ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

An Exploratory Study of Response Shift in Health-Related Quality of Life and Utility Assessment Among Patients with Osteoarthritis Undergoing Total Knee Replacement Surgery in a Tertiary Hospital in Singapore

Xu-Hao Zhang, PhD[1], Shu-Chuen Li, PhD[2], Feng Xie, PhD[3], Ngai-Nung Lo, FRCS[4], Kwang-Ying Yang, FRCS[4], Seng-Jin Yeo, FRCS[4], Kok-Yong Fong, FRCP[5,6], Julian Thumboo, FRCP[5,6,]*

[1]*Department of Pharmacy, National University of Singapore, Singapore;* [2]*Discipline of Pharmacy and Experimental Pharmacology, School of Biomedical Sciences, University of Newcastle, Newcastle, Australia;* [3]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada;* [4]*Departments of Orthopaedic Surgery and* [5]*Rheumatology and Immunology, Singapore General Hospital, Singapore;* [6]*Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore*

ABSTRACT

Objective: To investigate the influence of response shift (RS) on health-related quality of life (HRQOL) and utility assessment among patients undergoing total knee replacement. **Methods:** Consenting patients undergoing total knee replacement were interviewed to determine their HRQOL by using the six-dimensional health state short form, derived from SF-36, and the EuroQol five-dimensional questionnaire at baseline (pretest 1) and the six-dimensional health state short form, derived from SF-36, at 6 (pretest 2) and 18 months after surgery (posttest). RS was studied by using a "then-test" approach by contacting participants 18 months after surgery and asking them to evaluate their HRQOL at baseline (then-test 1) and at 6 (then-test 2) and 18 months after surgery. RS was calculated as the score difference between pretest and then-test scores for a given time point. Relationships between RS and external variables were explored by using univariate and multiple liner regression analyses. **Results:** In 74 subjects (63% response rate, median age 68 years), median (interquantile range) six-dimensional health state short form, derived from SF-36, scores for then-tests at baseline (0.48 [0.42–0.49]) and at 6 months (0.72 [0.66–0.79]) after surgery were significantly different from respective pretest scores (0.61 [0.58–0.68] at baseline, $P = 0.000$; 0.69 [0.63–0.72] at 6 months, $P = 0.000$), showing RS at both time points. RS at baseline (0.14 [0.08–0.20]) was significantly larger than that at 6 months (−0.05 [0.14 to 0.00], $P = 0.000$). EuroQol five-dimensional questionnaire pretest and then-test scores at baseline also differed significantly (0.69 [0.17–0.73] vs. −0.18 [−0.23 to 0.00], $P = 0.000$). RS at baseline was not affected by assessed demographic or medical variables. RS at 6 months was greater in subjects with more years of education (16% of variance in multiple liner regression, $P < 0.01$). **Conclusion:** RS was present and impacted HRQOL and utility assessment among patients undergoing total knee replacement before and 6 months after surgery.

Keywords: health-related quality of life, response shift, total knee replacement, utility assessment.

Introduction

Originating from self-report evaluation studies on organizational and educational psychology, a phenomenon termed *response shift* has been increasingly reported in patient-reported outcomes assessment among chronically or terminally ill patients over the last decade[1–3].

In the health-care arena, response shift has been regarded as an instinctive psychological mechanism of patients to adapt to the changes caused by illness or treatment [4]. Sprangers and Schwartz [5] defined *response shift* as a change in the meaning of one's self-evaluation of a target construct as a result of a) a change in the respondent's internal standards of measurement (scale recalibration); b) a change in the respondent's values (scale reprioritization); or c) a redefinition of the target construct (reconceptualization). Although distinguished as three types of response shift [5], reconceptualization, scale recalibration, and reprioritization are thought to occur in combination [6].

Health-related quality of life (HRQOL) measurement may be affected by response shift because it quantifies patient perceptions, which may change with time because of response shift. Paradoxes such as overestimation of health status or underestimation of treatment effects measured by HRQOL outcomes have been found across various patient groups, including those with cancer, stroke, mental illness, and so on [7–9]. A theoretical model has been built to illustrate the relationship between response shift and HRQOL, in which "changes in an individual's health status may prompt behavioral, cognitive and affective processes necessary for accommodating illness, which may be influenced by antecedents (e.g., sociodemographics, personalities, expectations, etc.) of

Conflicts of interest: The authors have indicated that they have no conflicts of interest with regard to the content of this article.

* *Address correspondence to:* Julian Thumboo, Department of Rheumatology and Immunology, Singapore General Hospital, Outram Road, Singapore 169608.

E-mail: julian.thumboo@sgh.com.sg.

1098-3015/$36.00 – see front matter Copyright © 2012, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

doi:10.1016/j.jval.2011.11.011

the individual; These processes have the potential to change an individual's standards, values and conceptualization of HRQOL" [7]. The presence of response shift calls into question the assumption that patients would perceive and value a self-reported item with entirely the same internal standards during longitudinal research. In other words, there may be situations where true change measured by HRQOL instruments may not be simply calculated as the difference between respective pre- and postintervention test scores [10,11].

In surgical interventions including total knee replacement (TKR), pre- and postintervention comparisons of HRQOL have been used as a standard method to evaluate patients' improvement in both generic and disease-specific health status, and consequently to determine the cost-effectiveness of the treatment [12,13]. Because of response shift's possible impact on such studies, its exploration has become an emerging area in HRQOL research of surgical interventions [14–16]. There is, however, limited information currently available on the impact of response shift in subjects undergoing TKR, with only one recent publication showing that response shift significantly affected postoperative function 6 months after TKR when measured by using a disease-specific HRQOL questionnaire, the Western Ontario and McMaster Universities Osteoarthritis Index [17]. The impact of response shift on generic HRQOL instruments, however, including preference-based HRQOL instruments (e.g., the EuroQol five-dimensional questionnaire [EQ-5D] or the six-dimensional health state short form, derived from SF-36 [SF-6D]), is not known; if present, this may lead to inaccurate or even invalid results when these instruments are used in utility assessment in this situation. In addition, the evidence of response shift's influence on patients undergoing TKR over a follow-up period longer than 6 months is also lacking. Neither is it clear whether response shift also affects comparisons between two postoperative time points for recovery assessment.

To address these gaps in the literature, the primary objective of the current study was to explore and compare the impact of response shift on HRQOL and utility scores measured by generic HRQOL instruments at baseline and 6 months after TKR when assessed 18 months after TKR. It was hypothesized that response shift at baseline would be larger than that at 6 months following TKR, given that there was no major intervention between 6 months and 18 months postoperatively. If response shift were demonstrated, potential demographic and health-related factors associated with response shift at that time point would be investigated. In addition, the agreement between the SF-6D and the EQ-5D in detecting response shift would also be explored. Based on a comparison study of the EQ-5D and the SF-6D across seven patient groups including those with osteoarthritis, it was hypothesized that the correlation of response shift between the two measures would be moderate as categorized by Cohen's criteria (a correlation coefficient within the range of 0.3–0.5) [18,19].

Methods

Subjects and study design

Contactable consenting patients undergoing TKR without cognitive problems seen at a tertiary referral center in Singapore were recruited in this Institutional Review Board–approved study. Because of difficulties in communication during the telephone survey (the third phase as mentioned below), dialect-speaking patients who could not speak either English or Mandarin Chinese ($n = 19$) were excluded. In addition, patients undergoing any additional surgery during the study period were also excluded to obviate any confounding physical and psychological impact caused by this additional surgery.

This Institutional Review Board–approved prospective study was carried out in three phases. Data for the first two phases were retrieved from an earlier Institutional Review Board–approved study, in which generic HRQOL and utility scores were determined by an interviewer by using the SF-6D and the EQ-5D at baseline before the surgery (pretest 1) and using the SF-6D 6 months after surgery (pretest 2) [20]. Response shift was studied by using the "then-test" approach in the third phase, in which eligible Mandarin- or English-speaking patients were interviewed through the telephone 18 months after their surgery. In this telephone interview, patients were asked to give their HRQOL scores for their current health status by using both the SF-6D and the EQ-5D (i.e., post-test scores). They were also asked to give their HRQOL scores at baseline (i.e., then-test 1 scores) and 6 months after TKR (i.e., then-test 2 scores). The rationale of the then-test approach is that at post-test using the same measure, respondents will provide their retrospective judgment of the health status at baseline and 6 months using the same internal standard [21]. In the scoring scheme of the then-test approach, response shift is calculated as the difference between pretest and then-test scores for each time point assessed, in this case at baseline and 6 months after TKR. True change or adjusted treatment effect is calculated as the difference between respective post-test and then-test scores. The difference between respective post-test and pretest scores was considered the observed change or unadjusted treatment effect [22]. Additional data collected during the telephone survey included demographics (age, gender, education level, work status, dwelling type), medical information (presence of acute or chronic illnesses, past knee surgery, number of knees operated), and general satisfaction with knee surgery (on a 0–10 Likert scale).

HRQOL measures

SF-6D

The SF-6D is a preference-based HRQOL measure assessing physical functioning, role limitations, social functioning, pain, mental health, and vitality, with four to six levels per dimension, allowing 18,000 health states to be identified [23,24]. An SF-6D health state is defined by selecting one level from each dimension. The SF-6D score or the SF-6D utility index is scaled from 0.26 to 1.00 continuously, with 0.26 representing the worst health state (all dimensions being at the worst level) and 1.00 representing full health (all dimensions being at full functional level). The validity and equivalence of the SF-6D in English and Chinese versions have been previously demonstrated in a population-based study in Singapore [25].

EQ-5D

The EQ-5D is a preference-based HRQOL measure with five domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) for respondents to self-classify and rate their health status [26]. For each item, there are three response levels (namely, with no problem, with some problems, and with extreme problems), which allow 243 unique health states to be identified. Scoring methods have been developed to assign each of these health states a utility score, in which 1 represents full health (no problem with all five items) and 0 represents being dead. The range of the final score or the utility index is from −0.594 to 1.00 [27]. The validity of English and Chinese versions of the EQ-5D has been previously demonstrated among patients with rheumatic diseases in Singapore [28].

Statistical analysis

Data were entered into a Microsoft Excel spreadsheet (Microsoft Corporation, Redmond, WA) and analyzed by using SPSS 13.0 (SPSS, Inc., Chicago, IL). All tests were two tailed and conducted at

Table 1 – Subject characteristics.

	Median (interquantile range), unless otherwise specified (N = 74)
Age (y)	68 (63–76)
Ethnicity, n (%)	
Chinese	66 (89.2)
Malay	3 (4.1)
Indian	4 (5.4)
Female, n (%)	60 (81.1)
Education, n (%)	
≤6 y	49 (66.2)
7–12 y	19 (25.7)
>12 y	6 (8.1)
Working, n (%)	10 (13.5)
Housing, n (%)	
Private	15 (20.3)
Public	59 (79.7)
Presence of acute disease(s),* n (%)	53 (71.6)
Presence of chronic disease(s),† n (%)	50 (67.6)
Past knee surgery, n (%)	7 (9.7)
Number of knee(s) operated on at baseline, n (%)	
1	59 (79.7)
2	15 (20.3)
Satisfaction with the operation(s)‡	8 (8–9)
SF-6D index for current health status at 18 mo	0.77 (0.66–0.90)
EQ-5D index for current health status at 18 mo	0.87 (0.71–1.00)

EQ-5D, EuroQol five-dimensional questionnaire; SF-6D, six-dimensional health state short form, derived from SF-36.
* Having had at least one of the acute disease(s) in the following five categories: a running nose/sore throat/cough, vomiting/diarrhea, a headache lasting more than 1 d, sleeping disorder, body injury in the past 4 wk.
† Having had at least one of the chronic disease(s) in the following nine categories: diabetes, hypertension, heart disease, stroke, asthma or other lung disease, cancer, rheumatism or back pain or other bone or muscle illness, mental illness, other chronic diseases.
‡ Self-reported satisfaction with the total knee replacement was measured on a 11-point Likert scale from 0 (totally unsatisfied) to 10 (totally satisfied).

a significance level of 0.01 to reduce the possibility of spurious tests of significance due to multiple comparisons. Descriptive analyses were used to characterize demographics (age, gender, education level, work status, dwelling type), medical information (presence of acute or chronic illness, past knee surgery, number of knees operated), and general satisfaction with knee surgery (on a 0–10 Likert scale). Data with a normal distribution were reported as mean (SD). Otherwise, medians (interquantile range or IQR) were reported. Wilcoxon signed rank tests were used to assess the significance of differences between pre- and then-test scores and between true change and observed change at baseline and 6 months after surgery, respectively. In addition, Wilcoxon signed rank tests were used for comparisons of response shift, observed change, and true change between baseline and 6 months after the surgery. Relationships between response shift and external variables were investigated by univariate analyses by using Mann-Whitney or Kruskal-Wallis tests for categorical independent variables, or Spearman's correlation for continuous independent variables. Independent variables with $P < 0.10$ in univariable analyses were then entered into respective multiple linear regression

Table 2 – Horizontal comparisons between pre- and then-test and between true change and observed change at baseline and 6 mo after total knee replacement.

		Median (interquantile range), unless otherwise specified							
Baseline	N	Pretest 1*	Then-test 1†	Magnitude of response shift‡	P	Posttest at 18 mo	Observed change§	True change‖	P
Baseline (SF-6D)	71	0.61 (0.58–0.68)	0.48 (0.42–0.49)	0.14 (0.08–0.20)	<0.001	0.77 (0.66–0.90)	0.16 (0.02–0.26)	0.30 (0.18–0.39)	<0.001
Baseline (EQ-5D)	68	0.69 (0.17–0.73)	−0.18 (−0.23 to 0.00)	0.72 (0.22–0.91)	<0.001	0.87 (0.71–1.00)	0.27 (0.02–0.62)	0.95 (0.65–1.16)	<0.001
At 6 mo	N	Pretest 2¶	Then-test 2#	Magnitude of response shift	P value	Posttest at 18 mo	Observed change	True change	P value
Six months after the surgery (SF-6D)	62	0.69 (0.63–0.72)	0.72 (0.66–0.79)	−0.05 (−0.14 to 0.00)	<0.001	0.77 (0.66–0.9)	0.10 (0.03–0.21)	0.03 (0.00–0.09)	<0.001

EQ-5D, EuroQol five-dimensional questionnaire; SF-6D, six-dimensional health state short form, derived from SF-36.
Please note that the result was based on the statistical analysis of response shift of 74 patients, and so it might be slightly different from the result of a simple subtraction.
* Pretest 1: measure taken at baseline, just prior to the TKR surgery.
† Then-test 1: retrospective measure of baseline score taken at 18 mo after the TKR surgery.
‡ Magnitude of response shift = respective (pretest score – then-test score).
§ Observed change = respective (post-test score – pretest score).
‖ True change = respective (post-test score – then-test score).
¶ Pretest 2: measure taken at 6 mo after the TKR surgery.
\# Then-test 2: retrospective measure of score of 6-mo postoperative score taken at 18 mo after the TKR surgery.

Table 3 – Response shift in domains of the SF-6D at baseline and 6 mo after total knee replacement*.

SF-6D domains†	Median (interquantile range)		P
	0 mo	6 mo	
Physical functioning (six levels)	1.00 (1.00 to 0.00)	1.00 (1.00 to 0.00)	0.059
Role limitations (four levels)	2.00 (2.00 to 0.00)	0.50 (1.00 to 0.00)	<0.001
Social functioning (five levels)	2.00 (3.00 to 0.00)	1.00 (1.00 to 0.00)	0.002
Pain (six levels)	1.00 (2.00 to 0.00)	0.00 (1.00 to 0.00)	<0.001
Mental health (five levels)	2.00 (2.00 to 1.00)	0.00 (1.00 to 0.00)	<0.001
Vitality (five levels)	1.00 (2.00 to 0.00)	1.00 (0.00–2.00)	<0.001

SF-6D, six-dimensional health state short form, derived from SF-36.
* Response shift = pretest level – then-test level.
† Each domain is measured by numeric levels from "1" onward, with higher number indicating worse condition.

(MLR) models to explore factors potentially impacting response shift. The dependent variable for each MLR was response shift at baseline or at 6 month after TKR, respectively. Because of the small number of subjects, we considered the results of MLR analysis as exploratory. The agreement between the SF-6D and the EQ-5D in detecting response shift was explored by Spearman's correlation and Bland-Altman plots.

Results

Response rate and subject characteristics

The response rate of this 18-month follow-up study was 63% (74 of 117 patients). Altogether 43 patients were not recruited for the following reasons: dialect speaking (n = 19), uncontactable (n = 10), declined participation (n = 5), cognitive problems (n = 4), overseas residence (n = 2), admission into hospital (n = 1), deafness (n = 1), and death (n = 1).

The majority of respondents were elderly women with few years of education and with chronic disease(s) (median [IQR] age: 68 [63–76] years, 81% females, 92% with <12 years of education, 68% with at least one chronic disease). Patients' satisfaction with the TKR was high, with a median (IQR) score of 8 (8–9). More detailed subject characteristics are presented in Table 1.

Presence and impact of response shift

As shown in Table 2, median (IQR) SF-6D scores of then-tests at baseline (0.48 [0.42–0.49]) and 6 months after TKR (0.72 [0.66–0.79]) were significantly different from the respective pretest scores (0.61 [0.58–0.68] at baseline, $P = 0.00$; 0.69 [0.63–0.72] at 6 months after TKR, $P = 0.00$), indicating the presence of response shift at both time points. Interestingly, response shift at baseline (0.14 [0.08–0.20]) was significantly larger than that at 6 months after TKR (0.05 [0.14 to 0.00], $P = 0.00$). When measured by the EQ-5D, significant difference was also detected between pretest and then-test scores at baseline (0.69 [0.17–0.73] for pretest vs. −0.18 [−0.23 to 0.00] for then-test, $P = 0.00$).

Thus, when response shift was considered in studying the impact of TKR, the adjusted improvement in health status became significantly greater between the pre- and 6-month postoperative period (true change of 0.30 [0.18–0.39] by the SF-6D and 0.72 [0.22–0.91] by the EQ-5D, $P = 0.00$). On the contrary, after adjustment, the treatment effect between the periods of 6 months and 18 months after surgery was considered as not clinically minimally important as per the minimally important difference for the SF-6D published by Walters and Brazier [32], though statistically significant (true change of 0.03 [0.00–0.09] by the SF-6D, $P = 0.00$) (Table 2).

Similarly, when the magnitude of response shift was further studied by using individual SF-6D items (Table 3), it was found that the degree of impairment in all six domains (each measured by one item) was rated as more severe during the then-test at baseline. A similar situation was also found in the then-test 6 months after TKR, except for the domain of "vitality" where the direction of response shift was toward less severe reduction, in contrast with that observed for the other domains.

Interestingly, contrary to the quantitative data showing the presence of response shift provided by patients, these same subjects generally did think that their then-test and pretest ratings were similar. At the end of the telephone survey, when asked whether their then-test ratings were different from pretest ratings, 70 of the 74 patients thought that these would be similar for both time points assessed (i.e., at baseline and 6 months postoperatively). At baseline, the magnitude of response shift of patients who thought their scores were similar (n = 70) (by EQ-5D: 0.72 [0.25–0.90]; by SF-6D: 0.15 [0.08–0.20]) was slightly larger than that of patients who were actually aware of the difference (n = 4) (by EQ-5D: 0.57 [0.04–1.02]; by SF-6D: 0.12 [0.07–0.25]). There seemed to be, however, no obvious difference between the two groups regarding response shift at 6 months after TKR (SF-6D scores of −0.05 [−0.14 to −0.00] vs. −0.05 [−0.19 to −0.05], respectively). As the sample size of the comparison group is very small (n = 4), the data should be interpreted as exploratory trending data.

Influence of external variables on response shift

Univariate analysis indicated that response shift at baseline measured by using the SF-6D was significantly influenced by education level, working status, and presence of chronic disease(s). Interestingly, patients with less education and chronic disease(s) who were not working experienced a larger degree of response shift than did those having higher education and a job but no chronic disease. The 6-month postoperative response shift measured by using the SF-6D, however, was significantly influenced only by education level. In contrast with the findings at baseline, patients with more education experienced a larger degree of response shift.

The exploratory MLR analysis (Table 4) suggested that the combination of education level, working status, and presence of chronic disease accounted for 8% of the variance in baseline response shift ($P = 0.05$). At 6 months postoperatively, educational level was the only external variable incorporated, accounting for 16% of the variance ($P < 0.01$).

Systematic difference between the SF-6D and the EQ-5D of detecting response shift

In the current study, the EQ-5D was found to have a significantly larger magnitude of response shift at baseline (0.72 [0.22–0.91]) as compared with the SF-6D (0.14 [0.08–0.20]). A moderate correlation (Spearman's correlation coefficient of 0.43 [$P = 0.00$]) in response shift at baseline for these two measures was found as hypothesized. Further comparison of the degree of response shift by

Table 4 – Influence of external variables on respective response shift at baseline and 6 mo after total knee replacement.

Independent variable(s)*	Regression coefficient (95% confidence interval)	P	Adjusted R^2
MLR on baseline response shift†			
Education level‡			0.08
7–12 y	−0.04 (−0.10 to 0.02)	0.179	
>12 y	−0.08 (−0.17 to 0.00)	0.068	
Working status‡	−0.02 (−0.10 to 0.06)	0.608	
Presence of chronic disease(s)‡	0.41 (−0.01 to 0.10)	0.110	
MLR on response shift at 6 mo after total knee replacement			
Education level‡			0.16
7–12 y	−0.05 (−0.10 to 0.00)	0.055	
>12 y	−0.12 (−0.20 to −0.05)	0.001	

SF-6D, six-dimensional health state short form, derived from SF-36.
* Only external variable(s) with $P < 0.05$ in univariable analyses were incorporated into (MLR) models as independent variable(s). No further external variables had $P < 0.1$.
† Multiple linear regression model; Response shift in the table was measured by the SF-6D.
‡ Reference categories of education level, working status, and presence of chronic disease(s) were ≤6 y of education, not working, and with no chronic disease(s), respectively.

using a Bland-Altman plot (Fig. 1) also showed moderate agreement between the two measures, with most of the data points falling in between the lower and upper 95% limit of agreement. A systematic difference in response shift between the SF-6D and the EQ-5D was also detected in the plot, demonstrated by the presence of a linear relationship between EQ-5D and SF-6D scores. The mean difference between the two measures (EQ-5D minus SF-6D) was 0.43 (SD 0.40; 95% CI (−0.36 to 1.22]), suggesting that the response shift detected by the EQ-5D might be potentially larger than that detected by the SF-6D.

Discussion

In this 18-month follow-up study of patients undergoing TKR, we detected the presence of response shift and quantified its impact on both HRQOL and utility assessment by using the then-test approach. We found that response shift was present at both baseline and 6 months after TKR and that it significantly influenced HRQOL scores. This may have an impact on the use of conventional pre- and post-test methods to assess improvement in HRQOL and utility scores in longitudinal studies. It also suggests that treatment effect may be masked by response shift, because of gradual adaptation to an improving health status of patients [29,30]. To the best of our knowledge, this is the first study exploring response shift by using generic measures and over a prolonged 18-month postoperative period among patients undergoing TKR. Given that TKR is a commonly performed procedure, our findings are important in several ways as detailed in the following paragraphs.

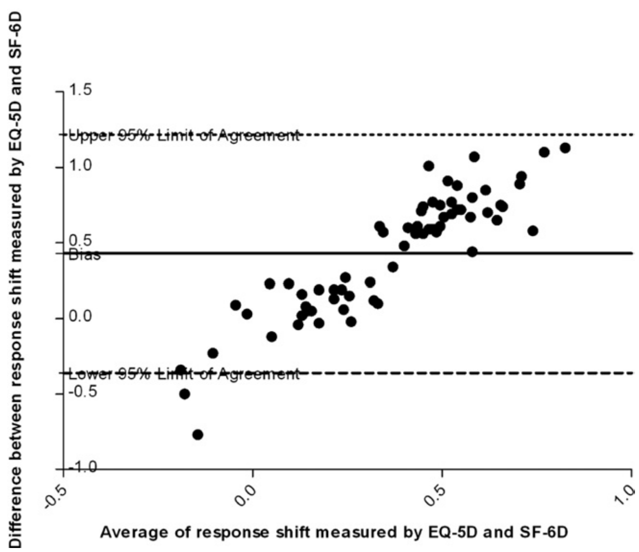

Fig. 1 – Bland–Altman plot: Difference versus average of response shift measured by the EQ-5D and the SF-6D. EQ-5D, EuroQol five-dimensional questionnaire; SF-6D, six-dimensional health state short form, derived from SF-36.

First, our data complement and extend the findings of the previously reported 6-month longitudinal study on another group of patients undergoing TKR using Western Ontario and McMaster Universities Osteoarthritis Index, further supporting the idea of incorporating the measurement of response shift to more accurately measure treatment effects [13].

Second, our results raise concerns over the accuracy of generating utility differences in the conventional post- and pretest manner to evaluate the effectiveness or cost-effectiveness of an intervention or to make comparisons between several interventions. The substantial changes caused by response shift may also have clinically important implications on drug subsidy or technology assessment, as illustrated below.

In addressing the clinical implications of response shift, the minimal important difference (MID) for an HRQOL or utility score needs to be considered. The *MID* is defined as the smallest difference in score that patients perceive as beneficial [31]. The MID of the SF-6D and the EQ-5D has been reported as 0.041 and 0.074, respectively [32]. Based on our results, the quantum of response shift as a percentage of MID changed substantially over time (response shift at baseline: 341% by the SF-6D and 97% by the EQ-5D; response shift at 6 months after TKR: 122% by the SF-6D). With regard to economic impact, after adjustment for the response shift observed in this study, the cost-utility ratio would be decreased by almost twofold, suggesting a substantial increase in the cost-effectiveness of TKR. For example, when a hypothetical value of US$10,000 is assigned to account for direct and indirect costs of TKR over 18 months, the impact of response shift in SF-6D scores on the cost-utility ratio is as high as US$29,167 per quality-adjusted life-year (QALY), changing the unadjusted ratio of US$62,500 per QALY, which would be considered not cost-effective (based on a commonly used cutoff point of US$50,000 per QALY), to US$33,333 per QALY, which would be considered cost-effective [33]. Therefore, the impact of response shift could potentially topple the decisions on the approval and subsidy of interventions, especially for cases whose cost-utility ratios are close to the cutoff point before adjusting for response shift. Furthermore, the presence and impact of response shift could potentially influence comparisons across different studies, if response shift were not fully assessed and adjusted for. Interestingly, although we have shown that the response shift generated from the SF-6D and the EQ-5D differed systematically, it is of note that the impact of the response shift on the cost-utility ratio in the above example if measured by the EQ-5D would be $26,511 per QALY, close to that of the SF-6D.

Third, by extending the study period beyond 6 months, we had the opportunity to further characterize the nature of response shift and its influence on health status both at baseline and 6 months after TKR. We found that during the first 6-month recovery period, patients may have experienced a larger degree of response shift because of a comparatively faster pace of recovery; subsequently, a plateau in health status could have been reached and maintained, leading to a much lesser degree of response shift. Such evidence was also supported by clinical impressions and comments from most of the patients during the telephone survey, who generally expressed a view that their health status had not improved substantially between 6 and 18 months after TKR [34]. The differing magnitude of response shift at baseline and 6 months also suggests that recall bias was not a major factor influencing the results of this study, given that a similar magnitude of response shift would be expected if recall bias were present.

The interview during telephone survey also provided some other important information for study design and interpretation of results. The discrepancies between quantitative and qualitative assessment of response shift further suggested that patients experienced response shift in an unconscious manner. Therefore, it is necessary to determine both pre- and then-test scores for reference purposes. It was also found that the SF-6D was preferred over the EQ-5D by 51 patients to assess health status, as the three-level EQ-5D items were felt by subjects to be less accurate of a description of three levels of health states when compared with the SF-6D, which had five to six levels for various items. Such information could partially explain the discrepancies between the two measures in detecting the response shift at baseline. Last but not the least, the moderate agreement between the SF-6D and the EQ-5D in measuring response shift suggested that it was feasible to detect response shift with both measures. The systematic difference between the two measures, however, raises a note of caution regarding the accurate quantification of response shift in HRQOL and utility assessment.

We also recognize some limitations of the current study. First, the small sample size (n = 74) prevented us from generalizing our results to all patients undergoing TKR. For the same reason, the factors identified in MLR analyses should be considered as exploratory rather than confirmatory. Second, as the EQ-5D pretest data at 6 months were not available, we were not able to study the agreement between two measures for this time point to provide more robust results. Third, although the then-tests were taken at 18 months to minimize the difference in the magnitude of recall bias, recall bias might still have affected the results of the "then-test" in our study to a certain extent because the retrospective design of the then-test approach is known to be susceptible to recall bias [35]. According to the proposed guideline by Schwartz and Sprangers (2010) on "improving the stringency of response shift research use the then test" that was published after the completion of our study, this exploratory study may also have some other limitations. For example, we did not include clinician-based measures in the study to assess patient's health states. We were also not able to utilize statistical approaches to estimate recall bias and the implicit theory of change, as we did not have an adequate sample size to conduct meaningful analysis. Also, we did not perform the psychometric tests of the then-tests, such as construct validity and stability [35].

In conclusion, response shift was present and impacted HRQOL and utility assessment among patients undergoing TKR both just prior to and 6 months after surgery. This suggests that HRQOL and utility evaluations should be performed bearing in mind the potential changes in patients' internal standards that lead to response shift.

Acknowledgments

The authors thank Chong Hwei Chi and William Yeo from the Orthopaedic Diagnostic Center, Singapore General Hospital for assistance in study coordination.

Source of financial support: Funding for this study was provided by the National University of Singapore. Publication of the study results was not contingent upon sponsor's approval.

REFERENCES

[1] Golembiewski RT, Billingsley K, Yeager S. Measuring change and persistence in human affairs: types of change generated by OD designs. J Appl Behav Sci 1976;12:133–57.
[2] Zwiebel A. Changing educational counsellors' attitudes toward mental retardation: comparison of two measurement techniques. Int J Rehabil Res 1987;10:383–9.
[3] Postulart D, Adang EM. Response shift and adaptation in chronically ill patients. Med Decis Making 2000;20:186–93.
[4] Wilson IB. Clinical understanding and clinical implications of response shift. Soc Sci Med 1999;48:1577–88.
[5] Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999; 48:1507–15.
[6] Ahmed S, Mayo NE, Corbiere M, et al. Change in quality of life of people with stroke over time: true change or response shift? Qual Life Res 2005;14:611–27.
[7] Breetvelt IS, Van Dam FS. Underreporting by cancer patients: the case of response-shift. Soc Sci Med 1991;32:981–7.
[8] Schweickhardt A, Larisch A, Fritzsche K. Differentiation of somatizing patients in primary care: why the effects of treatment are always moderate. J Nerv Ment Dis 2005;193:813–9.
[9] Ahmed S, Mayo NE, Wood-Dauphinee S, et al. Using the Patient Generated Index to evaluate response shift post-stroke. Qual Life Res 2005;14:2247–57.
[10] Howard GS, Ralph KM, Gulanick NA, et al. Internal invalidity in pretest–posttest self-report evaluations and a re-evaluation of retrospective pretests. Appl Psychol Meas 1979;3:1–23.
[11] Ahmed S, Mayo NE, Wood-Dauphinee S, et al. Response shift influenced estimates of change in health-related quality of life poststroke. J Clin Epidemiol 2004;57:561–70.
[12] Bernhard J, Lowy A, Maibach R, et al. Response shift in the perception of health for utility evaluation: an explorative investigation. Eur J Cancer 2001;37:1729–35.
[13] NIH Consensus Statement on total knee replacement. NIH Consens State Sci Statements 2003;20:1–34.
[14] Thompson-Fawcett MW, Richard CS, O'Connor BI, et al. Quality of life is excellent after a pelvic pouch for colitis-associated neoplasia. Dis Colon Rectum 2000;43:1497–502.
[15] Oort FJ, Visser MR, Sprangers MA. An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. Qual Life Res 2005;14:599–609.
[16] Roos EJ, de Graeff A, van Eijkeren MA, et al. Quality of life after pelvic exenteration. Gynecol Oncol 2004;93:610–4.
[17] Razmjou H, Yee A, Ford M, Finkelstein JA. Response shift in outcome assessment in patients undergoing total knee arthroplasty. J Bone Joint Surg Am 2006;88:2590–5.
[18] Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 2004;13: 873–84.

[19] Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd ed.). New York: Erlbaum, 1988.
[20] Xie F, Li SC, Lo NN, et al. Cross-cultural adaptation and validation of Singapore English and Chinese versions of the Oxford Knee Score (OKS) in knee osteoarthritis patients undergoing total knee replacement. Osteoarthritis Cartilage 2007;15:1019–24.
[21] Sprangers MA, Schwartz CE. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med 1999;48:1531–48.
[22] Brossart DF, Clay DL, Willson VL. Methodological and statistical considerations for threats to internal validity in pediatric outcome data: response shift in self-report outcomes. J Pediatr Psychol 2002;27: 97–107.
[23] Brazier J, Usherwood T, Harper R, et al. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol 1998; 51:1115–28.
[24] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271–92.
[25] Wee HL, Cheung YB, Fong KY, et al. Are English- and Chinese-language versions of the SF-6D equivalent? A comparison from a population-based study. Clin Ther 2004;26:1137–48.
[26] Rabin R, Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;33:337–43.
[27] Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
[28] Luo N, Chew LH, Fong KY, et al. Do English and Chinese EQ-5D versions demonstrate measurement equivalence? An exploratory study. Health Qual Life Outcomes 2003;1:7.
[29] Timmerman AA, Anteunis LJ, Meesters CM. Response-shift bias and parent-reported quality of life in children with otitis media. Arch Otolaryngol Head Neck Surg 2003;129:987–91.
[30] Ring L, Hofer S, Heuston F, et al. Response shift masks the treatment impact on patient reported outcomes (PROs): the example of individual quality of life in edentulous patients. Health Qual Life Outcomes 2005;3:55.
[31] Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–15.
[32] Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 2005;14:1523–32.
[33] Brauer CA, Rosen AB, Olchanski NV, Neumann PJ. Cost–utility analyses in orthopaedic surgery. J Bone Joint Surg Am 2005;87: 1253–9.
[34] Ethgen O, Bruyère O, Richy F, et al. Health-related quality of life in total hip and total knee arthroplasty: a qualitative and systematic review of the literature. J Bone Joint Surg Am 2004;86-A:963–74.
[35] Schwartz CE, Sprangers MA. Guidelines for improving the stringency of response shift research using the then test. Qual Life Res 2010;19: 455–64.