Published in final edited form as:

Qual Life Res. 2017 August ; 26(8): 2093–2102. doi:10.1007/s11136-017-1560-2.

Changes in quality of life after elective surgery: an observational study comparing two measures.

Vanessa L. Kronzer¹, Michelle R. Jerry², Arbi Ben Abdallah¹, Troy S. Wildes¹, Sherry L. McKinnon¹, Anshuman Sharma¹, and Michael S. Avidan³

¹Department of Anesthesia, Washington University School of Medicine, 660 South Euclid Avenue, Campus Box 8054, Saint Louis, MO, 63110, USA.

²Department of Biostatistics, University of Michigan, 534 Canton Street, Canton, MI, 48188, USA.

³Department of Anesthesia, Washington University School of Medicine, 660 South Euclid Avenue, Campus Box 8054, Saint Louis, MO, 63110, USA.

Introduction:

Assessing change in quality of life from surgery is important for patient decision-making and for healthcare evaluation [1], especially as healthcare becomes more metric-oriented [2]. Many studies use a validated measure to assess change in quality of life after surgery, such as the 36-item Short Form Health Survey (SF-36) [3], the EuroQol Health Survey (EQ-5D) [4], or the Veterans RAND 12-item Health Survey (VR-12) [5]. Another method is to simply ask patients how their quality of life changed after surgery, also called a “global assessment” measure [6, 7]. Both methods seek to measure the same construct, namely, health-related quality of life. Validated measures are more objective because they ask about concrete factors that contribute to quality of life, yet they can take a long time to administer. In contrast, the global measure is short, but subjective.

Comparing the change in a validated measure to a global assessment measure in a large population could help determine whether a simple, self-reported, global quality of life question could be an appropriate surrogate for more lengthy, psychometrically-validated questionnaires. The concern might be that a global measure would compromise granularity and fine discrimination, or contain bias. For example, patients who received kidney transplants gradually overestimated their improvement in quality of life with time [8]. This psychological phenomenon of feeling positive about one’s choices is called “choice-supportive bias” [9]. Another concern is “theory-driven recall bias,” where a patient’s report of well-being is influenced more by theory than by their actual experiences [6]. For example, patients who have had a colostomy reversal remember the colostomy as much more

Corresponding Author: Michael S. Avidan: avidanm@anest.wustl.edu.

Conflicts of interest: The authors declare that they have no conflict of interest.

Informed consent: Informed consent was obtained from all individual participants included in this study.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

miserable than what they reported when they still had the colostomy in place [10]. Because of these cognitive biases, a patient's perceived change in quality of life may be inaccurate—specifically, over-optimistic. One small study has compared the two types of measures. It found a statistically significant correlation between postoperative SF-36 and a global assessment measure, but it did not assess agreement, measure baseline SF-36 scores, or use a standardized time frame related to surgery [7].

Even though few studies have compared validated quality of life batteries to the simple global assessment measure, many have examined change in quality of life for specific surgery types [11, 12]. As a result, the minimum clinically important difference (MCID) for change in quality of life has been established for a range of specific surgical populations (approximately five points) [13, 14], but not for an unselected surgical population. Similarly, few studies have compared the change in quality of life across multiple surgical specialties.

To address these three gaps, we leveraged a large population of unselected patients undergoing elective surgery. Our primary aim was to compare the change in a validated quality of life measure (VR-12) to a global assessment measure. We hypothesized that there would be substantial agreement, with most of the disagreement arising from a higher report of quality of life on the global assessment compared to the VR-12. A secondary aim was to determine the MCID for change in VR-12 quality of life in this population. Finally, we aimed to describe the change in quality of life by surgical specialty.

Methods:

Study Design and Population:

This prospective, observational cohort study received ethics committee approval from the Washington University Human Research Protection Office (ID number 201505035) and is reported in compliance with the STROBE guidelines for observational studies [15]. It was a substudy of the Systematic Assessment and Targeted Improvement of Services Following Yearlong Surgical Outcomes Surveys (SATISFY-SOS) study at Barnes Jewish Hospital [16]. SATISFY-SOS is an ongoing registry that enrolls adult patients undergoing elective surgery who are able to provide written consent during their preoperative assessment visit. Nurses at the preoperative assessment clinic administer a baseline questionnaire to the patients who consent to participate (Online Resource 1). Reasons for non-participation include patient refusal, lack of invitation to participate, and lack of English literacy. Approximately two months after surgery, participants complete a follow-up questionnaire (Online Resource 2). To minimize bias from survey nonresponse, SATISFY-SOS uses an aggressive follow-up sequence consisting of email, two paper mailings, and up to five phone calls per patient.

This sub-study included patients who enrolled in SATISFY-SOS between January 16, 2014 and October 7, 2015 and responded to both the baseline and follow-up questionnaires. The baseline questionnaire response rate was 92%, with insufficient time cited as the main reason for non-completion. The follow-up questionnaire response rate was 57% during the study period. Possible reasons for non-response included patient refusal, change in contact information, or patient death. As a result, 9,097 complete records were available. As specified in the protocol [17], we excluded records where the baseline questionnaire was

completed more than 60 days before surgery (N = 375) or the follow-up questionnaire more than 120 days after surgery (N = 266), leaving 8,456 records eligible for inclusion. We included only the first record for each patient, excluding any subsequent records (N = 554). Therefore, 7,902 patients were included in this study.

Measures

The Veterans RAND 12-Item Health Survey (VR-12) is a validated survey, originally designed for veteran ambulatory populations [18]. Using questions about physical functioning, role limitations due to physical problems, bodily pain, general health perceptions, vitality, social functioning, role limitations due to emotional problems, and mental health, it calculates a physical component summary (PCS) and a mental component summary (MCS) [19]. Both summary scores are continuous on a scale of 0 to 100, where 50 is the general population mean, and higher scores represent better quality of life [20]. For the VR-12, change quality of life was calculated by subtracting the score on the baseline questionnaire from the score on the follow-up questionnaire. The global assessment question, obtained only from the follow-up questionnaire, simply asked, “How would you rate your quality of life **now**?” Choices included “better than before your procedure,” “the same as before your procedure,” and “worse than before your procedure.” To minimize potential bias from priming, the two different instruments were separated on the follow-up questionnaire. That is, the global assessment question was at the beginning (item #1), while the twelve VR-12 items were near the end (items 24 to 35).

History of falls (used for imputation) was also patient-reported on the surveys, and was based on a validated definition [21]. Patients also reported any in-hospital complications on the surveys [22]. Surgical specialty and all other variables were obtained from the medical record using MetaVision® (iMDsoft, Needham, MA).

Statistical Analysis:

Out of the 7,902 included patients, 7.0% were missing at least one question on the baseline VR-12 (1.5% to 3.7% per question), 18.3% were missing at least one question on the follow-up VR-12 (3.3% to 8.9% per question), and 4.9% were missing the global assessment question. Missing data analysis revealed differences in patient characteristics between those with missing data and those with complete data. To minimize bias from this missing data while also limiting the potential inaccuracy of imputation, we used imputed VR-12 values for patients missing just one or two questions out of the twelve. We did not use imputed global assessment values.

We performed imputation in SAS using the discriminant method of fully conditional specification [23], creating 20 imputed datasets. To improve imputation quality, we included the two quality of life measures, demographic variables, and predictors of quality of life in the imputation procedure (all of the variables shown in Table 1 except procedural cardiac risk and postoperative complication). Based on this broad set of demographic and confounder variables, we assumed the missing quality of life data to be at least partially missing at random (MAR). A logistic regression model of the missingness revealed an area under the curve of $c=0.65$. After imputation, 6,980 records had complete quality of life data

and were analyzed. We used un-imputed data to describe patient characteristics, and the first imputed dataset for all other analyses. Sensitivity analyses were performed comparing the results from this dataset to the results from the un-imputed dataset and the other imputed datasets.

We performed all analyses for both the PCS and MCS of the VR-12. To compare VR-12 scores across the three unevenly-sized global assessment groups (better, same, worse), we performed Welch's ANOVA, using a Bonferroni correction for comparisons between groups. For agreement between the PCS or MCS (transformed into ordinal scales using a MCID of five points) and the global assessment (ordinal scale), we used weighted kappa [17, 24], which penalizes disagreements in proportion to their severity. We also calculated percent agreement. For this calculation, patients were considered to "match" if they reported "better" quality of life and had a VR-12 change greater than zero, if they reported "same" and had a VR-12 change between -5 and +5 points, or if they reported "worse" and had a VR-12 change less than zero. These cutoffs were pre-specified in the protocol [17].

We calculated the MCID for this population using two anchor-based approaches, which use the global assessment as the basis for MCID calculation [25]. First, we determined the change in VR-12 where the proportion reporting either "better" or "worse" quality of life was equal to the proportion reporting "same." We refer to this method as the "proportional method." Due to some overlap in this pre-specified proportional method, we also performed a post hoc MCID calculation using a receiver operating characteristic (ROC) curve. Maximal Youden Index determined the optimal cut-off point [26].

For sensitivity analysis, we calculated agreement and MCID using the average of the PCS and MCS scores, defining this average as the "overall" VR-12 quality of life score. Note that this technique has not been validated for primary analyses. All analyses were pre-specified in a published protocol [17]. Statistical significance was set at alpha of 0.05, and clinical significance was set at a five-point change [17]. We performed all analyses with SAS version 9.4 (SAS Institute Inc., Cary, NC), and reported results using 95 percent confidence intervals (CI).

Results

At the time of the postoperative follow-up survey (median 56 days, interquartile range 41 to 73), 61% of patients reported better quality of life on the global assessment, while 10% reported worse quality of life (Table 1). However, by the VR-12 instrument, only 28% and 25% of patients experienced improvement in their PCS and MCS scores, while the identical proportions experienced deterioration. Sensitivity analyses using the average of the PCS and MCS scores, or the "overall" VR-12 score, showed that only 21% experienced improvement in quality of life after surgery, and 20% experienced deterioration. Among those whose overall VR-12 quality of life declined by more than five points, 40% still reported they were "better." In contrast, among those whose overall VR-12 quality of life improved by more than five points, only 2% reported that they were "worse."

The change in VR-12 scores between the better, same, and worse groups was significantly different for both the PCS (2.4, -2.4, -7.5) and MCS (1.2, -0.6, -4.7) scores ($p < 0.001$ for each; Standard Deviations [SD] 10, 8.7, 11, 10, 9.0, and 12, respectively). Agreement between the change in VR-12 quality of life and the patients' global assessment was slight, with a kappa of 0.20 for PCS (95% CI 0.18 to 0.22) and 0.10 for MCS (95% CI 0.08 to 0.12). Directionality of quality of life changes (better, same, worse) matched between the global assessment measure and the PCS in 57% and the MCS in 54% of patients. Finally, the ability of the change in PCS and MCS to discriminate each patient's global assessment response was low, ranging from a c-statistic of 0.54 to 0.64 (Table 3). Sensitivity analyses of the overall VR-12 score produced a slightly improved kappa of 0.27 (95% CI 0.25 to 0.29) and percent match of 64% (Table 2). Overall, the positive predictive value of reporting to be "better" was lower (60%) than the positive predictive value of reporting to be "worse" (80%) (Table 2).

By the proportional method, the MCID for improvement or deterioration in PCS or MCS was approximately five points (Table 3). However, these numbers were difficult to estimate due to overlap between the "worse," "same," and "better" curves on the graphs (Figure 1, Online Resource 3). Nevertheless, post hoc ROC analysis revealed similar MCID values (Table 3). In sensitivity analyses of the overall VR-12 quality of life, the MCID for improvement was 1.25 to 1.5, while the MCID for deterioration was 3.5 to 3.74 (Table 3, Online Resource 4).

Finally, the change in quality of life for the VR-12 and the global assessment varied by surgical specialty (Figure 2). For the global assessment instrument, patients who underwent orthopedic and general surgery reported the highest percentage of "better" quality of life (73% for each). The highest percentage reporting "worse" quality of life occurred in patients who underwent urologic (14%) and gastrointestinal/hepatobiliary surgery (13%). For the change in VR-12, patients who underwent orthopedic surgery again had the greatest overall improvement (+1.8 points; SD 11), but patients receiving neurosurgery showed the second highest overall improvement (+1.0 point; SD 11). Patients receiving urologic and gastrointestinal/hepatobiliary surgery had the greatest decline (-1.3 and -2.0 points; SD 9.2 and 10 respectively).

The characteristics of patients with missing quality of life data differed from those without, including the quality of life scores (appendix). Nevertheless, sensitivity analyses comparing the imputed dataset to the un-imputed and other imputed datasets revealed no major differences in any of the results. A post hoc sensitivity analysis using inverse probability weighting to account for survey nonresponse also showed no differences (Online Resource 5).

Post hoc analyses were also performed to generate MCID values by surgical specialty (appendix) and to examine the subgroup of patients with cancer ($N=971$). Overall, patients with cancer reported worse global quality of life (42% "better" and 18% "worse") and had worse change in VR-12 PCS scores (-4.6, SD 10) compared to the full group, though change in MCS (-0.2, SD 11) was similar. Agreement was lower for PCS (48% match, kappa 0.14), but similar for MCS (55% match, kappa 0.12). This worse agreement on PCS

occurred from a more negative global assessment relative to the patients' VR-12 scores, when compared to the full group. For example, among patients whose PCS improved by more than five points, 80% of the full group reported "better" quality of life, compared to only 54% in the cancer group. Change in quality of life by specialty was similar between the cancer and full groups, though there were some differences, especially in general surgery and neurosurgery (appendix).

Discussion

This study compared the change in VR-12 to a global assessment measure of quality of life approximately two months after elective surgery. As expected, most of the discordance came from a higher global assessment relative to the change in VR-12. For example, 61% of patients reported "better" quality of life on the global assessment, but only 21% showed improvement on the VR-12. Reflecting this inflation in the global assessment, the mean changes in PCS among the "better," "same," and "worse" groups (+2.4, -2.4, -7.5) appear to be shifted downward from zero by approximately 2.3 points. Previous studies in unselected surgical populations have also shown no change in VR-12[5] but overwhelming improvement by global assessment [7]. Thus, these results are likely accurate. It is possible that this optimism results from cognitive biases such as choice-supportive bias [9] or theory-driven recall bias [6], where patients believe they have experienced the improvement that they or others expected.

While some inflation on the global assessment was expected, the low agreement between the VR-12 and global assessment was a novel and surprising finding. The observed agreement for PCS and MCS (kappa 0.20, 0.10; percent match 57%, 54%) was much lower than hypothesized [17], even after combining PCS and MCS. Agreement was slightly better for PCS than MCS, perhaps because patients are able to better appraise their physical than their mental quality of life. Post hoc analyses on the subgroup of patients with cancer showed lower agreement for PCS (0.14) compared to the full group due to more negative global assessments, potentially because many patients might not have been cured of their cancer or were depressed. While a previous study concluded that the global measure was a good approximation of the SF-36, its results actually showed only slight correlation (R-square 0.02 to 0.12) [7]. Due to the poor agreement observed in this study, clinicians cannot assume that one measure reflects the other. Instead, they should choose the most appropriate instrument based on their objectives. While validated measures like the VR-12 are more objective and only ask patients to recall the past four weeks, the global assessment is easier to administer and captures the overall change that patients perceive, arguably the end-goal of any procedure.

Similar to previous studies in both general[28, 29] and surgical populations [13, 30], which found a MCID of approximately five points, we calculated the MCID for PCS to be approximately four and MCS approximately six. While MCIDs are traditionally reported separately for PCS and MCS, the global assessment asks patients to rate their overall quality of life, not the physical or mental components. Therefore, the sensitivity analysis calculating MCID for overall VR-12 quality of life (1.4 points for improvement, 3.6 points for deterioration), despite not being previously validated, may be more accurate. These values

are lower than those for PCS and MCS, likely because the overall VR-12 quality of life is closer to what patients are actually reporting, as shown by the improved c-statistics. The lower MCID for improvement than deterioration is consistent with prior studies [25]. Interestingly, the average of the two (2.5 points) is close to the shifting value described previously (2.3 points), supporting the conclusion that this is the VR-12 change threshold patients can subjectively detect. While specialty-specific MCIDs may be more applicable when available, the MCID of 2.5 points represents a useful standard for populations of patients undergoing multiple surgery types or surgeries where no MCID has been reported.

Finally, on an unadjusted basis, change in quality of life varied by surgical type. On both scales, patients who underwent orthopedic surgery had the greatest improvement in quality of life, perhaps because many of the problems are limited and correctable. The two scales also agreed that patients who underwent urologic and gastrointestinal/hepatobiliary surgery had the most deterioration in quality of life. Reasons for such deterioration might include a greater proportion of cancer and chronic conditions. Alternatively, the recovery time frame for these surgeries might be longer compared to other surgeries, as physical quality of life is known to decrease during recovery [31]. Many studies of change in quality of life after surgery currently use disease-based measures [32–34]. While general quality of life measures are criticized for being less responsive than disease-specific measures [33, 34], they take into account the patient's total experience and facilitate comparisons across specialties [1], as done in this study. Future studies should compare the change in quality of life after controlling for confounders, and should also measure the change in quality of life across different postoperative time periods.

Strengths of this study include its large sample size; multiple, concordant measures of agreement and MCID; comparison of quality of life across multiple surgical specialties, and pre-specified protocol [17]. It also has several limitations. First, generalizability is restricted since participants came from one institution with its own unique case mix. Changing the proportion of patients undergoing surgery in different specialties would likely affect the overall results. Generalizability is also limited by nonparticipation and nonresponse, though previous work shows that participants do not differ from non-participants [35], and responders do not differ significantly different from non-responders [22, 35]. Second, since postoperative quality of life increases with time, surgical specialties with either earlier patient responses or longer recovery times may have had artificially low changes in quality of life. Along similar lines, the VR-12 asks patients to consider the past four weeks, while the global assessment asks about “now,” which may have reduced the change in quality of life for the VR-12. Third, while the VR-12 has been validated in ambulatory care patients with all types of chronic diseases, it has not specifically been validated in every surgical specialty [19]. In addition, measures of agreement separated by PCS and MCS may be biased towards null since the global question asks about quality of life overall. Reassuringly, sensitivity analyses averaging PCS and MCS produced similar results, yet patients may have taken into account other spiritual, emotional, or cultural factors not accounted for by the mental and physical summaries. Another limitation is that anchor-based approaches like the ones used in this study rely on accuracy of patient response on the global assessment question, which has been shown to be potentially biased by this study and others [36, 25]. Nevertheless, including patient input is essential to the definition of MCID, so this remains

the favored approach [37]. Finally, using imputed values artificially decreases p-values. This effect was probably small in this study due to the low number of imputed values and consistency of results in the sensitivity analyses.

In conclusion, perhaps due to patient over-optimism, the two methods of quality of life assessment show poor agreement. Clinicians should select an instrument based on their goals, using 2.5 points as the MCID for the VR-12. Future studies should confirm the difference in quality of life changes across surgical specialties and explore the reasons for such differences.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: VLK was supported by the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH) [grant UL1TR000448, sub award TL1TR000449]. MSA was supported by the National Institute on Aging [grant 1UH2AG050312-01] and the Barnes Jewish Hospital Foundation [grant BJHF#7937-77]. The content is solely the responsibility of the authors and does not necessarily represent the official view of the NIH. The funding sources provided infrastructure and financial support but had no role in the design or conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Appendix:

Characteristics of the patients with complete data vs missing data

Characteristic	*Number (%)		p-value
	Complete (N=5,853)	Missing (N=2,049)	
Age (mean $\pm$ SD)	59 $\pm$ 14	60 $\pm$ 15	0.008
Female Sex	3431 (59)	1177 (57)	0.352
White Race	5270 (92)	1627 (81)	<0.001
BMI (median, IQR)	29 (25,34)	30 (26,34)	0.024
Baseline PCS (median, IQR)	41 (30,52)	38 (28,49)	<0.001
Baseline MCS (median, IQR)	57 (48,62)	55 (46,61)	<0.001
Follow-up PCS (median, IQR)	41 (32,50)	38 (29,47)	<0.001
Follow-up MCS (median, IQR)	57 (48,61)	55 (45,61)	<0.001
Mean PCS Change	-0.2 $\pm$ 10	-4.6 $\pm$ 10	<0.001
Mean MCS Change	0.1 $\pm$ 10	-2.0 $\pm$ 11	0.007
Better quality of life	3535 (60)	1050 (63)	0.115
Same quality of life	1732 (30)	453 (27)	-
Worse Quality of life	586 (10)	159 (10)	-
Charlson Index 3	1060 (18)	426 (21)	0.007
ASA Physical Status 3	2278 (39)	973 (48)	<0.001
Fell in the 6 Months Before Surgery	1438 (25)	557 (29)	<0.001
Fell in the 30 days After Surgery	329 (6)	151 (8)	<0.001
Low Physical Activity (< 4 METs)	1542 (26)	721 (36)	<0.001

Characteristic	*Number (%)		p-value
	Complete (N=5,853)	Missing (N=2,049)	
Chronic Pain	2562 (44)	959 (47)	0.018
Mood Disorder	1183 (20)	418 (20)	0.855
Low Procedural Cardiac Risk	2268 (40)	818 (41)	0.47
Intermediate Procedural Cardiac Risk	3310 (58)	1118 (56)	-
High Procedural Cardiac Risk	131 (2)	46 (2)	-
Postoperative Complication	1381 (24)	874 (43)	<0.001

SD = Standard Deviation; BMI = Body Mass Index; IQR = Inter-Quartile Range; PCS = Physical Component Summary of VR-12; MCS = Mental Component Summary of VR-12; ASA = American Society of Anesthesiologists; METs = Metabolic Equivalents

* Due to different amounts of missing data for each variable, denominator may differ from 7,902

MCID for change in physical and mental quality of life, by specialty*

Specialty	N	PCS, improvement	PCS, deterioration	MCS, improvement	MCS, deterioration
Cardiac	1021	1.1	15	7.0	6.1
General Surgery	414	5.5	9.8	1.1	4.2
GI/Hepatobiliary	488	0.9	8.5	1.9	5.6
Gynecology	712	1.7	11	6.5	5.1
Neurosurgery	526	2.5	18	4.3	7.5
Ophthalmology	519	1.4	1.9	1.0	2.2
Orthopedic	2111	6.2	1.5	6.1	3.9
Otolaryngology	557	1.3	6.3	2.5	4.0
Plastic Surgery	280	2.9	0.02	3.4	7.0
Urology	626	1.4	4.3	6.6	6.8
Other	642	3.5	11	3.7	9.2

MCID = Minimum Clinically Important Difference; PCS = Physical Component Summary of VR-12; MCS = Mental Component Summary of VR-12

* Using ROC method. Note these values may be inaccurate due to small sample sizes in each group

Change in quality of life by specialty in patients with cancer compared to the full group

Specialty	Global Better, All (%)	Global Better, Cancer (%)	Diff.	Global Worse, All (%)	Global Worse, Cancer (%)	Diff.	Change in PCS, All	Change in PCS, Cancer	Diff.	Change in MCS, All	Change in MCS, Cancer	Diff.
Orthopedic	73	66	-7	10	10	0	4.0	3.0	-1.0	-0.3	-1.4	-1.1
General Surgery	73	33	-40	7	17	10	0.5	1.7	1.2	-0.8	-10.3	-9.6
Ophthalmology	71	55	-16	4	15	11	-0.7	-5.2	-4.5	0.3	1.7	1.4
Neurosurgery	67	41	-26	11	23	12	0.7	-5.0	-5.7	1.2	1.9	0.7
Plastic Surgery	60	70	10	8	4	-4	-0.8	-4.9	-4.0	0.0	3.4	3.5

Specialty	Global Better, All (%)	Global Better, Cancer (%)	Diff.	Global Worse, All (%)	Global Worse, Cancer (%)	Diff.	Change in PCS, All	Change in PCS, Cancer	Diff.	Change in MCS, All	Change in MCS, Cancer	Diff.
Gynecology	59	51	-8	6	13	7	-1.0	-2.8	-1.7	0.7	-0.7	-1.4
Otolaryngology	56	37	-19	12	21	9	-2.8	-5.9	-3.1	0.4	0.5	0.1
Cardiac	53	34	-19	12	17	5	-1.0	-5.7	-4.7	0.0	1.2	1.3
GI/Hepatobiliary	52	42	-10	13	22	10	-4.0	-7.7	-3.7	0.0	-0.5	-0.5
Urology	49	39	-10	14	23	8	-2.4	-3.7	-1.3	-0.2	-0.3	-0.1
Other	40	35	-5	11	16	5	-2.2	-4.5	-2.2	0.4	-0.2	-0.7

GI = Gastrointestinal

References:

1. Urbach DR (2005). Measuring quality of life after surgery. *Surgical Innovation*, 12(2), 161–165. [PubMed: 16034507]
2. Lohr KN, & Zebrack BJ (2009). Using patient-reported outcomes in clinical practice: challenges and opportunities. *Quality of Life Research*, 18(1), 99–107. [PubMed: 19034690]
3. Suk KS, Baek JH, Park JO, Kim HS, Lee HM, Kwon JW, et al. (2015). Postoperative quality of life in patients with progressive neuromuscular scoliosis and their parents. *Spine J*, 15(3), 446–453. [PubMed: 25301022]
4. Noyez L (2014). Is quality of life post cardiac surgery overestimated? *Health Qual Life Outcomes*, 12, 62. [PubMed: 24773766]
5. Maillard J, Elia N, Haller CS, Delhumeau C, & Walder B (2015). Preoperative and early postoperative quality of life after major surgery - a prospective observational study. *Health Qual Life Outcomes*, 13, 12. [PubMed: 25649467]
6. Ubel PA, Loewenstein G, Schwarz N, & Smith D (2005). Misimagining the unimaginable: the disability paradox and health care decision making. *Health Psychology*, 24(4 Suppl), S57–62. [PubMed: 16045420]
7. Velanovich V, Younga J, Bhandarkar V, Marshall N, McLaren P, Ritz J, et al. (2012). A single, global patient-centered measure from the SF-36 instrument to assess surgical outcomes and quality of life: a pilot study. *World Journal of Surgery*, 36(9), 2045–2050. [PubMed: 22538393]
8. Smith D, Loewenstein G, Jepson C, Jankovich A, Feldman H, & Ubel P (2008). Mispredicting and misremembering: patients with renal failure overestimate improvements in quality of life after a kidney transplant. *Health Psychology*, 27(5), 653–658. [PubMed: 18823192]
9. Mather M, Shafir E, & Johnson MK (2000). Misremembrance of options past: source monitoring and choice. *Psychological Science*, 11(2), 132–138. [PubMed: 11273420]
10. Smith DM, Sherriff RL, Damschroder L, Loewenstein G, & Ubel PA (2006). Misremembering colostomies? Former patients give lower utility ratings than do current patients. *Health Psychology*, 25(6), 688–695. [PubMed: 17100497]
11. Mangione CM, Goldman L, Orav EJ, Marcantonio ER, Pedan A, Ludwig LE, et al. (1997). Health-related quality of life after elective surgery: measurement of longitudinal changes. *Journal of General Internal Medicine*, 12(11), 686–697. [PubMed: 9383137]
12. Turel MK, Thakar S, & Rajshekhar V (2015). Quality of life following surgery for large and giant vestibular schwannomas: a prospective study. *Journal of Neurosurgery*, 122(2), 303–311. [PubMed: 25479119]
13. Carreon LY, Bratcher KR, Canan CE, Burke LO, Djurasovic M, & Glassman SD (2013). Differentiating minimum clinically important difference for primary and revision lumbar fusion surgeries. *Journal of Neurosurgery: Spine*, 18(1), 102–106. [PubMed: 23157276]
14. Zhou F, Zhang Y, Sun Y, Zhang F, Pan S, & Liu Z (2015). Assessment of the minimum clinically important difference in neurological function and quality of life after surgery in cervical

- spondylotic myelopathy patients: a prospective cohort study. *European Spine Journal*, 24(12), 2918–2923. [PubMed: 26324283]
15. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, & Vandenbroucke JP (2014). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *International Journal of Surgery (London, England)*, 12(12), 1495–1499.
 16. Avidan MS ([cited 2016 2 10]). Systematic Assessment and Targeted Improvement of Services Following Yearlong Surgical Outcomes Surveys (SATISFY-SOS). In: ClinicalTrials.gov.
 17. Kronzer VL, Jerry MR, & Avidan MS (2016). Assessing change in patient-reported quality of life after elective surgery: protocol for an observational comparison study. *F1000Research*, 5, 976. [PubMed: 27635222]
 18. Kazis LE, Miller DR, Skinner KM, Lee A, Ren XS, Clark JA, et al. (2006). Applications of methodologies of the Veterans Health Study in the VA healthcare system: conclusions and summary. *Journal of Ambulatory Care Management*, 29(2), 182–188. [PubMed: 16552327]
 19. Kazis LE, Selim A, Rogers W, Ren XS, Lee A, & Miller DR (2006). Dissemination of methods and results from the veterans health study: final comments and implications for future monitoring strategies within and outside the veterans healthcare system. *Journal of Ambulatory Care Management*, 29(4), 310–319. [PubMed: 16985389]
 20. Selim AJ, Rogers W, Fleishman JA, Qian SX, Fincke BG, Rothendler JA, et al. (2009). Updated U.S. population standard for the Veterans RAND 12-item Health Survey (VR-12). *Quality of Life Research*, 18(1), 43–52. [PubMed: 19051059]
 21. Lamb SE, Jorstad-Stein EC, Hauer K, & Becker C (2005). Development of a common outcome data set for fall injury prevention trials: the Prevention of Falls Network Europe consensus. *Journal of the American Geriatrics Society*, 53(9), 1618–1622. [PubMed: 16137297]
 22. Kronzer VL, Jerry MR, Ben Abdallah A, Wildes TS, Stark SL, McKinnon SL, et al. (2016). Preoperative Falls Predict Postoperative Falls, Functional Decline, and Surgical Complications. *EBioMedicine*, 12, 302–308. [PubMed: 27599969]
 23. van Buuren S (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, 16(3), 219–242. [PubMed: 17621469]
 24. Liu H, & Hays RD Measurement of Interrater Agreement: A SAS/IML Macro Kappa Procedure for Handling Incomplete Data. In *Twenty-Fourth Annual SAS Users Group International Conference*, April 11–14, 1999 (pp. 1620–1625)
 25. Crosby RD, Kolotkin RL, & Williams GR (2003). Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology*, 56(5), 395–407. [PubMed: 12812812]
 26. Youden WJ (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35. [PubMed: 15405679]
 27. Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, et al. (1999). Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation*, 100(10), 1043–1049. [PubMed: 10477528]
 28. Norman GR, Sloan JA, & Wyrwich KW (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care*, 41(5), 582–592. [PubMed: 12719681]
 29. Coteur G, Feagan B, Keininger DL, & Kosinski M (2009). Evaluation of the meaningfulness of health-related quality of life improvements as assessed by the SF-36 and the EQ-5D VAS in patients with active Crohn's disease. *Alimentary Pharmacology and Therapeutics*, 29(9), 1032–1041. [PubMed: 19222413]
 30. Zhang Y, Zhou F, & Sun Y (2015). Assessment of health-related quality of life using the SF-36 in Chinese cervical spondylotic myelopathy patients after surgery and its consistency with neurological function assessment: a cohort study. *Health Qual Life Outcomes*, 13, 39. [PubMed: 25889912]
 31. Bahler CD, & Sundaram CP (2013). Quality of life following laparoscopic living-donor nephrectomy. *JSLs: Journal of the Society of Laparoendoscopic Surgeons*, 17(2), 273–278. [PubMed: 23925021]
 32. Dawson J, Doll H, Coffey J, & Jenkinson C (2007). Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and

- SF-36 assessments following surgery for hallux valgus. *Osteoarthritis and Cartilage*, 15(8), 918–931. [PubMed: 17383907]
33. Shi HY, Lee HH, Chiu CC, Chiu HC, Uen YH, & Lee KT (2008). Responsiveness and minimal clinically important differences after cholecystectomy: GIQLI versus SF-36. *Journal of Gastrointestinal Surgery*, 12(7), 1275–1282. [PubMed: 18454301]
34. Bilbao A, Quintana JM, Escobar A, Garcia S, Andradás E, Bare M, et al. (2009). Responsiveness and clinically important differences for the VF-14 index, SF-36, and visual acuity in patients undergoing cataract surgery. *Ophthalmology*, 116(3), 418–424.e411. [PubMed: 19168220]
35. Helsten DL, Abdallah AB, Avidan MS, Wildes TS, Winter A, McKinnon SL, et al. (2016). Methodological considerations for collecting patient reported outcomes from unselected surgical patients. *Anesthesiology*, 125(3), 495–504. [PubMed: 27355128]
36. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., & Schuler TC (2007). Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*, 17(5), 541–546. [PubMed: 17448732]
37. Wright A, Hannon J, Hegedus EJ, & Kavchak AE (2012). Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *Journal of Manual & Manipulative Therapy*, 20(3), 160–166. [PubMed: 23904756]

Fig. 1.
Proportion of patients experiencing change in physical quality of life, by global assessment response

Fig. 2.
Change in global (a) and VR-12 (b) quality of life, by surgical specialty

Table 1.

Characteristics of the 7,902 surgical patients included in the study

Characteristic	*Number (%)
Age (mean $\pm$ SD)	59.1 $\pm$ 14.1
Female Sex	4608 (58)
White Race	6897 (89)
BMI (median, IQR)	29.3 (25.5,34.1)
Baseline PCS (median, IQR)	40.2 (29.2, 51.3)
Baseline MCS (median, IQR)	56.6 (47.1, 61.4)
Follow-up PCS (median, IQR)	40.3 (31.3, 49.3)
Follow-up MCS (median, IQR)	57.1 (48.0, 61.3)
Mean PCS Change $\pm$ SD	-0.33 $\pm$ 10.3
Mean MCS Change $\pm$ SD	0.04 $\pm$ 10.1
Better quality of life, global measure	4585 (61)
Same quality of life, global measure	2185 (29)
Worse Quality of life, global measure	745 (10)
Charlson Index 3	1486 (19)
ASA Physical Status 3	3251 (41)
Fell in the 6 Months Before Surgery	1995 (26)
Fell in the 30 days After Surgery	480 (6)
Low Physical Activity (< 4 METs)	2263 (29)
Chronic Pain	3521 (45)
Mood Disorder	1601 (20)
Low Procedural Cardiac Risk[27]	3086 (40)
Intermediate Procedural Cardiac Risk	4428 (58)
High Procedural Cardiac Risk	177 (2)
Postoperative Complication	2255 (29)

SD = Standard Deviation; BMI = Body Mass Index; IQR = Inter-Quartile Range; PCS = Physical Component Summary of VR-12; MCS = Mental Component Summary of VR-12; ASA = American Society of Anesthesiologists; METs = Metabolic Equivalents

* Due to different amounts of missing data for each variable, denominator may differ from 7,902

Table 2.

Percent of patients who matched* on the two quality of life instruments, stratified by global assessment category

Category	PCS	MCS	Overall VR-12	Total
Better	2443 (57)	2227 (52)	2551 (60)	4251
Same	1002 (49)	1109 (55)	1357 (67)	2030
Worse	520 (74)	446 (63)	559 (80)	699
Combined	3965 (57)	3782 (54)	4467 (64)	6980

PCS = Physical Component Summary of VR-12; MCS = Mental Component Summary of VR-12

* See methods section for the definition of a “match.”

Table 3.

MCID in VR-12 quality of life for this surgical population, using two methods

Method	PCS change	MCS change	Overall change ^a
Proportional method, improvement	4	6.5	1.5
ROC, improvement (c-statistic)	2.59 (0.64)	6.72 (0.54)	1.25 (0.65)
Proportional method, deterioration	5 to 11 ^b	5.5	3.5
ROC, deterioration (c-statistic)	5.34 (0.64)	5.36 (0.61)	3.74 (0.70)

MCID = Minimum Clinically Important Difference; PCS = Physical Component Summary of VR-12; MCS = Mental Component Summary of VR-12

^a Performed on the average of PCS and MCS scores as a pre-specified sensitivity analysis. Note that this is not a validated measure.

^b A range due to overlap in the proportion curves