Clinical Study

Reconsidering the minimally important difference: evidence of instability over time and across groups

Carolyn E. Schwartz, ScD[a,b,*], Jie Zhang, MPH[a], Bruce D. Rapkin, PhD[c], Joel A. Finkelstein, MD, MSc, FRCS(C)[d]

[a] DeltaQuest Foundation, Inc., Concord, MA, USA
[b] Departments of Medicine and Orthopaedic Surgery, Tufts University School of Medicine, Boston, MA, USA
[c] Department of Epidemiology & Population Health, Division of Community Collaboration & Implementation Science, Albert Einstein College of Medicine, Bronx, NY, USA
[d] Division of Orthopedic Surgery, Spine Section Head, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

Abstract

BACKGROUND CONTEXT: Underlying cognitive factors have been found to influence patients' symptom experience. Current evidence suggests that concomitant changes in appraisal must be taken into account to accurately interpret change as measured by standard spine patient-reported outcomes (PROs).

PURPOSE: To investigate changes in patients' minimally important differences (MID) over recovery from spinal surgery; whether and how cognitive appraisal processes are implicated in the change trajectories.

STUDY DESIGN/SETTING: Longitudinal cohort study with up to 12 months follow-up.

PATIENT SAMPLE: Surgical patients (n = 167) with a diagnosis of disc herniation or spinal stenosis.

OUTCOME MEASURES: Standard spine patient-reported PROs were used (Rand-36, Oswestry Disability Index, Numerical Rating Scale for pain, PROMIS Pain Impact).

METHODS: This study was funded by the Feldberg Chair in Spinal Research, Sunnybrook Health Sciences Centre and the authors have no conflicts of interest. MID used an anchor technique and was computed by global assessment of change (GAC) grouping. Participants were binned into groups based on their GAC response patterns at all time points: Consistently better post-surgery, consistently worse post-surgery, and bouncers, whose GAC ratings fluctuate (ie, better-then-worse-then-better; or vice versa). Individuals' longitudinal quality of life (QOL) and appraisal slope scores were computed. QOL-appraisal slopes' correlations were computed by GAC group. Fisher's Z transformation tested the hypothesis that GAC groups differed in the QOL-appraisal relationship over time.

RESULTS: Moderate to large changes are recognized as clinically important in the early stages of recovery (ie, 6 weeks post-surgery), and over time smaller and smaller changes become important. The three pattern groups emphasized and deemphasized different standards of comparison over time, with the Better group emphasizing personal goals and the Worse and Bouncers deemphasizing doctors' input. These group differences translated to differential relationships between PRO change and appraisal changes over time.

CONCLUSIONS: The MID reflects increasingly subtle change over time in PROs. Appraisal may influence how patients experience the same (MID) change over time, with better outcomes associated with emphasizing long-term goals. PRO change seems to be driven by different standards of

---

* Corresponding author. DeltaQuest Foundation, 31 Mitchell Road, Concord, MA 01742, USA.

E-mail address: carolyn.schwartz@deltaquest.org (C.E. Schwartz).

Introduction

The acknowledgement that each patient brings a unique perspective to evaluating patient outcomes has led to increasingly sophisticated methods for quantifying reliability and validity in patient-reported outcomes (PRO) [1,2]. Seminal work on clinical significance [3] has helped to make important distinctions between statistical and meaningful change for longitudinal research and to propose a number of metrics that can be used to facilitate interpretation. The minimally important difference (MID) is one of several such metrics. Defined as the smallest change that patients perceive as important [4,5], the MID helps to guide clinicians and researchers in evaluating patient data over time. Often based on an anchor such as a transition question (ie, a patient-reported global assessment of change), the MID can be used to identify responder groups, to distinguish gradations of response to treatment, or to benchmark one's findings against other clinicians or research studies [6−8]. In the context of spine surgery outcomes, the current standard set of PROs has benefited from numerous studies aimed at quantifying the MID or similar metrics to aid interpretation of longitudinal data [9−12].

Despite these advances, there are concerns related to the MID and similar metrics that they generate large variation in values across studies and across different methods within studies [13]. While some of this variation may be due to measurement error [14,15], it may also be due to psychologically meaningful differences [16,17]. For example, Prospect Theory would predict that people value improvement differently from worsening [18,19]. Indeed, the MID differs in magnitude and direction in patients who gain versus lose functioning [20,21].

These documented anomalies have also been shown to be mediated by response shifts in QOL appraisal, which also may play a role in the vicissitudes of the MID [14,22,23,5]. In other words, when people experience changes in health, they may change their internal standards, values, or conceptualizations of quality of life (QOL) or other target constructs [24−27] While a number of studies have suggested that response-shift effects may be playing a role in the unexpected MID findings [22,23], to date no one has examined differences in underlying cognitive processes [26,28] that distinguish MID-based groups. The present study seeks to investigate if and how MID may change over time in patients after spine surgery, and how patterns of clinical trajectories differ in terms of underlying appraisal processes.

Methods

Sample and design

This longitudinal study included adults with who were recruited from three active spine surgery practices. Eligibility criteria included being over the age of 18, having a diagnosis of a degenerative lumbar spinal disorder of disc herniation, neurogenic claudication, degenerative spondylolisthesis, and lytic spondylolisthesis. Surgery was electively planned. The patient needed to be competent to complete self-report questionnaires, and English-speaking. Patients were recruited consecutively by the clinic research assistant, and study participation was explained. All patients provided written informed consent prior to completing the questionnaires. Data were collected online or by mail at presurgery (two baselines), and at 6-weeks, 3-months, 6-months, and 12-months postsurgery using a secure, Health Information Portability and Accountability Act-compliant interface (www.surveygizmo.com). The study was reviewed and approved by the hospital's Institutional Review Board.

Measures

Standardized PRO measures

The Medical Outcomes Study Short-Form (Rand-36) [29] was used to assess physical and emotional functioning via the physical and mental components scores (PCS and MCS, respectively). Condition-specific disability was measured using the 10-item Oswestry Disability Index (ODI) [30]. Pain impact was measured using four Numeric Rating Scale (NRS) items to assess pain at rest, pain with activity, back pain, and leg pain [31]; and the 6-item PROMIS Pain Impact short-form [32]. To facilitate interpretation of the PROs, all were scored such that high scores reflect better functioning; ie, ODI, NRS, and PROMIS Pain Impact were coded such that higher scores were reflective of better functioning.

Clinically Important Change was assessed using an anchor-based technique, the Global Assessment of Change (GAC) item: "Compared to 3 months ago, how would you rate your health in general now?" Response options ranged from 1 (much worse) to 5 (much better), with a score of 3 indicating same / no change. An option to decline to answer was coded as missing. This item was included in all but the first baseline data collection time points.

The cognitive appraisal processes involved in patient-reported outcome responses were assessed using the *QOL Appraisal Profile* (QOLAP) [33]. This measure has not been applied previously to a spine population. It is a validated measure that provides a descriptive context about how a patient thinks about or appraises his/her QOL. We focused on the eight-item Standards of Comparison domain within QOLAP because research on recovery from surgery as well as in other patient populations has suggested that this is a particularly salient domain of appraisal [34,35]. This domain examines how much the individual compare him/herself to relevant standards: (1) other people they know who are now living with their spinal condition; (2) people whose health does not limit them in any way; (3) the things their doctor told them would happen; (4) their ideal or dream of perfect health; (5) the kind of life they are really working for; (6) the way that the people in their life see them; (7) most people their age; (8) a time in their past before they had their spinal condition. Likert-scaled response options range from all of the time (5) to never (1). This tool is useful for identifying the comparisons that people consider when rating PRO items and whether these standards of comparison change over time [36].

Statistical analysis

t-tests or chi-squared analysis (for continuous or categorical variables, respectively) compared the subsample able to be included in study analysis from those with insufficient follow-up data for inclusion. This comparison allowed us to characterize the selection biases in the study sample.

Estimating clinically important change

The MID was computed as the mean change at each follow-up time point for patients who reported being "somewhat better" on the GAC. (The "somewhat worse" group had too small a sample size to allow robust inference [n = 6 to 10, depending on time point]). The MID was computed with two comparison time points: (a) compared to the last time point, ie, baseline to 6 weeks, or 6 weeks to 3 months, etc. and (b) compared to baseline.

Investigating patterns of change over time

Linear slope scores over time for each PRO scale score and each standards-of-comparison item were computed. These slope scores were created using regression models (DV = PRO or appraisal item; IV = time; computed separately by Study ID; beta coefficients reflected change over time, and were saved for subsequent analyses). We examined individuals' trajectory patterns by binning them into one of three groups based on their GAC response patterns at all time points: Better Post-Surgery, Worse Post-Surgery, and Bouncers (ie, better-then-worse-then-better; or vice versa).Analysis of Variance models were used to compare baseline standards-of-comparison item scores by pattern group. We then computed

correlations between PRO and appraisal slope scores (ie, PRO and appraisal trajectories over time) separately by GAC group. These correlation coefficients were compared using Fisher's Z transformation [37] to test the hypothesis that GAC groups differed in the relationships between QOL-change and appraisal-change over time.

To guide interpretation of mean differences and correlation coefficients, Cohen's published cut-offs were used [38]. Stata Release15 statistical software was used for all analyses [39].

Results

Sample

Table 1 shows the demographic characteristics of the sample. From a study sample of 258 people with baseline data, 167 had sufficient data to be included in the trajectory analysis (ie, data from pre- and postsurgery). The study sample had a mean age of 59 (SD = 16) and an equal proportion of males and females. The sample had an average of 2.4 comorbidities (range 0 to 7), with the most common being back pain, high blood pressure, and osteoarthritis. The majority of the sample had more than a college degree, and about a third of the sample was employed. Less than 5% of the sample reported being on worker's compensation currently or in the recent past.

Surgical indication was leg pain with or without back pain. Forty-nine (29%) had decompression and fusion and 118 (71%) had a decompression alone.

In order to test for selection bias, we compared the baseline demographics of the analytical sample with the patients who had insufficient longitudinal data and were thus excluded from the analysis. These analyses revealed that the analytical sample was more likely than the baseline-data-only sample to include past smokers and was less likely to include people with diabetes or liver disease (Supplemental Table 1).

Estimating clinically important change

Supplemental Table 2 shows the mean change for each PRO comparing each follow-up score to the last follow-up for the Somewhat-Better group as reported on the GAC. The Somewhat-Better group reveals that the MID varies substantially over time for each PRO. The compared-to-last effect sizes are generally largest at 6-weeks postsurgery, and then attenuate over follow-up. In other words, it takes a bigger change to be significant soon after surgery, but thereafter smaller improvements become more noticeable. Supplemental Table 3 shows similar statistics but the comparison is between each follow-up and baseline. In contrast to the compared-to-last mean changes, the compared-to-baseline MID generally grows larger with follow-up for all of the PROs. Thus, improvements build and are cumulative.

Fig. 1 shows these same findings in terms of effect sizes. Whereas the compared-to-last effect sizes are medium to

Table 1
Descriptive baseline statistics of study sample (n = 167)

Variable	Mean	(SD)
Age	58.28	16.98
Range	18−90	
BMI	32	7.03
Range	17.58−56.64	
Total number of comorbidities	2.42	1.36
Range	0−7	
	N	%
Surgical procedure		
Decompression alone	49	29.30
Decompression and instrumented fusion	118	70.70
Gender: N (%)		
Male	83	49.7
Female	83	49.7
Missing	1	0.60
Co-morbidities: N (%)		
Anemia or blood disease	6	3.6
Cancer	9	5.4
Depression	15	9.0
Diabetes	15	9.0
Heart disease	15	9.0
High blood pressure	53	31.7
Kidney disease	1	0.6
Liver disease	1	0.6
Lung disease	8	4.8
Osteoarthritis	43	25.8
Rheumatoid arthritis	17	10.2
Ulcer or stomach disease	8	4.8
Other	50	29.9
Marital status: N (%)		
Married	112	67.1
Widowed	7	4.2
Living with significant other	9	5.4
Single (never married)	19	11.4
Divorced/separated	11	6.6
Civil union/Domestic partner	1	0.6
Missing	8	4.8
Race: N (%)		
American Indian or Alaskan Native	3	1.8
Asian	11	6.6
Black or African-America	1	0.6
Native Hawaiian or Other Pacific Islander	2	1.2
White	144	86.2
Hispanic or Latino	1	0.6
Missing	5	3.0
Education: N (%)		
Less than High School	7	4.2
Graduated From High School or GED	23	13.8
Some College or Technical School	29	17.4
Graduated from College	40	24.0
Postgraduate School or Degree	40	24.0
Missing	28	16.8
Employment status at pre-surgical baseline: N (%)		
Working	55	32.9
On leave of absence	8	4.8
Retired(not due to ill health)	46	27.5
Disabled and/or retired because of ill health	9	5.4
Homemaker	6	3.6
Unemployed	4	2.4
Student	3	1.8
Other	10	6.0
Missing	26	15.6
Smoking status: N (%)		

Table 1 (*Continued*)

Variable	Mean	(SD)
Never smoked/used tobacco	65	38.9
Used to smoke/use tobacco	60	35.9
Current Smoke/use tobacco	12	7.2
Missing	30	18.0
Legal act: N (%)		
Not considering any legal action	135	80.8
My legal action is pending	2	1.2
legal action has been resolved in my	1	0.6
Missing	29	17.4
Worker's compensation status: N (%)		
I am not planning to apply for Workers Compensation	130	77.8
I used to receive it, but do not now	0	0.0
I am planning to apply for it	1	0.6
I have applied for it	1	0.6
I am currently receiving workers compensation	5	3.0
Missing	30	18.0

large at 6 weeks' postsurgery, they attenuate over follow-up to be medium to less-than-small by 3 months' postsurgery and remain at this level through 12 months' postsurgery. In contrast, the compared-to-baseline effect sizes grow to large and remain so over time. This suggests that while patients notice increasingly small changes over postsurgery follow-up, the magnitude of the changes are more pronounced with time.

Investigating patterns of change over time

The study sample eligible for inclusion in the patterns-of-change analysis included 167 people. While about half of the sample reported consistent improvement over follow-up (Better; n = 85), about 23% of the sample reported consistent worsening after surgery (Worse; n = 38) and 26% reported variability in status over time (Bouncers; n = 44).

Baseline differences

Looking at baseline differences among the three groups, an analysis of variance revealed that Bouncers were less likely to compare themselves to their health ideal or to how others see them (F = 5.7 and 3.0; $p \leq .001$ and $p \leq .06$, respectively).

Differences over time

Fig. 2 shows the PRO standardized slope trajectories by group. As expected, the Better group had larger slopes across all PROs, compared to the Worse and Bouncer groups. In particular, the Better group's ODI slope score was notably larger than the other groups, which would be expected for this spine-specific outcome measure. The Worse group had smaller slopes than the other two groups, and the Bouncers' PRO slopes were in between the Better and Worse groups.
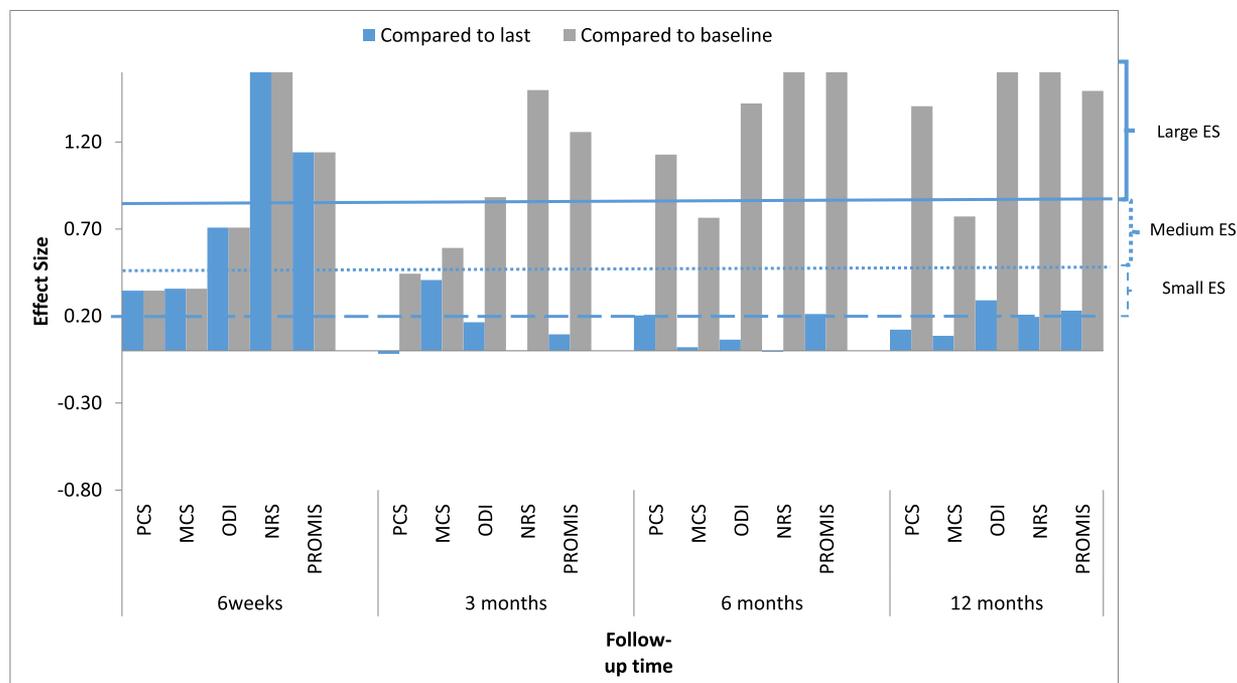
Fig. 1. **Effect sizes (ES) of change over time by comparison time point for somewhat better patients.** The mean difference ES differs by PRO and by comparison. For the compared-to-last comparisons (blue series), the changes are largest at 6-weeks postsurgery, and attenuate over time. In contrast, the ES for compared-to-baseline increases over time and smaller changes become important. Cohen's d-statistic characterizes ES as small (r = 0.20−0.49); medium (r = 0.50−0.79); and large (r ≥ 0.80). (Color version of figure is available online.).

ES, effect-size; PRO, patient-reported outcomes.

Correlations between PRO-change and change in standards of comparison

Fig. 3 shows the standardized slope trajectories by group for the standards-of-comparison appraisal scores. This figure reveals different patterns of appraisal in the three pattern groups. Compared to the other groups, the Better group increasingly emphasized the life they are working for and their ideal health. Compared to the other groups, the Worse group deemphasized the life they are working for and all other standards of comparison except for ideal health. Compared to the other groups, the Bouncer group increasingly de-emphasized the life they were working for, the past before their spinal condition, and what their doctor told them would happen. They increasingly emphasized comparing themselves to others with a spinal condition, how others see them, and ideal health. The Better and Worse groups were similar in their decreased comparison to others their age, whereas the Worse and Bouncers were similar in their decreased emphasis on the life they are working for.

Correlations between PRO- and appraisal-change trajectories differed between pattern groups (Supplemental Table 4). In the Better versus Worse contrasts, the most prominent differences were in how two standards of comparison related to the trajectories for MCS, ODI, and NRS (Supplemental Table 4). For the Worse group, their slower improvement on the ODI and NRS ($\beta = 0.34$ and 0.32, as compared to 0.58 and 0.80 for the Better group) was driven

by focusing more on what their doctor told them (Supplemental Table 5). Their slower improvement on the MCS ($\beta = 0.17$ vs. 0.44) was driven by focusing less on how others see them (Supplemental Table 5).

In the Better versus Bouncer contrasts, the most prominent differences were in how two standards of comparison related to the MCS trajectory (Supplemental Table 4). Both Better and Bouncer groups increasingly compared themselves to others with a spinal condition ($\beta = 0.04$ vs. 0.10, respectively) and decreasingly compared themselves to other people their age ($\beta = -0.11$ vs. -0.01, respectively). The Better group's MCS improvement was driven by focusing less over time on others with a spinal condition (Supplemental Table 5).

In the Worse versus Bouncers contrasts, the most prominent differences were in how three standards of comparison related to the trajectories for MCS, ODI, NRS, and PROMIS (others with a spinal condition, what their doctor told them, and how others see you; Supplemental Table 4). Compared to Worse, Bouncers' decreasing emphasis on what their doctor told them drove their low improvement on the MCS, ODI, NRS, and PROMIS ($\beta = 0.19$, 0.44, 0.54, 50 vs. 0.17, 0.34, 0.32, 0.60). Bouncers' decreasing focus on others with a spinal condition drove their low NRS improvement, and focusing more on how others see you drove their low MCS improvement (Supplemental Table 5). On the other hand, the Worse group's slower improvement on the ODI and NRS
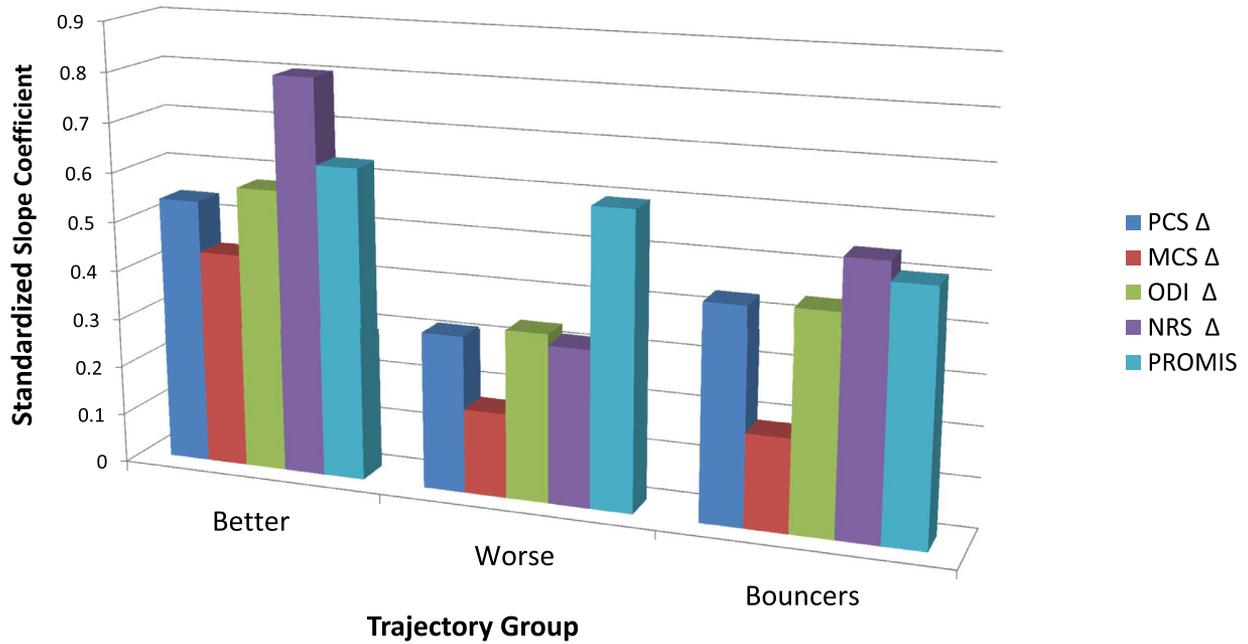
Fig. 2. **PRO slopes by pattern group.** The bar chart illustrates the standardized slope coefficients for each PRO over time by pattern group. The Better group had larger slopes across all PROs and especially for the ODI, compared to the Worse and Bouncer groups. In contrast, the Bouncer group's MCS score suggests that attenuated improvement in emotional functioning may characterize this group.

PRO, patient-reported outcome; ODI, Oswestry disability index.

was driven by focusing more on what their doctor told them (Supplemental Table 5). Their slower improvement on the MCS was driven by focusing less on how others see them (Supplemental Table 5).

For the Better group, their PCS, ODI and PROMIS improvement was driven by a decreasing focus on their past before their spinal condition (Supplemental Table 5). For the Worse group, their slower improvement on the ODI
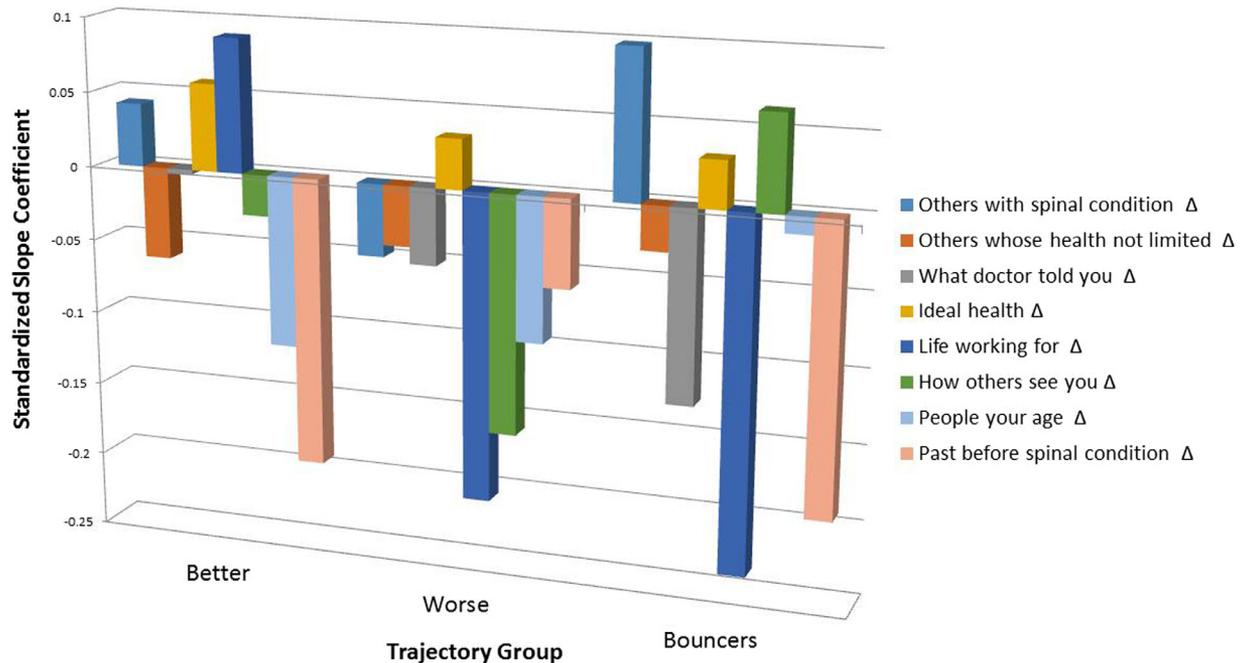

Fig. 3. **Appraisal slopes by pattern group.** The bar chart illustrates the standardized slope coefficients for each standards-of-comparison appraisal score by pattern group. Positive slopes (above the x axis) reflect increasing use of this standard of comparison, whereas negative slopes (below the x axis) reflect its decreasing use. This figure reveals different patterns of emphasis in terms of what patients compare themselves to when appraising their QOL. Thus, Better, Worse, and Bouncer patients are focusing on different standards of comparison, which may influence whether and how they perceive subjectively important change.

and NRS($\beta = 0.34$ and 0.32, as compared to 0.58 and 0.80 for the Better group) was driven by focusing more on what their doctor told them and focusing less on their past before their spinal condition (Supplemental Table 5).

The study cohort had 11(7%) complications that might have an impact on outcome. There was one CSF leak requiring a second surgery to repair. One patient had urinary retention requiring prolonged catheterization and a transurethral resection of the prostate at 3 weeks. There were two cases of adjacent level disease requiring secondary surgery. Five patients had recurrences of disc herniation or return of symptoms requiring a secondary procedure. One patient required prolonged rehabilitation for mobilization. The rate of complications in the Better group was 5/85 (6%), Bouncers 3/44 (7%), and Worse group 2/38 (5%).

Discussion

The present study documents that among spine surgery patients experiencing improvement over time, moderate to large changes are recognized as clinically important in the early stages of recovery (ie, 6 weeks postsurgery), and over time smaller and smaller changes become important. These small changes aggregate to large improvements in comparison to baseline (presurgery) functioning.

While the majority of patients evidenced improvement over time after surgery, about a quarter of the patients had worse outcomes and another quarter reported substantial variability over time. A closer examination of these three pattern groups suggested that they emphasized and deemphasized different standards of comparison over time. Those who improved tended to focus more on the life they are working for, whereas those who declined and Bouncers focused less on the life they are working for. Bouncers also focused more on how others see them, compared to the other two groups. These differences resemble the concept of locus of control [40,41], where the Better group considered internal locus of control concepts (ie, their own goals) and the Worse and Bouncers deemphasized personal goals. The three groups also differed in how much they considered what their doctor told them would happen. Worse and Bouncers de-emphasized their doctor's input, and Betters did not change in this emphasis over time.

These group differences also translated to differential relationships between PRO change and appraisal changes over time. These analyses suggested that Betters' improvements were driven by focusing less on others with a spinal condition and the past before their spinal condition. For the Worse group, lesser improvement on outcomes was driven by focusing less on what their doctor told them and the past before their spinal condition. For Bouncers, deemphasizing what their doctor told them would happen may drive their slower recovery trajectories.

The implications of the present study are that underlying differences in appraisal may influence how patients experience the same (MID) change over time. Depending on what standard(s) of comparison they use, they may experience change in emotional functioning, spine disability, and pain differently. Such appraisal processes may influence expectations of surgery, which have been found to influence preoperative PRO scores as well as postoperative improvement [42]. By identifying whether and which appraisal processes influence PRO trajectories, the present work provides potential avenues for clinical intervention to improve spine surgery outcomes.

For example, it may be worthwhile during the preoperative discussion with the patient to encourage specific standards of comparison to help patients' postsurgical trajectories. One might use a cognitive behavioral therapy [43] approach to help patients change their internal self-talk, to focus more on what they can control and less on what is out of their control. In other words, at each follow-up visit, one might provide and/or remind patients about their expected recovery trajectory (eg, "after 6 weeks, you should be able to do x, y, and z without pain"; "you can now re-introduce x. y. or z into your life") and encourage patients to focus more on long-term goals (ie, the life they are working for) and less on how others see them.

While our study revealed suggestive associations among PRO and appraisal trajectories, it is also worth noting one outcome that showed no differential correlation across pattern groups. PCS trajectories did not show differential correlations with appraisal trajectories across groups. In most cases, PCS and appraisal trajectories had small or negligible correlations. It may be that physical functioning is less subject to changes in standards of comparison because it is more straightforward so one's comparison is more fixed. For example, "It hurts when I bend over" involves different comparisons than "I am not social enough." Because physical functioning is less subjective or evaluative, it would be expected to be less subject to adaptation effects [27]. In contrast, all of the other measures in this standard spine PRO battery had such adaptation effects. This finding provides further support to the recommendation to encourage more adaptive and controllable standards of comparison as part of spine surgery rehabilitation practice.

The overall rate of complications that might impact outcome was 7%. The complication rate was very similar among the three groups (5% to 7%). The similar rate across groups suggests that complications had no effect on our results.

Limitations of our study include selection bias as the analyses indicated that the study sample was more likely to include past smokers. This may suggest that the patients in the sample listened to their doctors better than the attrition sample. It is well known that smoking impairs the body's ability to heal from spine surgery [44], and a common preoperative recommendation is for patients to quit smoking. The study sample was less likely to include people with diabetes or liver disease. Our findings may thus be more representative of a healthier spinal disorders population. Due to very small sample sizes, the Somewhat-Worse patients' MIDs

were not reported because we did not have confidence in the robustness of the findings. It is not unexpected that the Somewhat Worse group was small. Finally, the present analyses focused on one aspect of appraisal, standards of comparison, and did not consider the three other aspects known to play a role in adaptation (ie, frame of reference, sampling of experience, and combinatory algorithm) [26]. Future research might consider one or more of these other aspects of appraisal in longitudinal spine outcomes research.

Conclusions

The present work extends current theory related to clinical significance and interpreting change in PROs. This study demonstrated that the MID changes over 1 year after spine surgery. Whereas soon after surgery their compared-to-last MID is moderate to large, as time goes on smaller and smaller changes become clinically significant. At 12 months postsurgery, these changes would not reach Cohen's threshold for a small ES. In contrast, the MID compared to baseline increases over time, culminating in large ES at 12 months. The present study also revealed that among patients with different trajectories after surgery, the relationship between PRO change, and appraisal change differs by group. Better, Worse, and Bouncer patients are focusing on different standards of comparison over time despite overall similarity in appraisal at baseline. These appraisal trajectory differences may influence whether and how they perceive subjectively important change. Future research might investigate whether such appraisal processes are amenable to intervention, such as cognitive behavioral therapy or other coaching to help patients think in ways that can enhance their recovery and long-term health.

Funding

IRB Approval

Sunnybrook Health Centre Research Ethics Board.

Level of Evidence

Level II

Acknowledgment

Supplementary materials

Supplementary material related to this article can be found at http://dx.doi.org/10.1016/j.spinee.2018.09.010.

References

[1] Embretson SE, Reise SP. Item response theory for psychologists. London: Lawrence Erlbaum Associates; 2000.

[2] Lai JS, Dineen K, Reeve BB, Von RJ, Shervin D, McGuire M, et al. An item response theory-based pain item bank can enhance measurement precision. J Pain Symptom Manage 2005;30:278−88.

[3] Sloan JA, Frost MH, Berzon R, Dueck A, Guyatt G, Moinpour C, et al. The clinical significance of quality of life assessments in oncology: a summary for clinicians. Supportive Care Cancer 2006;14:988−98.

[4] Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clinic Proceedings, 77; 2002. p. 371−83.

[5] King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 2011;11:171−84.

[6] Rai SK, Yazdany J, Fortin PR, Aviña-Zubieta JA. Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. Arthritis Res Ther. 2015;17:143.

[7] Bedard G, Zeng L, Zhang L, Lauzon N, Holden L, Tsao M, et al. Minimal important differences in the EORTC QLQ−C30 in patients with advanced cancer. Asia−Pacific J Clin Oncol. 2014;10:109−17.

[8] Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003;56:395−407.

[9] Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J 2008;8:968−74. doi: 10.1016/j.spinee.2007.11.006.

[10] Djurasovic M, Glassman SD, Dimar JR, Crawford III CH, Bratcher KR, Carreon LY. Changes in the Oswestry disability index that predict improvement after lumbar fusion. J Neurosurg.: Spine 2012;17:486−90.

[11] Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 2008;33:90−4.

[12] Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF−36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Care Res 2001;45:384−91.

[13] Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol 2010;63:524−34.

[14] Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HC. Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol 2009;62:1062−7.

[15] de Vet HC, Terluin B, Knol DL, Roorda LD, Mokkink LB, Ostelo RW, et al. Three ways to quantify uncertainty in individually applied "minimally important change" values. J Clin Epidemiol 2010;63:37−45.

[16] Lee BB, King MT, Simpson JM, Haran MJ, Stockler MR, Marial O, et al. Validity, responsiveness, and minimal important difference for the SF−6D health utility scale in a spinal cord injured population. Value Health 2008;11:680−8.

[17] Nagl M, Farin E. Response shift in quality of life assessment in patients with chronic back pain and chronic ischaemic heart disease. Disability Rehabil 2012;34:671−80.

[18] Kahneman D, Tversky A. Choices, values, and frames. Am Psychol 1984;39:341.

[19] Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Handbook of the fundamentals of financial decision making: Part I 2013: 99–127. World Scientific.

[20] Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. Qual Life Res 2002;11:207–21.

[21] Winter L, Parker B. Current health and preferences for life-prolonging treatments: an application of prospect theory to end-of-life decision making. Soc Sci Med 2007;65:1695–707.

[22] Ousmen A, Conroy T, Guillemin F, Velten M, Jolly D, Mercier M, et al. Impact of the occurrence of a response shift on the determination of the minimal important difference in a health-related quality of life score over time. Health Qual Life Outcomes 2016;14:167.

[23] Kvam AK, Wisløff F, Fayers PM. Minimal important differences and response shift in health-related quality of life; a longitudinal study in patients with multiple myeloma. Health Qual Life Outcomes 2010;8:79.

[24] Sprangers MAG, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999;48:1507–15.

[25] Schwartz CE, Sprangers MAG. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med 1999;48:1531–48.

[26] Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. Health Qual Life Outcomes 2004;2:14.

[27] Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. Health Qual Life Outcomes 2004;2:16.

[28] Schwartz CE, Quaranto BR, Rapkin BD, Healy BC, Vollmer T, Sprangers MAG. Fluctuations in appraisal over time in the context of stable and non-stable health. Qual Life Res 2014;23:9–19. doi: 10.1007/s11136-013-0471-0.

[29] Hays RD, Sherbourne CD, Maze R. The RAND 36-item health survey 1.0. Health Econ 1993;2:217–27.

[30] Fairbank JC, Pynsent PB. The Oswestry disability index. SPINE 2000;25:2940–52.

[31] Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. SPINE 2005;30:1331–4.

[32] Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. Qual Life Res 2009;18:447–60.

[33] Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. Health Qual Life Outcomes 2004;2:14.

[34] Bochner B, Schwartz CE, Garcia I, Goldstein L, Zhang J, Rapkin B. Understanding the impact of radical cystectomy and urinary diversion in patients with bladder cancer: treatment outcomes clarified by appraisal process [abstract]. Qual Life Res 2017;26(Suppl 1):6.

[35] Fayers PM, Langston AL, Robertson C. Implicit self-comparisons against others could bias quality of life assessments. J Clin Epidemiol 2007;60:1034–9.

[36] Schwartz CE, Finkelstein JA, Rapkin BD. Appraisal assessment in patient-reported outcome research: Methods for uncovering the personal context and meaning of quality of life. Qual Life Res 2017;26:545–54. doi: 10.1007/s11136-016-1476-2.

[37] Meng X-L, Rosenthal R, Rubin DB. Comparing correlated correlation coefficients. Psychol Bull 1992;111:172.

[38] Cohen J. A power primer. Psychol Bull 1992;112:155–9.

[39] StataCorp. Stata Statistical Software: Release 15. College Station. TX: StataCorp LLC; 2017.

[40] Strudler Wallston B, Wallston KA. Locus of control and health: a review of the literature. Health Educ Monogr 1978;6:107–17.

[41] Shapiro Jr DH, Schwartz CE, Astin JA. Controlling ourselves, controlling our world: Psychology's role in understanding positive and negative consequences of seeking and gaining control. Am Psychol 1996;51:1213.

[42] Yee A, Adjei N, Do J, Ford M, Finkelstein J. Do patient expectations of spinal surgery relate to functional outcome. Clin Orthop Relat Res 2008;466:1154–61.

[43] Rothbaum, B. O., Meadows, E. A., Resick, P., & Foy, D. W. (2000). Cognitive-behavioral therapy. In E. B. Foa, T. M. Keane, & M. J. Friedman (Eds.), *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies* (pp. 320−325). New York, NY, US: Guilford Press.

[44] Andersen T, Christensen FB, Laursen M, Høy K, Hansen ES, Bünger C. Smoking as a predictor of negative outcome in lumbar spinal fusion. SPINE 2001;26:2623–8.