) Essentials of rehabilitation research: a statistical guide to clinical practice. F.A Davis, Philadelphia
21. Streiner DL, Cairney J (2007) What's under the ROC? An introduction to receiver operating characteristics curves. Can J Psychiatry 52:121–128. doi:10.1177/070674370705200210
22. Kamper SJ, Maher CG, Mackay G (2009) Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 17:163–170. doi:10.1179/jmt.2009.17.3.163
23. Garrison C, Cook C (2012) Clinimetrics corner: the Global Rating of Change Score (GRoC) poorly correlates with functional measures and is not temporally stable. J Man Manip Ther 20:178–181. doi:10.1179/1066981712z.00000000022
24. Ross M (1989) Relation of implicit theories to the construction of personal histories. Psychol Rev 96:341–357. doi:10.1037/0033-295X.96.2.341
25. 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 Rheum Dis 64:34–37. doi:10.1136/ard.2004.023028
26. 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 16:1221–1230. doi:10.1016/j.spinee.2016.06.010
27. 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 82:56–63. doi:10.3109/17453674.2010.548024