Qual Life Res (2011) 20:335–341
DOI 10.1007/s11136-010-9760-z

Response shift effect on gastrointestinal quality of life index after laparoscopic cholecystectomy

Hon-Yi Shi · King-Teh Lee · Hao-Hsien Lee · Yih-Huei Uen · Chong-Chi Chiu

Accepted: 25 September 2010 / Published online: 10 October 2010

Abstract
Purpose Traditional pre- and post-surgery quality of life assessments are inadequate for assessing change in health-related quality of life (HRQoL) after laparoscopic cholecystectomy (LC). This study examined whether a response shift, a change in the internal standards of a patient, occurs in patients who have received LC.
Methods Self-administered gastrointestinal quality of life index (GIQLI) was used to evaluate preoperative, postoperative, and retrospective postoperative HRQoL. Response shifts, unadjusted treatment effects, adjusted treatment effects, and their effect sizes were calculated.
Results In all GIQLI domains, a significant response shift was indicated by the significantly higher pre-test scores compared to then-test scores ($P < 0.05$). The effect size of the response shift ranged from 0.19 for the physical impairment domain of the GIQLI to 0.49 for the total GIQLI score. It was observed the treatment effect was greater after adjusting for the presence of response shift.
Conclusion Patients who have received LC undergo a response shift that affects their outcome measurement at 6 months postoperative. Response shift is a potentially confounding factor and should be considered when designing clinical studies that employ self-administered HRQoL measures. This evidence of confounding effects warrants further study of response shift at longer intervals after LC, after other health care interventions, and in patients with varying preoperative health status.

Keywords Laparoscopic cholecystectomy · Health-related quality of life · Response shift · Gastrointestinal quality of life index

H.-Y. Shi · K.-T. Lee
Graduate Institute of Healthcare Administration, Kaohsiung Medical University, 100, Shih-Chuan 1st Road, Kaohsiung 80756, Taiwan

K.-T. Lee
Division of Hepatobiliary Surgery, Department of Surgery, Kaohsiung Medical University Hospital, 100, Zihyou 1st Road, Kaohsiung 807, Taiwan

H.-H. Lee · C.-C. Chiu (✉)
Department of Surgery, Chi Mei Medical Center, Liou Ying Campus, 201, Taikang Village, Liou Ying Township, Tainan 736, Taiwan
e-mail: honyi6165@gmail.com

Y.-H. Uen
Department of Surgery, Chi Mei Medical Center, Tainan 702, Taiwan

C.-C. Chiu
Chia Nan University of Pharmacy and Science, Tainan, Taiwan

C.-C. Chiu
Taipei Medical University, Taipei, Taiwan

Introduction

Health-related quality of life (HRQoL) outcome assessments are important for evaluating the impact of health care interventions. Health researchers must consider not only biological and physical indicators of health but also the broader implications of self-perceived well-being [1–3]. Self-reported HRQoL outcomes for a given treatment are often inconsistent with actual health status. The literature indicates that patients with terminal diseases and deteriorating health status often consider their quality of life stable and comparable to that of healthy people [4, 5]. One implication of this phenomenon, known as the "response shift" effect [6–8], is that evaluations of the impact of health care interventions on HRQoL are, at worst, inaccurate [2, 9].

Compared to HRQoL assessments, measures that consider potential response shifts may yield more accurate estimates since they are not confounded by potential changes in individual standards, values, or conceptions. The first series of studies of response shift published in *Social Science and Medicine* in 1999 provided an initial glimpse of the theoretical, methodological, and applied implications of response shift [10]. Since then, further research and testing of methods across diseases has progressively refined the theoretical framework for the response shift effect [2, 8]. Inclusion of the appraisal concept in the theoretical framework of response shift further improves understanding of the thinking process of patients when rating their HRQoL [7–12]. Methodological approaches for assessing response shift have also been developed and tested. The then-test and other statistical approaches have been developed specifically for interpreting changes in HRQoL scores [11, 12].

Retrospective pre-tests are now administered in addition to the traditional pre- and post-tests. The underlying assumption is that simultaneously administering the then-test and post-test ensures that the patient applies a consistent internal standard of measurement. Comparison of post-test and then-test HRQoL scores therefore gives a measure of treatment effect without the confounding effect of change in internal standards. Moreover, the size and direction of the response shift can be assessed by subtracting then-test scores from actual pre-test scores [7, 11, 12].

Surgical options in cholecystectomy, which involves the surgical removal of the gallbladder, include the standard laparoscopic cholecystectomy (LC) procedure and open cholecystectomy (OC), which is an older, more invasive procedure. In patients with symptomatic cholelithiasis or cholecystitis, complication rates and mortality rates are lower and hospital stays are shorter after LC than after OC [13–15]. Compared to OC, LC is also associated with substantially higher patient-reported HRQoL scores at 3–6 months postoperatively [13, 14]. After a successful treatment, the change in then-test score is likely to be larger than the change in conventional score. Therefore, the change in conventional score may underestimate treatment effectiveness. As LC is considered a minor procedure with relatively short hospital stay and recovery time, LC patients are likely to be disappointed by longer than expected hospital stay and recovery time, unexpected loss of gainful employment, and the possibility of additional surgery. Although the impact of response shift on self-reported HRQoL outcomes has not been studied extensively in cholecystectomy patients, the literature does indicate that response shift occurs in LC patients. Shi et al. [14] reported that, in patients who had undergone cholecystectomy, the minimal clinically important differences (MCIDs) in scores for the symptoms, emotions, physical functions, and social function subscales of the GIQLI were 6.42, 6.86, 7.64, and 6.46 points, respectively. They also found that the effect sizes of four anchors in the "somewhat better" group (0.38–0.49) exceeded those of the "same" group (0.25–0.38). However, that study did not assess the impact of response shift on the traditional measures of treatment outcome such as the GIQLI. The impact of response shift on patient recovery has important implications in terms of cost-effectiveness and treatment efficacy as well as accuracy of outcome measurements [10–12].

As patient-reported scores have been widely used in recent trials to assess the effectiveness of health care interventions [11–14], response shift in patients may affect HRQoL outcomes. Therefore, the aim of this study was to determine whether a response shift, as indicated by patient-reported functional scores, occurs in patients who undergo LC. If so, a secondary aim was to determine the effect size of the response shift.

Methods

Study design and sample

All patients who had undergone LC performed between March 2007 and February 2009 by any one of three senior surgeons (KT, HH, YH) practicing at two tertiary academic hospitals in southern Taiwan were eligible for the study. Patients with cognitive impairment ($n = 1$), severe organ disease ($n = 1$), or psychiatric disease ($n = 1$) were excluded. Of the 359 eligible subjects who gave written consent and were enrolled in the study at baseline, three were excluded due to the conversion of LC to OC, and three were excluded because they did not undergo postoperative assessments. All involved institutions approved this study of human subjects.

Instruments and measurements

The Gastrointestinal Quality of Life Index (GIQLI) is recognized as a valid and reliable instrument for measuring functional status, especially in patients undergoing cholecystectomy [14]. The thirty-six item GIQLI assesses symptomatology (nineteen items), emotional impairment (five items), physical impairment (seven items), social impairment (four items), and medical treatment effects (one item). Each of its thirty-six items is scored from 0 to 4 with a higher score indicating better health status, and the total GIQLI score ranges from 0 to 144. The GIQLI questionnaire was administered to all eligible subjects immediately before surgery on the day of the preoperative assessment (pre-test). The questionnaires were completed by the patients without assistance from the physician or medical staff to ensure that all responses were entirely self-assessed.

All patients who completed the preoperative GIQLI were then asked to complete two additional GIQLI questionnaires at 6 months postoperatively. One questionnaire asked the respondents to rate their current condition; that is, it measured outcome at 6 months postoperatively (post-test). This assessment is consistent with traditional pre-test and post-test data collection. The second questionnaire required patients to re-evaluate their previous functional status; that is, how they perceived their condition before surgery (then-test). These questionnaires were completed under the same conditions of the pre-test questionnaire.

Because post-test and then-test evaluations were administered simultaneously, patients presumably applied consistent internal scales of measurement [10–12]. No further evaluations were performed until at least 6 months after surgery to allow rehabilitation and full recovery from surgery. This time period was also considered sufficiently long for the patients to forget their pre-test answers and to accurately reassess their pretreatment health status. The reliability of the then-test is equivalent to that of the outcome measure applied, which in this case was the GIQLI [13, 14]. The validity of the then-test is well established in the education field [16].

Statistical analysis

Response shift (RS) (pre-test − then-test), unadjusted treatment effect (UTE) (post-test − pre-test), and adjusted treatment effect (ATE) (post-test − then-test scores) were calculated for each domain of the GIQLI. By definition, a positive response shift implies that patients considered themselves better at the actual pre-test assessment than they did retrospectively at the then-test assessment. Paired group *t*-tests were performed to determine whether response shifts and treatment effects significantly differed. The effect size was calculated for all response shift scores by dividing the change in scores by the standard deviation in the change to evaluate their meaningfulness in terms of Cohen's effect size [17]. Effect sizes of 0.2, 0.5, and 0.8 are typically considered small, medium, and large changes, respectively. Due to the skewed distribution, bias-corrected and accelerated bootstrapping estimation was performed to identify significant differences between the two treatment effects [18]. Analysis of covariance in response shift was performed for associations with age (<65 years old vs. ≧65 years old), gender (female vs. male), body mass index (BMI) (<18.5 kg/m^2, 18.5–24.9 kg/m^2 vs. ≧25.0 kg/m^2), average length of stay (ALOS) (<4 vs. ≧4 days), duration of symptoms (<12 vs. ≧12 months), number of co-morbidities (0, 1, or ≧2), operation time (<90 vs. ≧90 min), number of complications (0 vs. ≧1), and extent of recovery (defined as post-test HRQoL score—pre-test HRQoL score) (1st Quartile, 2nd to 3rd Quartiles, 4th Quartile).

To determine which patients demonstrated a response shift, we calculated minimum detectable change (MDC), the minimum difference between two measurements for an individual that represents a "true" change beyond the measurement error [14]. The following formula was used to determine MDC for the HRQoL scores: $\mathrm{MDC} = 1.96 \times \mathrm{SD} \times \sqrt{[2(1 - \mathrm{ICC})]}$ [19]. Calculating intra-class correlation coefficient (ICC) is the preferred method for quantifying reproducibility given systematic variability. An ICC > 0.70 is considered acceptable [20]. Changes above this threshold are considered "true" changes with 95% probability. Based on the baseline standard deviations (SDs) for each GIQLI domain score for the patients, the MDCs for symptomatology, emotional impairment, physical impairment, and social impairment were 4.88, 1.65, 2.14, and 1.77, respectively.

The Stata Statistical Package, version 9.0 (Stata Corp., College Station, TX, USA) was used for all statistical analyses, and a P value < 0.05 was considered statistically significant.

Results

The study sample included 147 women (41.64%) and 206 men (58.36%) with a mean age of 55.58 ± 14.55 years (range, 24–87 years). Table 1 presents the demographic characteristics for the 353 analyzed patients. Analysis of covariance showed that the total response shift score was

Table 1 Analyzed demographic characteristics ($n = 353$)

Demographic characteristics	Mean ± SD/n (%)
Gender	
Female	206 (58.36%)
Male	147 (41.64%)
Age (years)	55.58 ± 14.59
Body mass index (BMI, kg/m^2)	24.65 ± 3.78
Average length of stay (ALOS, days)	4.73 ± 3.59
Duration of symptoms (months)	12.53 ± 29.29
One or more co-morbidity	165 (44.70%)
Operation time (min)	89.53 ± 47.85
ASA score	2.01 ± 0.66
One or more complication	29 (8.20%)

unaffected by differences in gender ($F = 0.97$, $P = 0.33$), age ($F = 1.93$, $P = 0.17$), BMI ($F = 0.22$, $P = 0.67$), ALOS ($F = 0.18$, $P = 0.71$), duration of symptoms ($F = 0.09$, $P = 0.80$), general co-morbidity ($F = 0.02$, $P = 0.89$), operation time ($F = 1.02$, $P = 0.32$), ASA score ($F = 1.70$, $P = 0.24$), or extent of recovery ($F = 2.98$, $P = 0.09$) after adjusting for the preoperative score in each GIQLI domain.

Table 2 presents the mean pre-test, post-test, and then-test scores for symptomatology, emotional impairment, physical impairment, and social impairment. The mean preoperative scores (pre-test) exceeded the mean scores for the reassessment of preoperative status (then-test), indicating that the subjects considered themselves worse off than if they had judged themselves preoperatively.

Table 3 presents the scores calculated for response shift, unadjusted treatment effect, adjusted treatment effect, and effect size. The domains of symptomatology ($P < 0.001$), emotional impairment ($P < 0.001$), physical impairment ($P < 0.001$), and social impairment ($P < 0.001$) showed the presence of a response shift. Moreover, in all domains, preoperative and 6-month postoperative scores significantly differed ($P < 0.001$), indicating a significant unadjusted treatment effect. Furthermore, in all domains, then-test scores and post-test scores significantly differed ($P < 0.001$), indicating a significant adjusted treatment effect. Additionally, the effect size of the response shift ranged from 0.19 for the physical impairment domain of the GIQLI to 0.49 for the total GIQLI score. Finally, the two treatment effects significantly differed in all GIQLI domains.

Table 2 Mean scores for GIQLI domains ($n = 353$)

Variable	Mean	SD	Lower 95% confidence interval	Upper 95% confidence interval
Symptomatology				
Pre-test	58.69	12.97	57.34	60.05
Then-test	56.36	11.68	55.44	59.28
Post-test	71.06	7.67	69.97	72.15
Emotional impairment				
Pre-test	13.22	4.82	12.72	13.73
Then-test	12.06	3.24	10.87	13.18
Post-test	17.76	3.98	17.19	18.32
Physical impairment				
Pre-test	18.28	6.04	17.65	18.92
Then-test	17.15	5.85	16.04	18.67
Post-test	24.27	5.50	23.49	25.05
Social impairment				
Pre-test	8.85	3.12	8.53	9.18
Then-test	7.48	2.90	6.74	8.29
Post-test	10.81	2.25	10.49	11.13

Table 3 Summary statistical data for response shift (RS), adjusted treatment effect (ATE), and unadjusted treatment effect (UTE)

Variable	Mean	SD	*T* value	*P* value	Effect size
Response shift[a]					
Symptomatology	2.33	4.29	10.21	<0.001	0.20
Emotional impairment	1.16	3.58	6.09	<0.001	0.36
Physical impairment	1.13	3.40	6.24	<0.001	0.19
Social impairment	1.37	3.81	6.76	<0.001	0.47
Total score	5.98	10.97	10.24	<0.001	0.49
Adjusted treatment effect[b]					
Symptomatology	14.70	22.01	12.55	<0.001	1.26
Emotional impairment	5.70	10.73	9.98	<0.001	1.76
Physical impairment	7.12	16.92	7.91	<0.001	1.22
Social impairment	3.33	5.25	11.92	<0.001	1.15
Total score	30.85	34.20	16.95	<0.001	1.78
Unadjusted treatment effect[c]					
Symptomatology	12.37	13.30	17.48	<0.001	0.95
Emotional impairment	4.54	4.84	17.62	<0.001	0.94
Physical impairment	5.99	6.54	17.21	<0.001	0.99
Social impairment	1.96	2.67	13.79	<0.001	0.63
Total score	24.84	22.18	21.05	<0.001	0.98

[a] Response shift = Pre-test score − Then-test score

[b] Adjusted treatment effect = Post-test score − Then-test score

[c] Unadjusted treatment effect = Post-test score − Pre-test score

In addition to determining the average response shift for the overall patient group, this study also assessed the effects of response shift on the GIQLI domains of symptomatology, emotional impairment, physical impairment, and social impairment at the individual patient level, based on the MDC. Only a minority of patients exhibited response shifts in GIQLI domain scores larger than the MDC (eight in symptomatology, twelve in emotional impairment, ten in physical impairment, and twenty-two in social impairment). Therefore, self-reported condition before undergoing surgery was better than that reported in retrospect. No patients exhibited a negative response shift larger than the MDC.

Discussion

The self-reported HRQoL scores obtained in this study indicated that a response shift occurs in patients who have undergone LC surgery. The direction of the response shift indicated that, in most patients, self-assessed preoperative condition was worse when the patients were asked before undergoing LC than it was when asked after undergoing

LC. The similar findings reported in the literature [21, 22] suggest that response shift may result from recalibration, which may explain why improved HRQoL outcomes are often undetected in pre- and post-test analyses. Understanding how the magnitude of the response shift and its direction affect perceived health is essential for comparing recovery outcomes among different studies and for measuring the effectiveness of interventions, particularly that of elective surgeries, which are generally only partially covered by private or national insurance.

Patient-reported HRQoL instruments used in clinical research are not designed to account for response shifts. They assume that patients respond consistently on measurement domains and that the domains are directly comparable across individuals and over time [7, 23]. This study employed the most common approach for assessing change in internal standards, which is retrospective preoperative test or ''then-test'' [7, 23]. A difference between the retrospective preoperative test and the actual preoperative test, taken at the time, is considered an indicator of change in internal standards [10–12]. The present study revealed that such a change in internal standards occurs in patients who have received LC surgery and that this change is measurable by HRQoL scores. The scores for all GIQLI domains (i.e., symptomatology, emotional impairment, physical impairment, and social impairment) revealed a significant shift. This positive response shift after surgery has been reported previously [6, 7]. A likely explanation is adaptation to the illness [7, 24]. That is, patients with chronic diseases are likely to judge their health status more positively when they do so at the time of illness than when they judge their health status retrospectively [25]. A similar influence occurs when the general public is asked to imagine having a particular disease and to assess the associated health status. Respondents typically give lower estimates of health status than do patients currently afflicted with the disease [25].

Only a few patients in this study revealed a positive response shift larger than the MDC, which suggests that they had undergone a detectable change in internal standards resulting from their adaptation to their diseased state, and none revealed a negative response shift larger than the MDC. This relative lack of adaptation may be related to time. The disease onset may have simply been too recent for adaptation in most patients [26, 27]. In this study, mean follow-up was 6 months. Moreover, asking patients specific questions about their disease rather than asking a general question about subjective wellbeing may evoke better recollection of the injury. Finally, these results are consistent with recently reported patient wellbeing studies that reveal individual differences in adaptation to new situations [26, 27].

In the present study, response shifts in the patients treated by LC were unaffected by gender, age, BMI, ALOS, duration of symptoms, operation time, extent of recovery, or co-morbidity when adjusted for preoperative score for each GIQLI domain. Therefore, other variables not considered in this study, such as personality variables (e.g., optimism) or social variables (e.g., social support) may affect the individual adaptation associated with response shift. Response shifts did not alter the interpretation of the clinical results since they had an insignificant confounding effect on unadjusted scores as evidenced by the much smaller effect sizes for response shifts in all domains than for adjusted and unadjusted treatment effects. Significant improvement in both adjusted and unadjusted treatment effects is not unusual when evaluating LC [6–9]. In the clinical context, however, if the reported or unadjusted treatment effect is insignificant, adjusting for a response shift may then identify a considerable improvement (adjusted treatment effect). This interpretation of outcome could occur in association with other health care interventions that do not provide a satisfactory treatment solution.

In the context of clinical studies, an important consequence of response shift is its relation to treatment effect. In this study, the influence of the response shift on all four scores was to increase the treatment effect. This increase was large for the total scores and for social impairment, moderate for emotional impairment, and small for symptomatology and physical impairment scores. As mentioned above, total scores and social impairment domains may be more sensitive to response shifts but are also more susceptible to recall error. However, the response shift was not a confounding factor in the interpretation of HRQoL scores. An earlier study of response shift in patients who had undergone knee micro-fracture reached a similar conclusion [7]. Part of the reason that the response shift had a limited influence on treatment effect in the current study was the relatively large effect size of surgery. Whether measured by UTE or ATE, surgery in this study was associated with significantly improved total scores and social impairment scores.

In a study of total knee replacement surgeries performed in Canadian patients, Razmjou et al. [21] reported a response shift in HRQoL scores obtained 6 months after total knee replacement surgery. The effect size of the response shift ranged from 0.04 to 0.35. A study by Balain et al. [22] analyzed patients who had received microfracture treatment for full thickness cartilage defects in United Kingdom. They also reported an effect size of response shift ranging from 0.35 for Lysholm score and Internal Knee Documentation Committee (IKDC2) symptom score to over 0.90 for the visual analogue scale (VAS) pain score. However, as noted above, the effect size of the response shift (0.19–0.49) in the Taiwan population analyzed in this study was within the ranges reported in the literature.

Although this suggests that response shift is unaffected by cultural differences (i.e., Eastern versus Western), further confirmation is needed.

As noted in Ubel et al. [12], reprioritization and reconceptualization do not necessarily invalidate HRQoL change. Moreover, researchers can minimize response shifts associated with "reconceptualization" or "reprioritization" of concepts by using questionnaires to measure narrower constructs [12, 28]. However, the present study investigated how response shift resulting from recalibration affects the GIQLI domains of symptomatology, emotional impairment, physical impairment, and social impairment. The patients apparently changed their internal standards and adopted more ambitious goals after taking the pre-test. Additionally, a recent study reported that the magnitude of recalibration response shift increased as patients adapted to lower levels of pain and dysfunction [21]. Therefore, recalibration may be more likely after LC than after other surgical procedures.

We recognize several limitations of this study. One is the use of then-test to measure response shift. This method is the most widely used test, mainly because of its simplicity [7, 10]. It assumes that respondents apply their postoperative internal standards when providing a re-evaluation or 'then-test' rating of their preoperative score. It therefore relies on the ability of the patients to accurately recall their preoperative health status. Incorrect recall increases random error as well as systematic error (recall bias), which is the main disadvantage of the method [10–12]. Possible recall bias of the then-test design was studied by Visser et al. [23], who compared convergent validity in three measures of response shift: structured equation modeling, anchor recalibration, and the then-test design. In the present study, response shift adaptation varied among the GIQLI domains, which indicates that recall bias was not a major contributor to our findings. Second, this approach does not indicate whether the apparent response shift is actually attributable to other factors such as recall bias, measurement error, etc. Without additional evidence from cognitive interviews and advanced statistical methods (such as structural equation modeling), interpreting and confirming the above findings are problematic. Third, the magnitude of response shift may be attributable to factors other than those examined in this study. Further studies are needed to determine how response shift is associated with treatment type and other psychological and clinical variables (such as personality traits and disease severity). Sensitivity analysis would also elucidate the impact of response shift by accounting for different surgical treatment scenarios. For example, future studies may determine how response shift in LC patients differs from that in patients who undergo other procedures.

Conclusions

To the best of our knowledge, this study is the first to assess the response shift effect on patient-reported HRQoL after LC surgery at the group level as well at the individual patient level. At the group level, the influence of response shift was noted in all GIQLI domains. However, analysis of individual patients confirmed that some patients may not exhibit response shift effects. Response shift is a potential confounding effect and should be considered when designing clinical studies that employ self-administered HRQoL measures. These positive results warrant further study of response shift at longer intervals after surgery, after other health care interventions, and in patients with varying preoperative health status.

References

1. Lohr, K. N., & Zebrack, B. J. (2009). Using patient-reported outcomes in clinical practice: Challenges and opportunities. *Quality of Life Research, 18*(1), 99–107.
2. Ahmed, S., Schwartz, C., & Ring, L. (2009). Applications of health-related quality of life for guiding health care: Advances in response shift research. *Journal of Clinical Epidemiology, 62*(11), 1115–1117.
3. Valderas, J. M., Kotzeva, A., Espallargues, M., Guyatt, G., Ferrans, C. E., Halyard, M. Y., et al. (2008). The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. *Quality of Life Research, 17*(2), 179–193.
4. Breetvelt, I. S., & Van Dam, F. S. (1991). Underreporting by cancer patients: The case of response-shift. *Social Science and Medicine, 32*(9), 981–987.
5. Richards, A., & Folkman, S. (2000). Response shift: A coping perspective. In C. E. Schwartz & M. A. G. Sprangers (Eds.), *Adaptation to changing health: Response shift in quality-of-life research* (pp. 25–36). Washington, DC: American Psychological Association.
6. Schwartz, C. E., & Finkelstein, J. A. (2009). Understanding inconsistencies in patient-reported outcomes after spine treatment: Response shift phenomena. *The Spine Journal, 9*(12), 1039–1045.
7. Balain, B., Ennis, O., Kanes, G., Singhal, R., Roberts, S. N., Rees, D., et al. (2009). Response shift in self-reported functional scores after knee micro-fracture for full thickness cartilage lesions. *Osteoarthritis and Cartilage, 17*(8), 1009–1013.
8. Barclay-Goddard, R., Epstein, J. D., & Mayo, N. E. (2009). Response shift: A brief overview and proposed research priorities. *Quality of Life Research, 18*, 335–346.
9. Robertson, C., Langston, A. L., Stapley, S., McColl, E., Campbell, M. K., Fraser, W. D., et al. (2009). Meaning behind measurement: Self-comparisons affect responses to health-related quality of life questionnaires. *Quality of Life Research, 18*, 221–230.
10. Schwartz, C. E., & Sprangers, M. A. G. (1999). Introduction to symposium on the challenge of response shift in social science and medicine. *Social Science and Medicine, 48*, 1505–1506.
11. Schwartz, C. E., & Sprangers, M. A. G. (2010). Guidelines for improving the stringency of response shift research using the then test. *Quality of Life Research, 19*(4), 455–464.

12. Ubel, P. A., Peeters, Y., & Smith, D. (2010). Abandoning the language of “response shift”: A plea for conceptual clarity in distinguishing scale recalibration from true changes in quality of life. *Quality of Life Research, 19*(4), 465–471.
13. Shi, H. Y., Lee, K. T., Lee, H. H., Uen, Y. H., Tsai, J. T., & Chiu, C. C. (2009). Post-cholecystectomy quality of life: a prospective multicenter cohort study of its associations with preoperative functional status and patient demographics. *Journal of Gastro-intestinal Surgery, 13*(9), 1651–1658.
14. Shi, H. Y., Lee, K. T., Lee, H. H., Uen, Y. H., Na, H. L., Chao, F. T., et al. (2009). The minimal clinically important difference in the gastrointestinal quality-of-Life Index after cholecystectomy. *Surgical Endoscopy, 23*, 2708–2712.
15. Keus, F., Gooszen, H. G., & Van Laarhoven, C. J. (2009). Systematic review: open, small-incision or laparoscopic cholecystectomy for symptomatic cholecystolithiasis. *Alimentary Pharmacology Therapeutics, 29*(4), 359–378.
16. Hoogstraten, J. (1982). The retrospective pretest in an educational training context. *Journal of Experimental Education, 50*, 200–204.
17. Kazis, L. E., Anderson, J. J., & Meenan, R. F. (1989). Effect sizes for interpreting changes in health status. *Medical Care, 27*(3), 178–189.
18. Effron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.
19. Briggs, K. K., Lysholm, J., Tegner, Y., Rodkey, W. G., Kocher, M. S., & Steadman, J. R. (2009). The reliability, validity, and responsiveness of the Lysholm score and Tegner activity scale for anterior cruciate ligament injuries of the knee: 25 years later. *The American Journal of Sports Medicine, 37*(5), 890–897.
20. Beckerman, H., Roebroeck, M. E., Lankhorst, G. J., Becher, J. G., Bezemer, P. D., & Verbeek, A. L. (2001). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research, 10*, 571–578.
21. Razmjou, H., Schwartz, C. E., Yee, A., & Finkelstein, J. A. (2009). Traditional assessment of health outcome following total knee arthroplasty was confounded by response shift phenomenon. *Journal of Clinical Epidemiology, 62*(1), 91–96.
22. Balain, B., Ennis, O., Kanes, G., Singhal, R., Roberts, S. N., Rees, D., et al. (2009). Response shift in self-reported functional scores after knee micro-fracture for full thickness cartilage lesions. *Osteoarthritis and Cartilage, 17*(8), 1009–1013.
23. Visser, M. R., Oort, F. J., & Sprangers, M. A. (2005). Methods to detect response shift in quality of life data: A convergent validity study. *Quality of Life Research, 14*, 629–639.
24. Postulart, D., & Adang, E. M. (2000). Response shift and adaptation in chronically ill patients. *Medical Decision Making, 20*(2), 186–193.
25. Ubel, P. A., Loewenstein, G., & Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research, 12*, 599–607.
26. Mayo, N. E., Scott, S. C., Dendukuri, N., Ahmed, S., & Wood-Dauphinee, S. (2008). Identifying response shift statistically at the individual level. *Quality of Life Research, 17*, 627–639.
27. Diener, E., Lucas, R. E., & Napa Scollon, C. (2006). Beyond the hedonic treadmill Evising the adaptation theory of well-being. *American Psychologist, 61*(4), 305–314.
28. Reeve, B. B. (2010). An opportunity to refine our understanding of “response shift” and to educate researchers on designing quality research studies: Response to Ubel, Peeters, and Smith. *Quality of Life Research, 19*(4), 473–475.