Changing values, changing outcomes: the influence of reprioritization response shift on outcome assessment after spine surgery

Carolyn E. Schwartz, Tolulope T. Sajobi, Lisa M. Lix, Brian R. Quaranto & Joel A. Finkelstein

Quality of Life Research
An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation - Official Journal of the International Society of Quality of Life Research

ISSN 0962-9343

Qual Life Res
DOI 10.1007/s11136-013-0377-x

Qual Life Res
DOI 10.1007/s11136-013-0377-x

Changing values, changing outcomes: the influence of reprioritization response shift on outcome assessment after spine surgery

Carolyn E. Schwartz · Tolulope T. Sajobi · Lisa M. Lix · Brian R. Quaranto · Joel A. Finkelstein

Accepted: 19 February 2013

Abstract

Background When patients experience a substantial health state change, they may undergo changes in the underlying meaning of their self-report to standardized outcome measures. These response shifts can reflect changes in the patient's internal standards, values or conceptualization of quality of life. We investigated the presence of changing values (reprioritization response shift) in a longitudinal cohort of spine surgery patients.

Methods Spinal decompression surgery patients (mean age 52 years; 39 % female, 36 % working) provided visual analogue scale (VAS) back and leg pain items, the Short-Form-36 (SF-36v1), and the Oswestry Disability Index (ODI) data pre- and post-surgery ($n_{\text{pre}} = 169$; $n_{\text{6weeks}} = 102$; $n_{\text{3months}} = 106$; $n_{\text{6months}} = 68$). Improved and No-Effect patient groups were compared using the VAS minimally important difference (±15 points) as a cutoff. Reprioritization response shift detection was based on change in the relative importance of the SF-36 domains for group discrimination pre- and post-surgery.

Results The Improved group evidenced significant post-surgery differences from the No-Effect group on bodily pain, general health, physical functioning, social functioning, vitality, and the ODI. The relative importance analysis showed a differential effect with bodily pain ($p < 0.01$) and physical functioning ($p < 0.05$) becoming more important, and role physical ($p < 0.01$) becoming less important post-surgery in distinguishing the Improved group as compared to the No-Effect group. The Improved patients also evidenced stronger associations between bodily pain and physical functioning, vitality and general health ($p < 0.05$). The No-Effect group evidenced increased inter-correlations of bodily pain with social functioning, mental health, and general health ($p < 0.05$).

Conclusions Patients who report clinically significant change in leg and back pain post-surgery using VAS pain scores are also distinguished by increased importance of bodily pain and physical functioning, and decreased importance of role physical. Bodily pain is primarily reflective of physical item response post-surgery among Improved patients, but reflects physical, social, and emotional item response among No-Effect patients. These changes in values may reflect a "moving goal post" in outcome assessment that complicates the interpretation of mean differences over time on standard spine outcome measures.

Electronic supplementary material The online version of this article (doi:10.1007/s11136-013-0377-x) contains supplementary material, which is available to authorized users.

C. E. Schwartz (✉) · B. R. Quaranto
DeltaQuest Foundation, Inc., 31 Mitchell Road, Concord, MA 01742, USA
e-mail: carolyn.schwartz@deltaquest.org

C. E. Schwartz
Departments of Medicine and Orthopaedic Surgery, Tufts University Medical School, Boston, MA, USA

T. T. Sajobi
Department of Community Health Sciences, University of Calgary, 3280 Hospital Drive, NW, Calgary, AB T2N 4Z6, Canada

L. M. Lix
Department of Community Health Sciences, University of Manitoba, 750 Bannatyne Avenue, Winnipeg, MB R3E 0W3, Canada

J. A. Finkelstein
Division of Orthopaedics, Sunnybrook Health Sciences Center and the University of Toronto, 2075 Bayview Avenue, Room MG361, Toronto, ON, Canada

Published online: 22 March 2013

Keywords Response shift · Patient-reported outcomes · Interpretation · Quality of life · Values · Spinal decompression surgery

Abbreviations

DRC	Discriminant ratio coefficient
LPI	Logistic Pratt's index
ODI	Oswestry Disability Index
QOL	Quality of life
R	Relative importance rank
SDFC	Standardized discriminant function coefficient
SF-36v1	Short-Form-36 version 1
SLRC	Standardized logistic regression coefficient
VAS	Visual analogue scale
W	Relative importance weight

Introduction

With the substantial advances in measurement science over the past two decades [1–3], patient-reported outcome measures are increasingly required as metrics for testing and gaining approval for new pharmaceutical and medical device products [4, 5]. Indeed, within therapeutic areas, there is increasing uniformity in outcome data collection so that cross-study comparisons are more feasible and interpretable [6, 7]. This standardization is particularly important in spine surgery outcomes, as advocated by Deyo et al. [8] which made it possible for increased comparability across research studies and for defining clear metrics of clinical significance [9–12]. This heightened sophistication facilitates a stronger evidence-base for spine treatments [13–15].

Despite this important growth in the science and application of patient-reported outcome measures in spinal treatments as well as other chronic diseases, pervasive paradoxical and counterintuitive findings raise questions about what such evaluative rating scales actually assess and how scores should be interpreted: For example, people with chronic illnesses report quality of life (QOL) equal or superior to less severely ill or healthy people [16–18], and in the spine surgery population, it is not uncommon for there to be discrepancies between clinical measures and patients' own ratings [19–21]. These inconsistent findings support the notion of underlying differences in the appraisal of well-being that affect comparisons among individuals and within-persons over time.

At the core of patient-centered research is the imperative to understand individuals' ways of appraising and evaluating their own health and QOL. Aristotle noted that "what constitutes happiness is a matter of dispute …. Very often the same man says different things at different times: when he falls sick he thinks health is happiness, when he is poor, wealth" [22]. These differences in perspective can reflect experience-induced changes in the ways one thinks about symptoms and functional health, in one's values, and the way one conceptualizes QOL. This psychological phenomenon known as *response shift* has been shown to influence estimates for treatment benefit in a range of medical conditions [23, 24] and to undermine the underlying assumptions of psychometric science [25]. While these differing appraisals are indeed the very justification for subjective measures of outcomes as a complement to clinical measures, it is important to utilize patient-reported measures in a way that helps to highlight rather than obfuscate differences in appraisal. Response shift theory and methods provide a means for investigating how health state changes catalyze changes in internal standards, values, and conceptualizations of QOL-related concepts. In the context of spine surgery, the catalyst would be cure of back and leg symptoms as a result of surgery.

Despite the long-standing theoretical acknowledgment that individuals use markedly different personal standards [26–28], frames of reference, and plausible reference points to evaluate their lives [29, 30], assessment strategies generally do not make these underlying criteria explicit. For example, a standard single-item global QOL item asks respondents to anchor their pain level according to the best and worst circumstances imaginable, but these circumstances are never articulated. Similarly, in rating items on multidimensional [31, 32] and domain-specific [30] scales, the criteria individuals consider remain unspecified. Each of these approaches attempts to quantify levels of well-being, while ignoring differences in the "yardsticks" individuals use to evaluate their lives [33]. This work applies a newly developed response shift detection method—the Relative Importance Method [34, 35]—to investigate the presence of changing values—or reprioritization response shift—in a longitudinal cohort of spine surgery patients. This work represents a unique contribution to the response shift literature because it examines how response shift effects are differentially associated with treatment outcomes among a cohort of patients undergoing a strong catalyst, that is, spinal surgery. This response shift detection method has not been used heretofore to examine surgery-related response shift effects.

Methods

Design and eligibility criteria

The study is a longitudinal observational cohort study. All patients were over 18 years of age. There was no upper age exclusion criterion. Patients undergoing elective posterior lumbar spinal decompression surgery for either spinal stenosis at one or two levels or for disk herniation were candidates for recruitment. The clinical indication for surgery was leg-dominant pain without lumbar instability. Patients requiring spinal fusion were excluded. Patients were treated by three fellowship-trained spinal surgeons in the Division of Orthopaedics at Sunnybrook Health Sciences Center, a tertiary care hospital associated with the University of Toronto. Patients were excluded from the study if they were unable to complete the questionnaires in English, if they had visual or cognitive impairment or any other disability that prevented them from completing the outcome measures independently, or if they were unable to give consent for participation.

Procedure

Eligible patients were approached by research personnel not involved in their care, and informed consent was obtained. This study was approved by the research ethics review board at Sunnybrook Health Sciences Center. Participants completed the questionnaires preoperatively and postoperatively at 6 weeks, 3, and 6 months. Preoperative questionnaires were completed in the clinic within 2 weeks prior to surgery and took approximately 20–30 min to complete. After completion, questionnaire booklets were checked for completeness prior to the patient leaving the clinic and the study assistant gathered any missing information.

Measures

Standardized spine outcome measures were collected in this study: (1) the generic Short-Form-36 v1 (SF-36v1) [36] comprising eight domains assessing evaluative functional health, with higher scores reflecting better functional health; (2) two Likert-scaled visual analogue scale (VAS) items measuring back and leg pain on a 100-point scale, with higher scores indicating worse pain [37]; (3) The 10-item disease-specific Oswestry Disability Index (ODI) [36] measuring perceived pain during activities of daily living. The tool asks patients to describe the impact of pain of 10 distinct life domains, using descriptions of observable behaviors (e.g., "Pain prevents me from lifting heavy weights off the floor but I can if they are conveniently positioned, for example on a table."). Other baseline data collected were patient demographics (age and sex), duration of symptoms, employment status (working at present, retired, student, homemaker), and compensation status (currently on disability or compensation), as well as associated comorbid health conditions and other musculoskeletal conditions.

Hypotheses

Our focus was to understand the possible response shifts underlying mean change over time. Past work done by our group has noted that despite the spinal surgery's successfully treating a patients' leg pain, patients express dissatisfaction at the unmasked back pain that remains post-surgery [33]. This clinical observation led to the idea that response shifts in internal standards, values, and conceptualizations of QOL are revealed in these contexts of "partial" versus "total cured" (i.e., reducing leg pain alone versus leg and back pain). We hypothesize that patients who experience clear benefit in both leg and back pain (Improved) as compared to those who do not benefit from surgery (No-Effect) will not differ pre-surgically in terms of the relative importance of QOL domains; however, we expect that these two patient groups will demonstrate value differences over time as a function of surgical outcome. Thus, patients who experience improvement in both leg and back pain will emphasize different domains post-surgery, as compared to patients who do not benefit from surgery.

Statistical methods

A recently proposed *Relative Importance Method* for reprioritization response shift detection [34, 35] was implemented as follows. First, the *study sample was divided into two patient groupings* for response shift hypothesis-testing using a cutoff based on the published minimally important difference (MID; i.e., ±15 points) on the VAS in the spine patient population [12]. This grouping defined patients who were *improved* by the spine surgery (i.e., change on both leg and back VAS greater than the 15-point MID) as compared to those with *no effect* (i.e., change on neither leg nor back VAS greater than the MID). It should be noted that while the primary expectation of posterior lumbar spinal decompression surgery is to cure leg pain and not back pain, there can be improvement in back pain as well, although the primary expectation and goal of surgery is leg pain improvement and not necessarily back pain improvement.

Second, *missing data imputation methods* [38] were utilized to allow inclusion of the maximum number of respondents. In our sample, patients had variable follow-up so imputation methods maximized the analytic sample size. Visual examination of individual patient plots of VAS scores over time revealed that once surgery had taken place, patient scores were relatively stable over the follow-up period. We thus computed the within-patient mean of post-surgery scores as the follow-up score on the patient-reported outcome measures examined.

Third, descriptive analyses for the SF-36v1 domains at pre- and post-surgery occasions were conducted for each level of the grouping variable. We did not utilize the ODI scores in the relative importance analysis for two reasons. First, it does not have subscale or domain scores and thus is not a viable measure for looking at changes in relative importance of domains over time. Second, it is likely more of a perception-based measure rather than an evaluative measure and thus would not be amenable to response shift effects [25]. Schwartz and Rapkin [25] make a distinction between different types of items and how discrepancies between expected and observed scores reflect different measurement concepts. Perception-based items reflect a process that involves judgment where a knowledgeable Other's rating (e.g., spouse or close friend) would be expected to converge with the patient's own rating. Discrepancy in perception would reflect a response bias. In contrast, evaluation-based items reflect a judgment process based on subjective and idiosyncratic criteria; one would not expect knowledgeable Others to give the same answer as the patient because both parties would have unique evaluative criteria. Discrepancy in evaluation for these types of items would reflect response shift.

Mean scores pre- and post-surgery were computed by group for each SF-36v1 domain, and correlations between domain scores were computed within groups to assess potential collinearity and changes in association. The two groups were also compared regarding the magnitude of the inter-domain correlation coefficients pre-surgically, post-surgically, as well as changes in the magnitude of the correlation coefficients over time using Zhou's [39] confidence interval method for testing differences in domain correlations. This method takes into account the dependence between correlations that are repeatedly measured. As noted above, "post-surgically" refers to within-patient mean of post-surgery scores over 6 months of follow-up.

Finally, the relative importance analysis was conducted comparing Improved versus No-Effect groups on the SF-36v1 domains for pre-surgery measurements and the average post-surgery measurements. Relative importance analysis for two-group designs uses discriminant analysis or logistic regression models to estimates weights (i.e., variable coefficients) or ranks (i.e., rank order of the coefficients based on magnitude) for the variables that discriminate between groups [40–47]. This is a marginal analysis, so it does not focus on the importance to the individual respondents but rather on factors distinguishing known groups. Discriminant analysis evaluated domain importance pre- and post-surgery, using the standardized discriminant function coefficients (SDFCs) and the discriminant ratio coefficients (DRCs) for all domains. Logistic regression analysis evaluated domain importance pre- and post-surgery using the standardized logistic regression coefficient (SLRCs) and the logistic Pratt's index (LPI) for all domains. Measures of relative importance based on discriminant analysis provide information on a variable's importance in its contribution to multivariate group separation, while measures of relative importance based on logistic regression analysis provide information about a variable's importance in terms of its independent contribution to group discrimination (i.e., explained variation in group membership). Measures of relative importance based on discriminant analysis (the SDFCs and the DRCs) take into account the collinearity among domain scores, such that high collinearity will result in larger standard error terms and a lower likelihood of statistical significance. In contrast, the logistic regression-based relative importance measures (the SLRC and the LPI) assume that the explanatory variables are independent; these indices focus on explained variance in distinguishing groups. For each of the relative importance measure (i.e., SDFCs, DRCs, SLRCs, or LPI), *reprioritization response shift was considered present* in a domain if the change in signed values of the relative importance weights or relative importance ranks on the domain is statistically significant at a pre-specified α level. Thus, these analyses utilize a model where the outcome is the group membership variable, and the predictors are the domain scores; a significant effect reflects the idea that a domain score shows a significant change in its ability to distinguish the group membership (i.e., the weight or rank changes significantly from pre-surgery).

Results

Descriptive statistics

This longitudinal study focused on patients undergoing elective posterior lumbar spinal decompression surgery for leg and back pain. One hundred seventy-eight patients met inclusion criteria (13 of whom were excluded for language reasons), and 169 individuals (95 %) provided baseline data, 102 patients (60 % of baseline sample) with 6-week follow-up, 106 patients (63 %) with 3-month follow-up data, and 68 patients (40 %) with 6-month follow-up data. The mean age of the sample was 51.96 (SD $= 16.46$, range 20–84), and 39 % of the patients were female. The sample was well-educated, with over half having completed college or a graduate degree. Approximately one-third of the sample was currently working. Most patients reported using pain medication at baseline. Table 1 summarizes the proportion of patients in each group, as well as the number of people excluded from analysis because of missing data on the grouping variable.

A high proportion of our sample evidenced leg pain improvement irrespective of back pain, as shown by mean changes on the VAS items and the ODI items (Sample Mean Change in VAS Leg $= 45.31$, $t = 13.57$, $p \ll 0.0001$; Sample Mean Change in VAS Back $= 31.40$, $t = 8.94$, $p < 0.0001$; Sample Mean Change in

Table 1 Sample characteristics

Variable	N (%)
Time points	
Baseline	169
6-week follow-up	102
3-month follow-up	106
6-month follow-up	68
Grouping variable	
No-Effect	58 (34.3)
Improved	64 (37.9)
Missing	47 (27.8)
Gender	
Gender (% male)	94 (55.6)
Surgical diagnosis	
Disk herniation	64 (61.5)
Spinal stenosis	40 (38.5)
Comorbidities	
Depression	17 (10.1)
Cardiac conditions	12 (7.1)
Diabetes	11 (6.5)
Thyroid conditions	6 (3.5)
Cancer	8 (4.7)
Pulmonary conditions	5 (2.3)
Stroke	1 (0.5)
Peripheral neuropathy	1 (0.5)
Other	23 (13.6)
Education	
Less than high school	13 (8.3)
Graduated from high school or GED	27 (17.2)
Some college or technical school	33 (21.0)
Graduated from college	37 (23.6)
Postgraduate school or degree	47 (29.9)
Employment status at baseline	
Working	55 (35.8)
On leave of absence	14 (11.8)
Unemployed	7 (4.1)
Retired	39 (20.7)
Disabled	14 (11.8)
Homemaker	7 (4.1)
Pain medication use at baseline	
Narcotic	37 (35.6)[a]
NSAIDS or other anti-inflammatory	42 (40.4)[a]
Other pain medications	18 (17.3)[a]
Smoking status: (%) current smoker	51 (30.6)
Variable	**Mean (SD)**
Age	51.96 (16.46)
Range	[20–84]
Baseline outcome scores	
SF-36v1	
Bodily pain	29.35 (18.5)
Physical functioning	35.1 (22.8)
Role physical	22.6 (30.4)
Role emotional	52.64 (41.6)
Social functioning	49.17 (30.1)
Mental health	61.78 (21.9)
Vitality	36.74 (20.7)
General health	67.82 (20.3)
Physical component score	43.61 (12.94)
Mental component score	32.93 (8.24)
ODI	23.00 (9.05)

[a] Sample size for these variables is 104

ODI = 11.29, $t = 5.54$, $p \ll 0.0001$). A comparison of between-group differences pre- and post-surgery suggested that the Improved group evidenced larger changes in bodily pain, physical functioning, social functioning, and vitality on the SF-36v1 and on the ODI, as compared to the No-Effect group (Table 2, Fig. 1).

In addition to these mean differences between groups, we noted differences in the inter-correlations between domains over time as a function of group. Figure 2 illustrates this pattern using the bodily pain domain. The Improved group evidenced a greater increase in the correlation between bodily pain and physical functioning, vitality, and general health ($p < 0.05$ in all cases) from pre-surgery to post-surgery. In contrast, the No-Effect group evidenced a greater increase in the correlation between bodily pain and social functioning, mental health, and general health ($p < 0.05$ in all cases). [Supplemental Tables S1–S4 provide the full inter-correlation matrices.]

The relative importance analyses evaluated more formally whether these inter-relationships among the domains reflect a reprioritization response shift. Specifically, these analyses evaluated whether there were differences in the discriminative performance of the domains, based on their weights or ranks, for the Improved and No-Effect groups. Table 3 shows the discriminant function and logistic regression coefficients' (i.e., weights) and their ranks for each of the SF-36 domains before and after surgery. For the SDFC and SLRC measures, there were statistically significant differences in the relative importance weights for the bodily pain domain over time. Also, there were statistically significant changes in domain ranks for bodily pain and role functioning due to physical health (role physical) between pre- and post-surgery periods for these two relative importance measures. For the DRC and LPI measures, statistically significant change in domain importance was noted for the physical functioning domain. Whereas pre-surgery, bodily pain and physical functioning were of low

Table 2 Descriptive statistics for HRQOL domains by patient group

	Improved	No-Effect	*t* statistic
Bodily pain			
Pre-intervention	25.06 (18.27)	32.10 (19.63)	−2.05*
Post-intervention	65.76 (19.41)	48.99 (26.11)	4.05*
Physical functioning			
Pre-intervention	31.09 (23.22)	35.26 (21.77)	−1.02
Post-intervention	71.27 (20.87)	54.92 (25.08)	3.85*
Role physical			
Pre-intervention	16.70 (26.14)	14.66 (25.48)	0.44
Post-intervention	33.06 (35.51)	22.54 (29.62)	1.78
Role emotional			
Pre-intervention	29.56 (35.48)	27.23 (30.13)	0.39
Post-intervention	41.68 (35.71)	31.92 (33.20)	1.56
Social functioning			
Pre-intervention	42.77 (28.94)	55.17 (31.07)	−2.28*
Post-intervention	80.08 (19.59)	65.91 (27.36)	3.31*
Mental health			
Pre-intervention	57.65 (23.01)	60.82 (21.46)	−0.82
Post-intervention	73.58 (14.73)	68.20 (17.35)	1.94
Vitality			
Pre-intervention	36.14 (22.70)	36.10 (22.10)	0.01
Post-intervention	61.62 (16.23)	51.96 (21.18)	2.86*
General health			
Pre-intervention	67.14 (20.58)	67.06 (21.98)	0.01
Post-intervention	72.47 (18.67)	65.37 (19.92)	2.03
ODI			
Pre-intervention	24.14 (9.46)	22.29 (8.85)	1.04
Post-intervention	7.69 (6.56)	15.04 (9.99)	−4.48****

* Statistically significant difference between "Improved" and "No-Effect" groups

* $p < 0.05$, **** $p < 0.0001$

importance in discriminating Improved and No-Effect groups based on the signed values, and role physical were of high importance, the opposite was true post-surgery.

Discussion

Our results suggest that there were notable differences between Improved and No-Effect groups in mean reported levels of bodily pain, physical functioning, social functioning, and vitality. The relative importance analysis suggests that participants in Improved group were also more likely to evidence reprioritization response shifts in the importance of bodily pain, physical functioning, and role functioning due to physical health domains as compared to those in No-Effect group over the period between pre- and post-surgery occasions. Thus, as a function of a good surgical outcome, these patients changed their values about the importance of various domains of health, relative to the No-Effect group, moving toward emphasizing bodily pain, and physical functioning but de-emphasizing that their role functioning was due to physical health. Other domains such as an appraisal of their role functioning due to emotional health did not change substantially in importance pre- and post-surgery for either group.

To our knowledge, this is a first demonstration of differential response shift effects on patient values after undergoing a relatively strong catalyst that is surgical success or lack thereof. Using the newly proposed application of relative importance methods for reprioritization response shift detection, this study focused on how QOL domain importance changed as a function of surgery outcome, and how these changes distinguish patient groups. In contrast to the univariate mean comparisons shown in Table 2 which ignore any dependencies in the data, these methods based on discriminant analysis consider inter-correlations among domains. The convergence of the two methods in our data suggests that the inter-correlations among domains are not high enough to be problematic and the rank changes explain enough variance to matter. There were also underlying changes in the inter-correlations among SF-36v1 domains that support this inference of differential reprioritization. As an example, the relationship

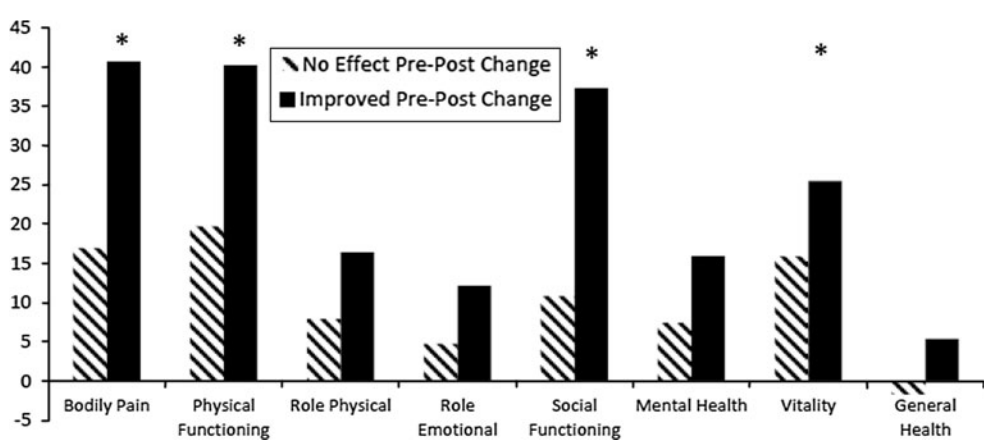

Fig. 1 Mean changes pre- and post-surgery in the Improved and No-Effect groups. There were significant mean differences between the Improved and No-Effect groups on bodily pain, vitality, social functioning, and physical functioning

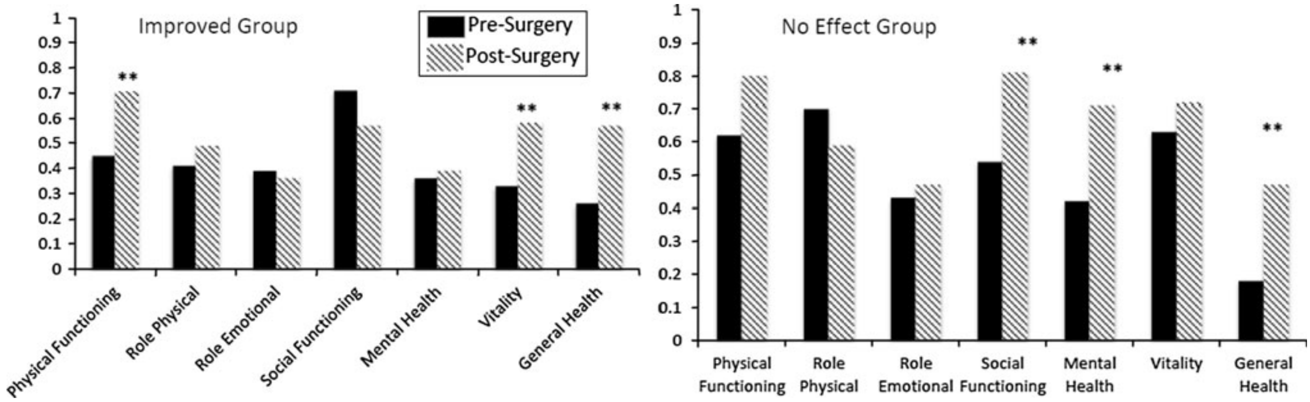

Fig. 2 Bodily pain domain inter-correlations with other SF-36v1 domains pre- and post-surgery. The inter-correlations among SF-36v1 domains were different over time in the Improved and No-Effect Groups. This bar chart illustrates differences in the correlation coefficients for bodily pain and the other SF-36v1 domains by group pre- and post-surgery. The Improved group evidenced change in the inter-correlations of bodily pain with physical functioning, vitality, and general health. The No-Effect group evidenced change in the inter-correlations of bodily pain with social functioning, mental health, and general health. Supplemental Tables S1–S5 provide the full inter-correlation matrices

Table 3 Relative importance analysis results

SF-36v1 domain	Pre-surgery								Post-surgery							
	Discriminant function-based approach				Logistic regression-based approach				Discriminant function-based approach				Logistic regression-based approach			
	SDFC		DRC		SLRC		LPI		SDFC		DRC		SLRC		LPI	
	W	R	W	R	W	R	W	R	W	R	W	R	W	R	W	R
Bodily pain	−0.73	8	0.45	1	−0.24	8	0.44	1	**0.72***	**1****	0.67	1	**0.29***	**1****	0.68	1
Physical functioning	0.02	5	0.00	8	0.00	5	0.00	7	0.36	3	0.33	**2***	0.14	3	0.31	**2***
Role physical	0.53	1	0.07	3	0.18	1	0.07	3	−0.47	**8****	−0.20	8	−0.21	**8****	−0.21	8
Role emotional	0.16	3	0.02	5	0.06	3	0.02	5	0.37	2	0.14	4	0.16	2	0.14	4
Social functioning	−0.61	7	0.41	2	−0.20	7	0.40	2	0.20	4	0.16	3	0.10	4	0.20	3
Mental health	−0.22	6	0.06	4	−0.08	6	0.06	4	−0.22	7	−0.10	7	−0.11	7	−0.12	7
Vitality	0.48	2	0.00	6	0.17	2	0.01	6	0.03	5	0.02	5	0.01	5	0.02	5
General health	0.03	4	0.00	7	0.01	4	0.00	8	−0.03	6	−0.01	6	−0.02	6	−0.02	6

W weight, *R* rank

Statistically significant change in relative importance between pre-surgery and post-surgery: * $\alpha = 0.05$, ** $\alpha = 0.01$

between bodily pain and other domains changed over time differentially between the two groups: The Improved group appeared to reconceptualize bodily pain post-surgery to place a greater emphasis on physical functioning, vitality, and general health. In contrast, the No-Effect group reconceptualized bodily pain post-surgery to emphasize social functioning, mental health, and general health.

Theoretical and clinical implications

These findings suggest that among Improved patients, bodily pain is primarily reflective of physical item response post-surgery, whereas in the No-Effect group, bodily pain appeared to reflect physical, social, *and* emotional item response. This change not only makes sense clinically, but also may reflect a "moving goal post" in spine surgery outcome assessment. These findings suggest that there is a change in the way patients are using the outcome scales pre- and post-surgery. Thus, despite the psychometric standardization of tools used to assess spine surgery outcomes, patients may be using the tools differently as a function of the surgery's effectiveness at reducing their symptom burden. Thus, the theory underlying the QOL measurement model would need to accommodate and consider response shift effects when estimating change over time as well as measurement tool responsiveness. Schwartz and colleagues note that an individual's appraisal processes influence components of a QOL measurement model in distinct ways, with symptom amplification affecting Symptom Experience; goal attainment motivation affecting Symptom Impact and value preferences affecting Global QOL [48]. The present work supports the idea that differential treatment outcomes have reverberating and synergistic effects on the interpretation of change scores. Future research might seek to understand these effects in more depth in other surgical cohorts.

The clinical implications of these findings are also important. First, if an individual uses the same outcome measure differently pre- and post-surgery, it is likely that the MID will shift differentially across patients who experience cure of all symptoms as compared to those whose symptoms are less affected by surgical surgery. The clinician's use of MID should thus be re-evaluated as to its clinical utility. Second, these findings suggest that patients' interpretation of their symptom experience is differentially affected by their surgery outcome. It may be advisable to develop coaching strategies to help manage expectations and recalcitrant symptoms post-surgery among No-Effect patients. Rehabilitation treatments of patients who do not experience surgical success may benefit from exploring the patient's symptom experience, their evaluation of its importance to their functional health and well-being, and their ideas about how to cope with or compensate for symptom burden while retaining activities that are meaningful and life-enhancing. Helping patients to "unpack" bodily pain from social and emotional functioning may be useful for improving their global QOL.

Limitations

The limitations of this work should be acknowledged. First, the period of follow-up is relatively short, and the findings might differ with longer follow-up. Future research should investigate how changes in relative importance influence estimates of mean differences over time. If one were able to adjust for this reprioritization response shift, would the estimate of change over time be larger than the unadjusted comparison? What are the implications on the MID? Do they suggest that we are not seeing significant mean differences in some domains because of changing values? Second, there may be issues of social desirability responding such that patient responses may be influenced by their informed consent and knowledge of the study aims (i.e., they may seek to please the surgeon by reporting better surgical outcomes). This possible response bias would likely lead to an underestimation of response shift effects. We do not, however, believe that social desirability responding would influence group differentiation. Social desirability is a distinct construct from response shift. Third, we note that our findings may be influenced by multiple comparisons. Table 3 shows our study results using a less conservative ($\alpha = 0.05$) and more conservative ($\alpha = 0.01$) nominal significance value. Two of the three statistically significant differences are retained despite the more conservative testing approach. We thus believe that our primary findings are robust. It should be noted that adopting a more conservative alpha can mitigate against Type I errors, but at the expense of a potential loss of power to detect true effects. Given that the sample size is small and treatment effects tend to be small in similar clinical studies, it is advantageous to focus on maximizing power to detect effects. Future research might, nonetheless, replicate this study in an independent sample. Fourth, missing data may lead to some biased estimates of response shift effects, particularly among those patients who were not classifiable as "improved" or "no effect." Future research should examine how missing data patterns influence the estimated prevalence and direction of response shift effects. Finally, the grouping variable for these analyses were created because: (1) patients who do not experience significant reductions in both leg and back pain are dissatisfied with surgical outcomes; (2) it is sometimes hard for patients to distinguish leg and back pain; thus, what may be identified as back pain actually originates in the leg. Supplemental figures show plots by group of baseline levels versus change in leg and back pain, respectively. These plots illustrate that the Improved group is clearly distinguished, whereas the No-Effect group is more "noisy." We thus expect that our findings underestimate the true differences.

In summary, our results suggest that patients who report clinically significant change in leg and back pain using VAS pain scores are also distinguished by an increase in the importance of bodily pain and physical functioning and a decrease in the importance of role functioning due to physical health in the 6 months following spine surgery. Bodily pain appears to be more reflective of physical item response in the Improved group, whereas it appears to reflect physical, social, and emotional item response for the No-Effect group. These findings suggest that underlying response shifts may complicate the interpretation of mean differences over time on standard spine outcome measures.

Acknowledgments We gratefully acknowledge Albert Yee, M.D., and Michael Ford, M.D., for providing access to patients; and Aimee Gallant for her assistance with data collection. Dr. Tolulope Sajobi is supported by an operating grant from the Canadian Institutes of Health Research. Dr. Lisa Lix is supported by a Manitoba Research Chair and previously was supported by a University of Saskatchewan Centennial Chair.

References

1. Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
2. Snyder, C. F., Aaronson, N. K., Choucair, A. K., Elliott, T. E., Greenhalgh, J., Halyard, M. Y., et al. (2011). Implementing

patient-reported outcomes assessment in clinical practice: A review of the options and considerations. *Quality of Life Research*. doi:10.1007/s11136-011-0054-x.

3. Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., et al. (2010). The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology, 63*(11), 1179–1194.
4. FDA. (2009). Guidance for industry. Patient-reported outcome measures: Use in medical product development to support labeling claims. In F. A. D. Administration (Ed.). Washington, DC: U.S. Dept. of Health and Human Services.
5. Sloan, J. A., Halyard, M. Y., Frost, M. H., Dueck, A. C., Teschendorf, B., & Rothman, M. L. (2007). The mayo clinic manuscript series relative to the discussion, dissemination, and operationalization of the Food and Drug Administration guidance on patient-reported outcomes. *Value Health, 10*(Suppl 2), S59–S63. doi:10.1111/j.1524-4733.2007.00267.x.
6. Cleeland, C. S., & Sloan, J. A. (2010). Assessing the symptoms of cancer using patient-reported outcomes (ASCPRO): Searching for standards. *Journal of Pain and Symptom Management, 39*(6), 1077–1085. doi:10.1016/j.jpainsymman.2009.05.025.
7. Bombardier, C. (2000). Outcome assessments in the evaluation of treatment of spinal disorders: Summary and general recommendations. *SPINE, 25*(24), 3100–3103.
8. Deyo, R. A., Battie, M., Beurskens, A. J., Bombardier, C., Croft, P., Koes, B., et al. (1998). Outcome measures for low back pain research. A proposal for standardized use. *Spine (Phila Pa 1976), 23*(18), 2003–2013.
9. Sloan, J. A., Cella, D., & Hays, R. D. (2005). Clinical significance of patient-reported questionnaire data: another step toward consensus. (Editorial). *Journal Clinical Epidemiology, 58*, 1217–1219.
10. Sloan, J. A., Aaronson, N., Cappelleri, J. C., Fairclough, D. L., & Varricchio, C. (2002). Assessing the clinical significance of single items relative to summated scores. *Mayo Clinic Proceedings, 77*(5), 479–487.
11. Sloan, J. A., Cella, D., Frost, M., Guyatt, G. H., Sprangers, M., & Symonds, T. (2002). Assessing clinical significance in measuring oncology patient quality of life: Introduction to the symposium, content overview, and definition of terms. *Mayo Clinic Proceedings, 77*(4), 367–370.
12. Ostelo, R. W., Deyo, R. A., Stratford, P., Waddell, G., Croft, P., Von Korff, M., et al. (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*(1), 90–94. doi:10.1097/BRS.0b013e31815e3a10.
13. Freeman, M. D. (2010). Clinical practice guidelines versus systematic reviews; which serve as the best basis for evidence-based spine medicine? *The Spine Journal, 10*(6), 512–513. doi:10.1016/j.spinee.2010.04.006.
14. Fisher, C. G., Vaccaro, A. R., Mulpuri, K., Angevine, P. D., Thomas, K. C., Patel, A. A., et al. (2011). Evidence-based recommendations for spine surgery. *Spine (Phila Pa 1976)*. doi:10.1097/BRS.0b013e31823dccc0.
15. Vaccaro, A. R., Fisher, C. G., Whang, P. G., Patel, A. A., Prasad, S. K., Angevine, P. D., et al. (2010). Evidence-based recommendations for spine surgery. *Spine (Phila Pa 1976)*. doi:10.1097/BRS.0b013e3181d9391d.
16. Drum, C. E., Horner-Johnson, W., & Krahn, G. L. (2008). Self-rated health and healthy days: examining the "disability paradox". *Disability and Health Journal, 1*(2), 71–78. doi:10.1016/j.dhjo.2008.01.002.
17. Kutner, J. S., Nowels, D. E., Kassner, C. T., Houser, J., Bryant, L. L., & Main, D. S. (2003). Confirmation of the "disability paradox" among hospice patients: Preservation of quality of life despite physical ailments and psychosocial concerns. *Palliative and Supportive Care, 1*(3), 231–237.
18. Albrecht, G. L., & Devlieger, P. J. (1999). The disability paradox: High quality of life against all odds. *Social Science and Medicine, 48*(8), 977–988.
19. Kagawa-Singer, M. (1993). Redefining health: Living with cancer. *Social Science and Medicine, 37*, 295–304.
20. Padilla, G. V., Mishel, M. H., & Grant, M. M. (1992). Uncertainty, appraisal and quality of life. *Quality of Life Research, 1*(3), 155–165.
21. Wilson, I. B. (1999). Clinical understanding and clinical implications of response shift. *Social Science and Medicine, 48*(11), 1577–1588.
22. Aristotle (384-322 BC). *Nicomachean Ethics.*
23. Sprangers, M. A., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. *Social Science and Medicine, 48*(11), 1507–1515.
24. Schwartz, C. E., & Sprangers, M. A. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science and Medicine, 48*(11), 1531–1548.
25. Schwartz, C. E., & Rapkin, B. D. (2004). Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and Quality of Life Outcomes, 2*, 16.
26. Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being: Americans' perceptions of life quality*. New York: Plenum Press.
27. Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542–575.
28. George, L. K. (1979). The happiness syndrome: Methodological and substantive issues in the study of social-psychological well-being in adulthood. *The Gerontologist, 19*, 210–216.
29. George, L. K., & Bearon, L. B. (1980). *Quality of life in older persons: Meaning and measurement*. New York: Human Sciences Press.
30. Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The quality of American life*. New York: Russell Sage Foundation.
31. Lawton, M. P. (1975). The Philadelphia geriatric center morale scale: A revision. *Journal of Gerontology, 30*, 85–89.
32. Neugarten, B. L., Havighurst, R. J., & Tobin, S. S. (1968). Personality patterns of aging. In B. L. Neugarten (Ed.), *Middle age and aging* (pp. 173–177). Chicago: University of Chicago Press.
33. Finkelstein, J. A., Razmjou, H., & Schwartz, C. E. (2009). Response shift and outcome assessment in orthopedic surgery: Is there is a difference between complete vs. partial treatment? *Journal of Clinical Epidemiology, 82*, 1189–1190.
34. Lix, L. M., Sajobi, T. T., Sawatzky, R., Liu, J., Mayo, N. E., Huang, Y., et al. (2012). Relative importance measures for reprioritization response shift. *Quality of Life Research*. doi:10.1007/s11136-012-0198-3.
35. Sajobi, T. T., Lix, L. M., Clara, I., Walker, J., Graff, L. A., Rawsthorne, P., et al. (2012). Measures of relative importance for health-related quality of life. *Quality of Life Research, 21*, 1–11. doi:10.1007/s11136-011-9914-7.
36. Hagg, O., Fritzell, P., & Nordwall, A. (2003). The clinical importance of changes in outcome scores after treatment for chronic low back pain. *European Spine Journal, 12*(1), 12–20. doi:10.1007/s00586-002-0464-0.
37. Gudex, C., Dolan, P., Kind, P., & Williams, A. (1996). Health state valuations from the general public using the visual analogue scale. *Quality of Life Research, 5*(6), 521–531.
38. Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New Jersey: Wiley.
39. Zhou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods, 12*(4), 399–413.
40. Huberty, C. J., & Wisenbaker, J. M. (1992). Variable importance in multivariate group comparisons. *Journal of Educational Statistics, 17*, 75–91.

41. Thomas, D. R. (1992). Interpreting discriminant functions: A data analytic approach. *Multivariate Behavioral Research, 27*, 335–362.
42. Thomas, D. R., & Zumbo, B. D. (1996). Using a measure of variable importance to investigate the standardization of discriminant coefficients. *Journal of Educational and Behavioral Statistics, 21*, 110–130.
43. Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.
44. Menard, S. (2004). Six approaches to calculating standardized logistic regression coefficients. *The American Statistician, 58*, 218–223.
45. Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In T. Pukkila & S. Puntanen (Eds.), *Second international conference in statistics* (pp. 245–260). Tampere, Finland: University of Tampere.
46. Thomas, D. R., Hughes, E., & Zumbo, B. D. (1998). On variable importance in linear regression. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, 45*, 253–275.
47. Thomas, D. R., Zumbo, B. D., Zhu, P., & Dutta, S. (2008). On measuring the relative importance of explanatory variables in a logistic regression. *Journal of Modern Applied Statistical Methods, 7*, 21–38.
48. Schwartz, C. E., Bode, R. K., & Vollmer, T. (2012). The symptom inventory disability-specific short-forms for multiple sclerosis: Reliability and factor structure. *Archives of Physical Medicine and Rehabilitation*. doi:10.1016/j.apmr.2012.03.006.