Quality of Life Research (2006) 15:1533–1550
DOI 10.1007/s11136-006-0025-9

The clinical significance of adaptation to changing health: A meta-analysis of response shift

Carolyn E. Schwartz[1,2,3], Rita Bode[4], Nicholas Repucci[5], Janine Becker[6], Mirjam A. G. Sprangers[7] & Peter M. Fayers[8,9]
[1]*DeltaQuest Foundation, Inc., Concord, MA, USA (E-mail: carolyn.schwartz@deltaquest.org);* [2]*Social Sectors Development Strategies, Inc., Boston, MA, USA;* [3]*Division of Palliative Medicine, Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA;* [4]*Department of Physical Medicine & Rehabilitation, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA;* [5]*Health Assessment Lab, Waltham, MA, USA;* [6]*QualityMetric Incorporated, Waltham, MA, USA;* [7]*Department of Medical Psychology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands;* [8]*University of Aberdeen Medical School, Aberdeen, UK;* [9]*Norwegian University of Science and Technology, Trondheim, Norway*

Accepted in revised form 22 May 2006

Abstract

Aims: When individuals experience changes in their health states, they may alter their internal standards, values, or conceptualization of quality of life (QOL). Such 'response shifts' can affect or distort QOL outcome measurement, which is of particular concern when evaluating medical or psychosocial interventions. Although clinicians and researchers acknowledge the occurrence of response shifts, little is known about the magnitude and clinical significance of those effects. To fill this gap in knowledge about response shift phenomena, we performed a meta-analysis on published QOL articles on response shift. *Methods:* Extensive literature searches and multiple contacts with researchers yielded a collection of 494 articles for potential reviewing. We retained only published longitudinal studies that measured response shift, resulting in 26, of which 19 reported the requisite data for computing an effect size (ES). We calculated and compared the ESs for each study with regard to potential moderator variables: the QOL domains measured, disease group investigated, sample size, and response shift method used. We rated studies for quality to allow ES weighting. *Results:* When we examined ES absolute values, we found that ES magnitude was small, with the largest ESs detected for fatigue, followed by global QOL, physical role limitation, psychological well-being, and pain (mean $|ES_{weighted}|$ = 0.32, 0.30, 0.24, 0.12, and 0.08, respectively). ESs varied considerably in direction. Aggregating raw ES scores over all studies led to positive and negative values canceling each other out (mean directional $ES_{weighted}$ = 0.17, 0.02, −0.01, 0.06, and 0.02, respectively). We found little evidence of an effect for the moderator variables examined. *Conclusions:* A definitive conclusion on the clinical significance of response shift cannot currently be drawn from existing studies. For a number of reasons, ES estimates were primarily based on then-test results, a method that is not without criticism, such as its susceptibility to recall bias. We recommend a standardized approach for reporting results of future response shift research to advance the field and to facilitate interpretation and comparisons across studies.

Key words: Adaptation, Clinical significance, Meta-analysis, Quality of life, Response shift

Introduction

Clinicians and researchers commonly recognize that people change their internal standards, values or conceptualization of quality of life (QOL) when they experience changes in health. Nonetheless, researchers and clinicians tend to neglect or even ignore such 'response shift' phenomena, when measuring QOL in repeated assessments of a scale or item. In the past decade, however, we have witnessed an increase in the amount of interest in understanding and exploring response shift phenomena, as evidenced by over 100 articles published since 1999 referring to or measuring response shifts. In reviewing this literature, however, one notices that the clinical importance of response shift phenomena has not yet been systematically addressed. Synthesizing the research findings to date can be important for understanding the relations and implications of research findings over a broad research area [1]. The purpose of the present work is a meta-analysis to evaluate the magnitude and clinical significance of response shifts across published QOL studies, and to provide guidelines to advance future QOL measurement in this respect.

Methods

We performed the meta-analysis according to the Cochrane recommendations [2] using the following six steps: (1) Literature search; (2) Selection of eligible articles; (3) Quality Rating of the articles; (4) Extraction of data for effect size (ES) computations; (5) ES computation and interpretation; (6) Moderator analyses to examine associations between detected ESs and specific study parameters.

Literature search

We sought to identify all empirical research related to QOL and response shift for inclusion in the meta-analysis. We used two strategies for retrieving relevant articles: (A) an extensive literature databank search and (B) a comprehensive survey of QOL researchers. For the literature search, we queried the widely-used online Ovid platform to screen for articles containing the key words 'quality of life,' 'response shift,' 'framing,' 'frame of reference,' 'change,' and 'shift.' Ovid automatically searches the following databases: *Medline*® [In-Process & Other Non-Indexed Citations]; *Old Medline*® [1950–1965]; *Journals@Ovid*, *PsycINFO*® [1872 to present]; the *Cumulative Index to Nursing & Allied Health Literature* (CINAHL) [1982 to November Week 3 2004]; *Medline*® [1966 to Present with Daily Update]; and all *Evidence Based Medicine (EBM) Reviews* [1991 to the 4th Quarter of 2004]. EBM reviews included the ACP Journal Club, the Cochrane Central Register of Controlled Trials (CCTR), the Cochrane Database of Systematic Reviews (Cochrane DSR), and the Database of Abstracts of Reviews of Effects (DARE).

We emailed a survey inquiring about relevant QOL research on response shift to those QOL researchers who had either (a) published abstracts or full-length papers on response shift found in the literature search, or (b) personally communicated with one of the authors (either CES or MAGS) over the past decade for consultation or collaboration on response shift research.

Selection of eligible articles

We first examined the resulting set of articles to assess their eligibility. To be eligible for inclusion in the meta-analysis, articles had to meet the following four requirements: (1) adequate description of empirical QOL measurement; (2) explicit measurement of response shift of QOL variables, (3) longitudinal assessment of QOL data; and (4) publicly available as a full-length published article (i.e., published abstracts were excluded).

Quality rating of the articles

Three trained reviewers (CES, JB, NR) then rated the quality of the studies using the eight criteria described below. Our internal training ensured that we consistently applied criteria across reviewed articles. For each criterion a study scored '1,' if the criterion was met and '0' if not. Each reviewer performed an initial rating independently, and we then discussed all of the ratings to come to a consensus. When the salient information was absent, a score of '0' was given. Thus we treated missing data as if the quality criteria were not met. We used the following quality criteria:

1. *Sample size*. The sample size reported in the article is large enough to detect a large ES with an alpha level of 0.05 (using Cohen's [2] criteria, such as n = 26 per group for comparing means, n ≥ 200 for Structural Equation Modeling, etc.).
2. *Response rate*. The response rate at baseline is reported in the published article as ≥70%.
3. *Control*. A control or comparison group is described.
4. *Randomized*. The described study is randomized and the analysis is done by treatment arm.
5. *External criterion*. The study involves a clinical criterion variable (e.g., the Barthel Index for stroke rehabilitation patients).
6. *Tool psychometrics*. The QOL measure(s) used in the studies is/are well-established, reliable and valid.
7. *Planned comparisons*. The study reports hypotheses and an analyses plan for hypotheses testing (i.e., not *post hoc*).
8. *Type I error rate*. The ratio of significant findings to the number of comparisons is ≥0.10 (i.e., the statistical significance is not likely to be caused by chance, taking into account the increase in type I error rate due to multiple comparisons).

We calculated the weighted mean ES per domain using standard meta-analyses principles: we weighted each study according to its quality rating multiplied by the inverse variance of its ES estimate.

Extracting data for ES computations

Many studies used several QOL measures that varied in popularity and psychometric quality. To perform the meta-analysis efficiently, we focused on extracting outcome data on: (a) five major QOL domains that were commonly assessed across studies (global QOL, fatigue, psychological well-being, pain, and physical role limitations), which were (b) measured by established and validated QOL instruments. We took this straightforward approach to maximize the chance of getting stable and meaningful ES estimates and to avoid calculating ES on a plethora of uncharacterized QOL instruments with questionable validity, which could obfuscate the interpretation of the meta-analyses results.

When extracting QOL data for ES computations from longitudinal studies using the 'then-test' for response shift measurement in more than two endpoints, we extracted only QOL data collected at baseline and the first time point (if it was at least 1 month). We used this strategy to minimize the potential bias of multiple comparisons, and to maximize the comparability of ES estimates across studies. For studies using more than one response shift method (e.g., then-test and anchor recalibration to assess recalibration response shift), we calculated ESs for as many response shift methods as data allowed to enable comparisons between different methods.

The vast majority of studies used the then-test method, which is also known as retrospective pretest–posttest design method. It asks respondents at the posttest to think back to how they were doing at baseline and to retrospectively rate their QOL at that time (or any construct(s) they already had rated at baseline). The method assumes that respondents will use their posttest internal standards when providing a re-evaluation or 'then-test' rating of their baseline score. Figure 1 depicts how difference scores are then used to compute a 'response shift' (then-minus-pre) score.

Not only did almost all studies use the then-test, but only then-test studies consistently provided the requisite data for ES computation. For example, response shift studies using individualized measures, such as the Patient-Generated Index, the Schedule for the Evaluation of Individual QOL (SEIQOL), or qualitative interviews, did not provide the data necessary for calculating ES. Thus, we did not include these studies in the meta-analysis. Additionally, we also did not include studies reporting on samples that had been reported in previously published work, to avoid the bias of over-representing one particular sample. Consequently, one study reporting on a structural equation modeling approach [3] was excluded because the same sample had been used in another publication using the then-test.

ES computation and interpretation

Most studies estimated response shift from the comparison of the baseline score ('pre'-test) against a retrospective-pretest score ('then'-test),

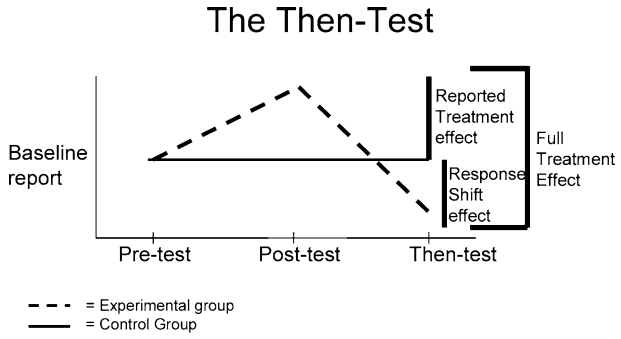

Figure 1. Then-test design method showing a hypothetical trial with experimental group (dashed lines) and control group (solid line).

and reported the raw difference 'pre' minus 'then'. While the standardized response mean is arguably a better indicator of the magnitude of effects, the data needed for its computation were unavailable in most of the articles reviewed. Thus we used the ES, which is defined as the mean difference between tests (i.e., 'then' minus 'pre') divided by the standard deviation (SD) of the pre-test. We used the original population estimate (i.e., baseline SD) to avoid violating the assumption of independence (since the two samples are correlated) [4]. SEM approaches were not included in the calculation of average ESs, but only for illustrative purposes, due to sample overlap with other studies included or because the same study used multiple methods and we included ES from only one method per study to reduce the multiplicity of endpoints.

For calculating the standard error (SE) of the ES values, we required the correlation between 'pre' and 'then.' Since the SE was rarely reported, we assumed a correlation of 0.5 because this association is typical of values reported for repeated QOL measurements. We adopted the usual scoring convention for QOL measures, so that high scores mean 'good' QOL and low QOL scores indicate 'poor' QOL. The scoring direction needed special attention when computing ES scores, because we wanted to maximize the comparability of ES estimates across studies.

For standardized interpretation, we relied on Cohen's criteria [2], and considered an ES >0.2 small, >0.5 moderate, and >0.8 large. Cohen's guidelines were intended to provide a rule of thumb for the magnitude of an ES rather than its clinical significance, and researchers in the field of QOL have been actively debating which of these magnitudes is equivalent to 'clinical significance.' A recent editorial by Sloan and colleagues [5] concludes that 0.5 ES is a conservative estimate that is likely to be clinically meaningful. They note that the evidence to date suggests that all approaches of estimating clinical significance converge more than they diverge. Thus, defining a 0.5 ES to be a clinically significant effect is a good and defensible starting point.

We used SPSS v13 and STATA for data analyses. We measured heterogeneity with Cochrane's Q-statistic. We used random-effects models in the meta-analysis, which computes mean ES and 95% confidence intervals (CI). For the estimates of ES that were plotted in the 'forest plots,' we combined studies using a weighting of study-quality multiplied by the inverse variance of study ESs. The forest plots show the mean estimates of ES, with 95% CI.

Moderator analyses

We investigated the aggregated ES statistics across all studies by QOL domain, and by the following study variables that we had hypothesized to be potential moderators of ES: sample size, design (observational vs. randomized), disease group studied (cancer, other seriously ill patients [e.g., AIDS, dialysis patients, hospice patients, and liver transplant recipients], neurological disease, and primary care), and methods used to evaluate response shift (e.g., then-test [6], structural equation modeling [3, 7]). Because the Oort [7] and Visser et al. [8] studies used the same patient samples, only the Visser et al. [8] estimates were included in the mean ES estimates [8]. We retained Oort (2005)

ES in the Forest Plots to show the difference in estimated ES by method. Visser used multiple methods and only the then-test was retained for ES calculation; Oort used SEM only.

We hypothesized that larger studies, randomized trials, and then-test studies would yield smaller ES, and that studies of cancer or other seriously ill patients would yield higher ESs due to salient and rapid health state changes in the context of life-threatening disease. To assess associations between the moderator variables and ES estimates, we used meta-analysis regression separately by domain.

Results

Sample

A total of 494 potential articles for the meta-analysis was found: 454 of them were revealed in the Ovid literature search and 40 were retrieved by QOL researchers input. Fifty-seven QOL researchers participated in the email survey, of whom 42 responded (74%), among the 40 papers they provided were: 28 published papers, 3 published abstracts, 9 manuscripts under review or in press, and 9 ongoing studies. Of the total 494 articles, 373 remained after removing duplicates. By applying the four eligibility criteria described above, we retained 28 articles for quality rating and meta-analysis. However, we were only able to extract data for ES calculation from 19 out of these 28 articles, because 9 did not report sufficient data for computing ESs.

The retained studies included: one study by Adang [9]; three studies by Ahmed et al. [10–12]; one study by Bernhard et al. [13]; Jansen et al. [14]; Joore et al. [15]; Lepore et al. [16]; Oort et al. [3]; Rapkin [17]; and Rees et al. [18]; four studies by Schwartz et al. [19–22]; one study by Sprangers et al. [6] and Timmerman et al. [23]; and two studies by Visser et al. [8, 24]. See Figure 2 for Exclusion Tree and Appendix for list of studies considered for inclusion in meta-analysis.

Those 19 studies were published in North America, Europe, the Middle East, and Asia between 1998 and 2005. These response shift studies addressed domains generally considered important in health-related QOL research. Most of them included an indicator of global QOL, followed by specific QOL domains like well-being, physical role limitations, fatigue, and pain. The studies were heterogeneous in patient groups, including both chronic and terminal conditions. Sample sizes ranged from small to moderately large (n = 21 to n = 199). The studies generally had sound baseline and follow-up rates that would not lead one to

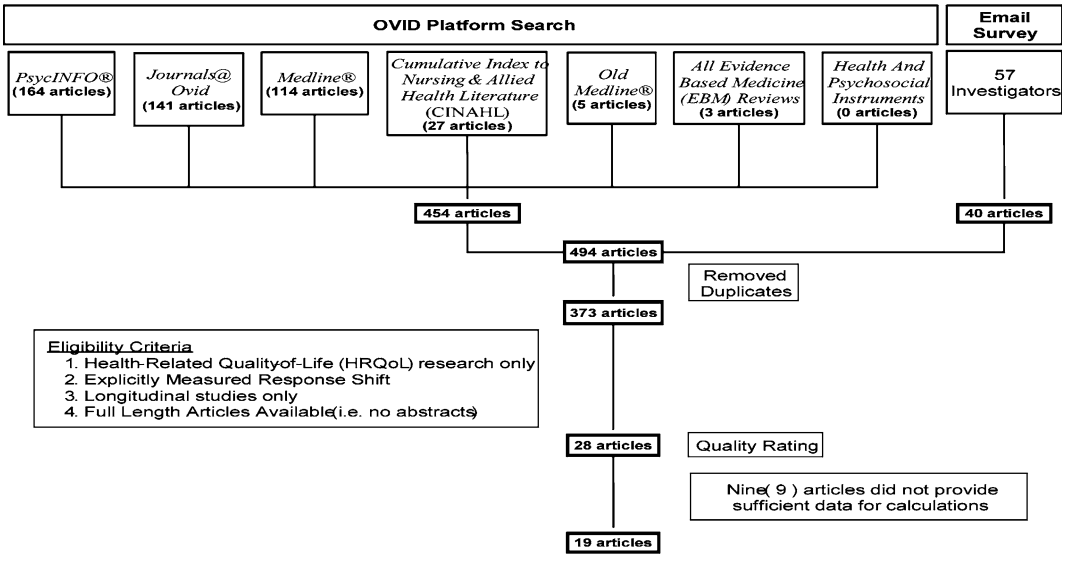

Figure 2. Exclusion tree for response shift studies.

suspect bias, although three of the 19 studies had follow-up response rates below 70% (see Appendix, column labeled 'Patient Population and Response Rate' for details).

Magnitude of ES

The magnitude of the ESs varied by QOL domain. ES was largest for fatigue, followed by global QOL, physical role limitations, psychological well-being, and then pain (aggregate mean |ES| = 0.31, 0.30, 0.23, 0.12, and 0.07, respectively; Table 1 shows directional ESs from which these absolute value ESs were computed). These ESs can be considered small, using Cohen's criteria [2]. Figure 3a–e provide a graphic representation of the raw ESs of the five QOL domains investigated in this meta-analysis. The 'forest plots' show the weighted mean estimates of ES with 95% CI.

An examination of the ESs presented in Figure 3a–e reveals that the ESs varied considerably in direction. Indeed, there was considerable and statistically significant ($p < 0.001$) evidence of heterogeneity for all domains. Aggregating ESs across all studies led to positive and negative values cancelling each other out (mean directional $ES_{weighted}$ = 0.17, 0.17, −0.01, 0.06, and 0.02, for fatigue, global QOL, physical role limitations, psychological well-being, and pain, respectively).

Moderators of ES

We examined five potential moderators of ES: sample size, quality rating, study design, disease group, and response shift method (Tables 1 and 2). We found little evidence of a moderator effect in these analyses. After adjusting for multiple comparisons, the only significant difference we found was for 'method in the domain physical role limitations' ($p < 0.001$), and this analysis compared the ESs for six then-test studies to one study that used a different method.

Discussion

This meta-analysis systematically reviewed response shift phenomena in QOL research for the first time. The study reveals a substantial body of literature on response shift phenomena

Table 1. Random effects model estimates of weighted effect size (wtd ES) with 95% CI

	N	Weighted mean ES	CI		N	Weighted mean ES	CI		N	Weighted mean ES	CI		N	Weighted mean ES	CI		N	Weighted mean ES	CI	
Sample size																				
n < 100	6	−0.13	−0.75	0.50	3	0.53	0.16	0.89	5	0.04	−0.34	0.41	2	0.12	−0.25	0.50	4	0.07	−0.34	0.48
n > 100	6	0.14	−0.15	0.42	4	−0.08	−0.35	0.19	4	0.08	−0.01	0.17	4	−0.03	−0.26	0.21	3	−0.09	−0.35	0.16
Study design																				
Randomized	3	−0.05	−0.62	0.52	2	−0.27	−0.75	0.22	3	0.08	−0.06	0.23	2	−0.20	−0.51	0.11	3	−0.22	−0.34	−0.09
Observational	9	0.05	−0.28	0.37	5	0.34	0.07	0.61	6	0.05	−0.22	0.31	4	0.13	−0.02	0.27	4	0.16	−0.17	0.49
Method																				
Then-test	8	−0.08	−0.34	0.17	7	0.17	−0.12	0.45	7	−0.03	−0.15	0.09	5	0.01	−0.22	0.24	6	−0.12	−0.26	0.03
Other	4	0.26	−0.38	0.90	0	–	–	–	2	0.44	−0.32	1.19	1	0.07	−0.10	0.24	1	0.73	0.48	0.98
Disease group																				
Cancer	7	0.16	−0.12	0.44	6	0.07	−0.19	0.32	5	−0.02	−0.19	0.15	4	0.03	−0.26	0.32	4	−0.12	−0.32	0.08
Neurological	2	−0.52	−1.00	−0.05	1	0.76	0.52	0.99	1	−0.03	−0.23	0.18	0	–	–	–	1	−0.24	−0.45	−0.03
Other serious	2	−0.47	−1.60	0.66	0	–	–	–	1	0.06	−0.11	0.23	1	0.07	−0.10	0.24	0	–	–	–
Primary care	1	1.09	0.81	1.37	0	–	–	–	2	0.35	−0.57	1.28	1	−0.05	−0.25	0.15	2	0.37	−0.34	1.07
All studies	12	0.02	−0.24	0.29	7	0.17	−0.12	0.45	9	0.06	−0.11	0.23	6	0.02	−0.17	0.20	7	−0.01	−0.23	0.22

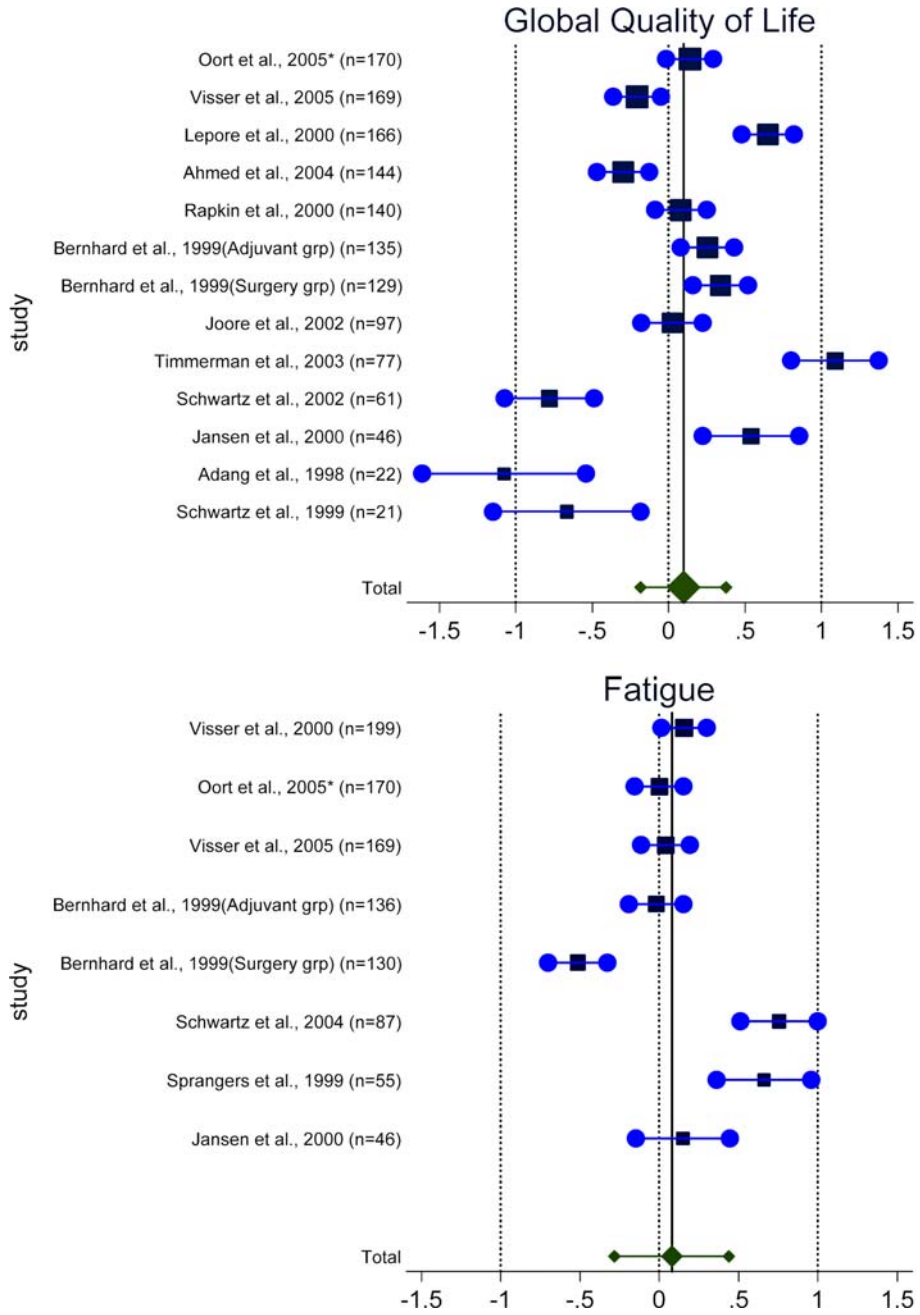

Figure 3. Forest plots showing weighted ESs and CI for global QOL [Mean |ES| = 0.30; Mean directional ES = 0.02] (3a), fatigue [Mean | ES| = 0.32; Mean directional ES = 0.17] (3b), psychological well-being [Mean |ES| = 0.12; Mean directional ES = 0.06] (3c), pain [Mean |ES| = 0.08; Mean directional ES = 0.02] (3d), and physical role limitations [Mean |ES| = 0.24; Mean directional ES = 0.01] (3e), from most influential (at top of *y*-axis) study to least (at bottom of *y*-axis) on the basis of sample size (i.e., inverse-variance). *Oort et al. [3] was not included in the mean calculations because of sample overlap with Visser et al. [8]. It is included in the forest plots to illustrate how method affected ES estimates. In these plots, Oort's et al. [3] estimate is based on Structural Equation Modeling, and Visser's et al. [8] is based on the then-test.

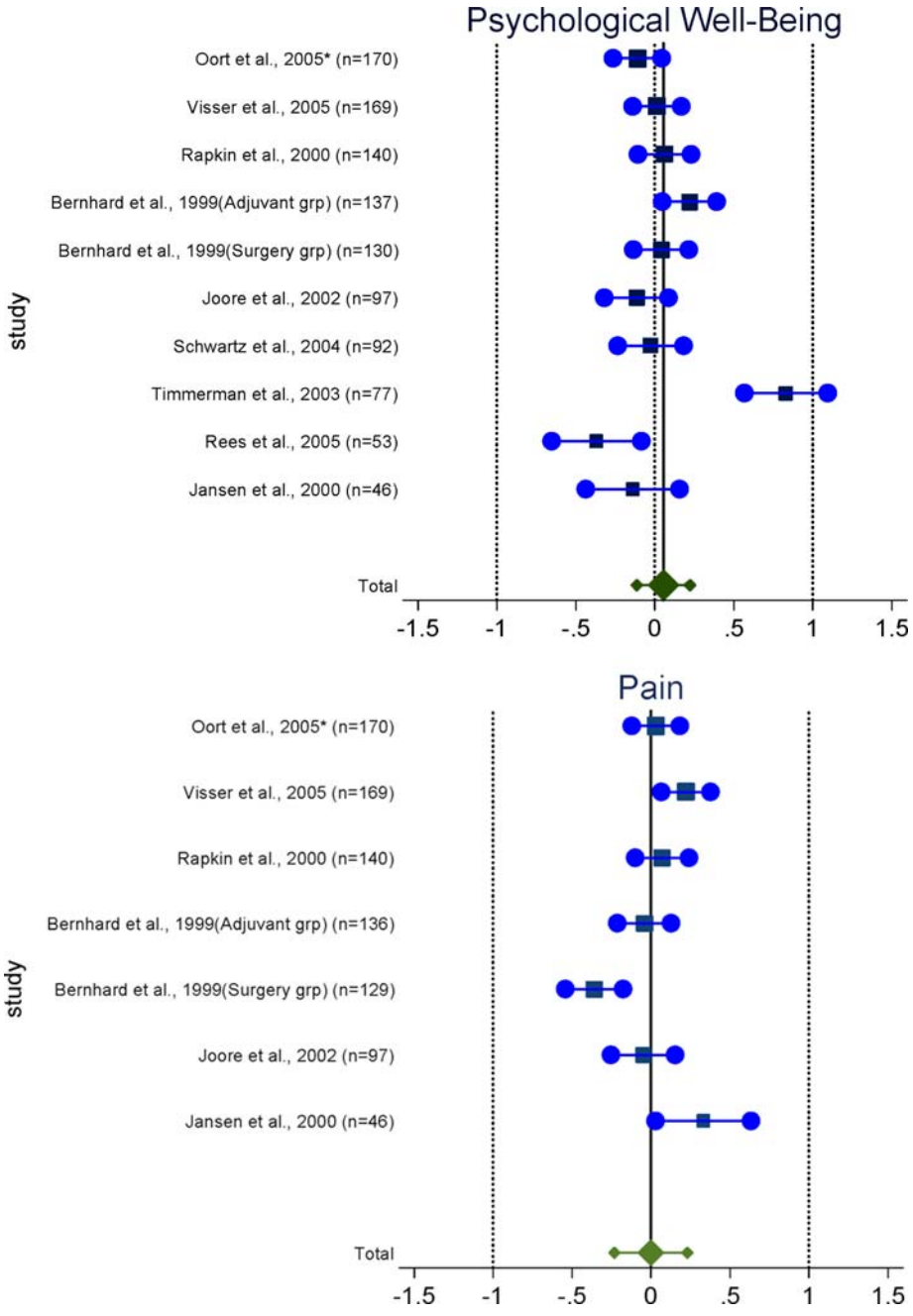

Figure 3. Continued

and one may tend to conclude that response shifts are a common and significant phenomenon in QOL measurement, implying that people adapt their internal standards of QOL in response to a changing health state. We found that overall the ESs of the response shift phenomena published to date are relatively small according to Cohen's [2] criteria. Even a small response shift may, however, result in an underestimation of the true QOL change, i.e., concluding that it is small when it is moderate, or moderate when it is large [3].

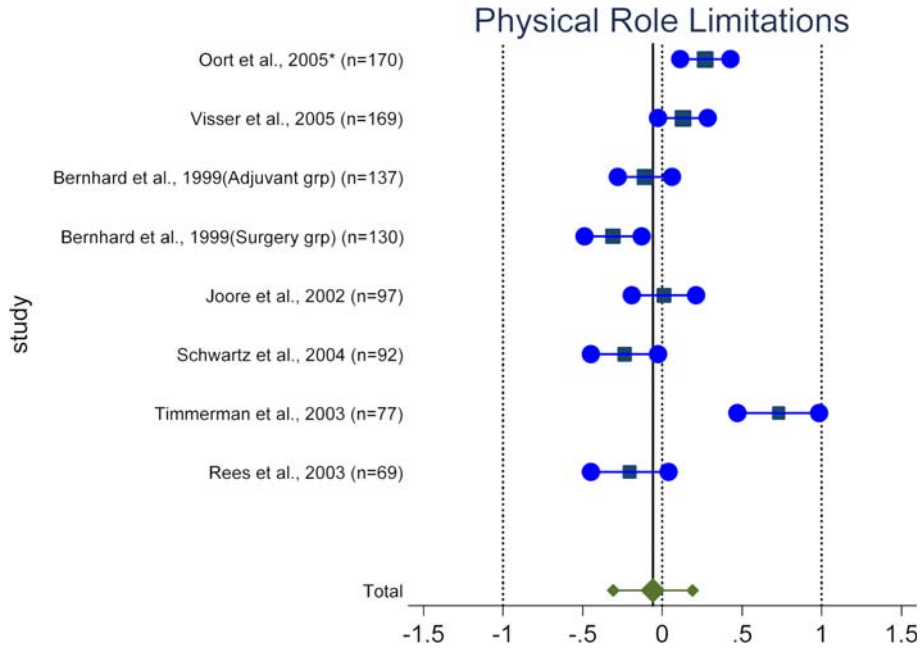

Figure 3. Continued

We found the largest ESs for fatigue, global QOL, and physical role limitations when considering the mean absolute ES. The results suggest that for example in cancer studies, patients with an objectively declining QOL may report no decrease in QOL due to a positive adaptation (response shift). Accounting for response shift would make the 'true' QOL change visible, which is of particular interest when evaluating the efficacy of cancer treatments, for example. Further, accounting for response shift is crucial for making QOL change scores comparable to change scores of other patient groups/diseases, and thus more standardized. Lastly, disentangling response shift in the sense of positive or negative adaptation to a disease may be useful information for clinicians in itself.

Alternatively, if a study reports only a slight increase in QOL for one patient sample, which is considered to be very healthy, the same amount of increase in QOL may be substantial for another less healthy sample, whose internal scaling has not yet been compressed due to a good health state.

Although most response shift research has focused on cancer to date, the growing number of studies addressing other patient groups documents that response shift is prevalent, salient in one or more QOL domains (i.e., fatigue, global QOL, physical role limitations), and was clinically important in some studies, although not in the

Table 2. Meta-analysis regression examining potential moderators of ESs

Moderator	Global QOL		Fatigue		Psychological well-being		Pain		Physical role limitations	
	N = 12		N = 7		N = 9		N = 6		N = 7	
	Coefficient	*p*	Coefficient	*p*	Coefficient	*p*	Coefficient	*p*	Coefficient	*p*
Sample size	0.0044	0.20	−0.0033	0.27	0.0011	0.70	−0.0013	0.65	−0.0015	0.71
Quality	−0.11	0.52	0.05	0.84	−0.06	0.60	−0.10	0.44	−0.24	0.07
Study design	0.09	0.83	0.61	0.03	−0.03	0.89	−0.20	0.03	0.38	0.10
Disease = cancer	0.32	0.38	−0.69	0.09	−0.21	0.32	0.01	0.95	−0.28	0.28

aggregate ES index. Overall, this meta-analysis showed considerable heterogeneous ES results across all QOL domains. The lack of homogeneity may simply reflect the considerable variation between populations, samples, and measures. It may also be caused by different response shift methods applied. The results of our meta-analysis regression suggest that response shift method may be a significant moderator in one domain (i.e., physical role limitations), although the evidence is marred by the fact that only one of the six studies compared used a method other than the then-test. Most studies evaluated in this meta-analysis used the then-test, which has not been without criticism as it may be susceptible to recall bias [10, 21]. There is, however, some research that disputes this susceptibility to recall bias and putatively demonstrates the convergent validity of the then-test [8].

Another possible explanation for heterogeneity could be that the magnitude of response shift depends on the direction and/or magnitude of the true mean change of the outcomes. This information was not reported for most of the studies. Response shifts of different sign may have a similar meaning when comparing patient groups who deteriorated or improved. For example, patients lowering their standards (i.e., then-minus-pre is a negative value) as a result of health state deterioration could be considered conceptually similar to patients raising their standards (i.e., then-minus-pre is a positive value) as a result of health state improvement. Such results might be two faces of the same phenomenon as interpreted by set point theory [25] or prospect theory [26]. If, however, two studies studying similar circumstances (e.g., health state deterioration) reveal different ES directions, then the sign of the ESs is relevant and would lead one to question the validity of the findings.

The limitations of this meta-analysis should be noted. A notable limitation is that we relied on information provided in the published papers for calculating ESs and for interpreting them. Most papers did not provide adequate information for calculating ESs (e.g., mean, SD) or for interpreting the findings (e.g., Were patients' health states improving or deteriorating?). This limitation was a notable constraint that made it difficult to determine whether results could be combined meaningfully. For this reason, we present both combined absolute values of ESs and combined raw values of ESs. The two presentations yield very different conclusions. In the former, estimates of response shift ES are close to the half-SD benchmark [27]. In contrast, when raw values are combined, the ES distribution appears to be normal with a mean weighted ES that is close to zero. This latter presentation thus suggests no overall response shift effect. Future research should provide as much information as possible for ES computation and interpretation.

Additionally, we could not correct for multiple comparisons that may have been made prior by authors selectively reporting statistically significant results. We also do not know the item characteristics of the various tools used and of the then-test versions of the tools, although we have standardized the metric of the comparisons so that higher scores indicate better functioning. Thus this meta-analysis provides no more than an indication of the possible range of ES values. Finally, our ES estimates are based on studies that used a range of outcomes tools and response-shift methods. Studies that used generic QOL measures would be more likely to report smaller ESs because generic measures are generally less sensitive than disease-specific measures [28].

Imprecise response shift measurement may also be a caveat of the present work. Although clearly a dominant method in response shift research, the then-test approach has considerable limitations. The method may be negatively affected by recall bias [10, 21], and may contain a substantial portion of noise [29, 30] which reduces its power and interpretability. Future research should include more than one response shift method assessing recalibration so that it is possible to compare the magnitude of ESs by method. Future research should focus on understanding the then-test method better to elucidate what patients are thinking when they answer then-test items (e.g., via cognitive interviewing). Further, clear interpretation guidelines for the approach need to be provided (e.g., what does a negative score reflect in tangible terms about how a person recalibrates their QOL appraisal?) to aid other investigators in interpreting then-test study findings.

Perhaps the most salient limitation of the present work is that it is difficult to state the findings in a simple sentence. For example, we cannot state 'response shifts lead to larger (or smaller) estimates of effects.' If response shift methods used were more directly comparable (i.e., directionality was clearly interpretable) then such a statement would be possible. This caveat also makes it difficult to know what the implications are for 'controlling for' response shift, teaching response shift, and developing more qualitative methods to assess response shift so that we understand the phenomenon better.

Conclusions and recommendations

A definitive conclusion on the clinical significance of response shift cannot currently be drawn from existing studies. This uncertainty is due to the heterogeneity of ESs as well as of studies and patient populations included, missing information that would allow for ES calculation from a larger sample of studies and from studies which include measures other than the then-test, and a current inability to aggregate ESs such that the directionality of the estimate has the same meaning across samples (e.g., a positive or negative recalibration response shift). Further research is warranted. Our experience implementing this systematic review suggests that the field of response shift research would benefit from clear reporting standards for published articles to facilitate comparisons across studies.

We recommend the following reporting standards for future response shift publications:

(1) providing statistics like the *mean*, *SD, and standardized response mean (SRM)* for *all* outcomes measured used for calculating ESs;
(2) defining the *scoring direction* of the outcome tool(s) used or (ideally) rescaling the scores in a standard manner across tools (e.g., higher scores reflect better QOL);
(3) reporting information on and (ideally) fulfilling all criteria we listed as '*quality criteria*' (sample size, baseline and follow-up response rates, control group, randomization, external criterion, tool psychometrics, no *post hoc* analyses, type I error rate);
(4) explaining the *meaning* of the study results in terms of changing internal standards, values, or conceptualization of QOL (i.e., recalibration, reprioritization, and reconceptualization response shifts)and (ideally) giving *clear interpretation guidelines* for the response shift approach and results.

Future research attempts to explore reasons (i.e., moderators) for the ES heterogeneity. If the suggested response shift reporting standards for future publications would be followed, then future research would be well-positioned to systematically advance our knowledge on response shift phenomena.

This is not only of interest for researchers but also of particular interest for clinicians, because response shift phenomena first need to be understood for their clinical significance before the knowledge can be transferred to the clinicians' daily practice. Potential benefits of stable and empirically supported response shift knowledge for the patient would be that the clinician, who administers his/her QOL using a specific QOL tool may know the potential response shift, which may occur and how to disentangle the 'objective' QOL change from the patient's adaptation to the disease/treatment. This may substantially help evaluating treatment efficacy and patients' adaptation.

Appendix A

Table A.1 Response shift studies considered in the meta-analysis review

Study (reference)	Study design	Response shift method(s) used	Outcome assessed	Patient population and response rate[a]	Quality rating	Reason for exclusion
Adang et al. [9]	Longitudinal prospective observational study	Then-test	Recalibration in overall QOL, using VAS (10-point scale)	N = 22 pancreas–kidney transplant patients	3	
Ahmed et al. [10]	Longitudinal observational study; Randomized clinical trial	Then-test	Recalibration in perceived health status: using VAS of the EQ-5D	N_{total} = 146–154 stroke patients (N_{stroke}/baseline evaluation = 146, N_{stroke}/6-week evaluation = 148, N_{stroke}/24-week evaluation = 154) and $N_{caregivers}$ = 50 controls; [N_{stroke}: 58% m; 42% f; mean age: 71 yrs; $N_{controls}$: 20% m, 80 f; mean age: 61 yrs]. Baseline response rate not provided; 100% follow-up rate	4	
Ahmed et al. [11]	Longitudinal observational study; Randomized controlled trial	Confirmatory factor analysis (individualized measure of HRQOL)	Recalibration in overall QOL, psychological well-being etc. using SF-36	N = 238 patients with stroke; N = 392 controls [61%/33% m, 39%/67% f; Mean age: 67/62 (SD = 13/12)]	6	Data not available for ES computation
Ahmed et al. [12]	Randomized clinical trial [Short follow-up study (6/24 weeks)]	Qualitative method (changes in number and weight of domain of Patient Generated Index (PGI)	Reprioritization and reconceptualization measured by PGI and semi-structured interview	N = 92 stroke patients during the first 6 months of recovery [61% m, 39% f; Mean age: 69 yrs (SD = 15)]	7	PGI data cannot be used to calculate ES
Bar-on et al. [31]	Longitudinal observational study	Qualitative method	Subjective QOL measured by semi-structured interview	N_{total} = 450 [$N_{hypertensives}$ = 295, $N_{normotensives}$ = 155; Only males]	3	Excluded 25% of patients from analysis so severe biases likely
Bernhard et al. [13]	Longitudinal observational study; Randomized clinical trial	Then-test	Recalibration in overall QOL, using LASA	N = 187 patients with colon cancer [59% m; 41% f; 57% <65; 43% ≥ 65] 74% response rate; 87% follow-up rate	5	

Bernhard et al. [32]	Longitudinal observational study; Randomized controlled trial	Then-test	Recalibration response shift in subjective health score; Perception of health for utility evaluation measured by Linear Analogue Self-Assessment (LASA)	N = 132 colon cancer patients undergoing adjuvant chemotherapy [59% m, 41% f; median age: 62 yrs; age range: 27–88 yrs]	5	Excluded because of sample overlap with Bernhard (1999)
Bernhard et al. [33]	Longitudinal observational study	Then-test; Linear regression models and multilevel analysis	Recalibration in overall QOL and range of disease- and treatment-related domains measured by LASA; Reconceptualization in overall QOL	N = 186 patients with colon cancer under chemotherapy [59% m; 41% f; 56% <65 yrs; 44% ≥ 65 yrs] 91% baseline response rate; 53% follow-up rate	6	
Cella et al. [34]	Longitudinal observational study	Then-test	Recalibration in QOL measured by Functional Assessment of Cancer Therapy (FACT-G) and a Global Rating of Change	N = 308 cancer patients [$\sim$50% f, mean age: 58.8 yrs, 1/4 ethnic minorities]	4	Did not provide data that could be used to compute ESs
Hagedoorn et al. [35]	Longitudinal observational study	Linear regression analysis	Self-reported quality of life (i.e., emotional and global QOL) and physical functioning as measured by the EORTC	N = 193 cancer patients under chemotherapy with complete data; N = 224 'significant others' [All >18 yrs, 55% f, 45% m; mean age: 50.5 (SD = 14.4)]	6	Did not provide data that could be used to compute ESs
Jansen et al. [14]	Longitudinal observational study	Then-test	Recalibration in overall QOL, psychological well-being, and pain measured by then-test items adapted from the SF-36 and the Rotterdam symptom checklist (RSCL)	N = 46 patients with early diagnosed breast cancer under radiotherapy with complete data [Only females, 60% housewives; mean age: 55 yrs (SD = 10) range of age: 28–77 yrs] 66% baseline response rate; 96% follow-up rate	4	
Jansen et al. [36]	Non-randomized clinical trial	Utility measure approach	Health states measured by VAS, chained Time Trade-Off (TTO) and chained Standard Gamble (SG)	N = 55 patients with breast cancer [Only females, 60% housewives; median: 57 yrs; range of age: 33–82 yrs]	3	Did not provide data that could be used to compute ESs

Table A.1 Continued

Study (reference)	Study design	Response shift method(s) used	Outcome assessed	Patient population and response rate[a]	Quality rating	Reason for exclusion
Joore et al. [15]	Longitudinal observational study	Then-test	Recalibration response shift in hearing-related generic and specific QOL measured by EQ-5D VAS and Audiological Disabilities Preference Index (ADPI)	N = 98 hearing impaired adults after hearing aid fitting [52% m, 48% f; Mean age: 67 yrs (SD = 12)]. Baseline response rate not given; 78% follow-up rate	4	
Lepore et al. [16]	Longitudinal observational study	Then-test; Ideographic assessment of personal goals (Rapkin & Fisher, 1992)	Recalibration using SF-12 and the Prostate Cancer Index (PCI); Reprioritization	N = 166 prostate cancer patients [Only men, mean age: 65 yrs]. Baseline response and follow-up rates not provided	5	
Oort et al. [3]	Longitudinal observational study	SEM	Recalibration, reprioritization, reconceptualization in QOL measured by SF-36. ESs provided as published in paper	N = 170 cancer patients undergoing invasive surgery [51.2% m, 48,8% f; Mean age: 57.5 yrs (SD = 14.1), range > 27–83 yrs]	3	Excluded because of sample overlap with Visser et al. [8]
Postulart and Adang [37]	Longitudinal observational study	Then-test	QOL in the face of adaptation of illness measured by VAS, Time Trade-Off (TTO) and Standard Gamble (SG)	N = 22 kidney–pancreas transplantation patients; N = 55 proxy student sample [81.8%/41.8 m, 18.2/58.2% f; Mean age: 39.9/ 26.7 yrs]	4	Excluded. Same sample as Adang [9]
Rapkin [17]	Longitudinal observational study	Ideographic assessment of personal goals (Rapkin and Fisher 1992); Multiple regression model	Reconceptualization and reprioritization using hierarchical regression: ΔR^2 for response shift effects (i.e., goal-by-catalyst interactions): 0.08, 0.06, 0.07, $p < 0.001$, 0.01, and 0.05, for global well-being, relative emotional, and relative pain, respectively. IFSA structured interview, SF-20, 64-item AIDS Checklist	N = 140 AIDS patients with follow-up data [87.2% m, 13.8% f: age range: 24–66 yrs, 50% ethnic minorities] 90% baseline response rate; 63% follow-up rate	6	
Rees et al. [38]	Longitudinal observational study	Then-test; SEIQOL	Recalibration using *t*-test comparing patients and controls; Lower urinary tract symptoms measured by the International Prostate Symptom Score (IPSS) and Symptom Problem Index (SPI)	N = 76 patients with prostate cancer; N = 17 controls [Only men, age range: 60–85 yrs] 100% baseline response rate; 89% follow-up rate	3	

Rees et al. [18]	Longitudinal observational study	Then-test	Recalibration of QOL measured by the Prostate Cancer Patient and Partner (PPP) questionnaire	N = 55 patients with prostate cancer; N = 41 partners [Only men; Mean age: 72.9 yrs]. Baseline response rate not given; 96% follow-up rate	4
Schwartz et al. [19]	Short-term longitudinal study	Then-test; ANOVA; Covariance analysis	Recalibration, reprioritization, reconceptualization of QOL as measured by SF-12 after psychosocial intervention	N = 22 young adult cancer patient survivors; N = 54 controls; Mean age = 22 yrs (SD = 3.5)	5
Schwartz et al. [20]	Short-term empirical psychometric validation study with retest	Preference change; Linear regression analysis	Reprioritization of treatment preference and goals of advance care planning, both measured by modified Emanuel and Emanuel Medical Directive	N = 168 seriously ill patients [60.1% m, 39.9% f; Mean age: 66.3 (SD = 14.6)]. Baseline response rate not provided; 91% follow-up rate	6
Schwartz et al. [21]	Longitudinal observational study	Then-test; Longitudinal factor analysis	Recalibration of QOL (i.e., physical role and well-being) measured by Expanded Disability Status Scale (EDSS), Multidimensional Assessment of Fatigue (MAF), Multiple Sclerosis Self-efficacy (MSSE) scale and the Sickness Impact Profile (SIP)	N = 93 multiple sclerosis patients; N = 39 non-participants [23%/11% m, 70%/28% f, Mean age: 43.0/42.7 yrs]. Baseline response rate not provided; 68% follow-up rate	5
Schwartz et al. [22]	Randomized controlled trial (short-term pilot trial)	Preference change	Reprioritization of treatment preference using modified Emanuel and Emanuel Medical Directive; Reconceptualization of beliefs and values using Beliefs and Values Questionnaire	N = 61 ambulatory geriatric patients [83.6% f, 16.4% m; Mean age: 80] 18% response rate; 92% follow-up rate	6

Table A.1 Continued

Study (reference)	Study design	Response shift method(s) used	Outcome assessed	Patient population and response rate[a]	Quality rating	Reason for exclusion
Sprangers et al. [6]	Longitudinal observational study	Then-test	Reprioritization of QOL measured by EORTC QLQ-C30; Response shift in MFI-20 fatigue scores; and Response shift in EORTC QLQ-C30 fatigue	$N_{baseline}$ = 105 breast or prostate cancer patients undergoing radiotherapy; N = 11undergoing response shift. [40% m, 60% f; median age: 63 yrs; range: 28–89 yrs] 83% response rate; 94% follow-up rate	5	
Timmerman et al. [23]	Short-term longitudinal study	Then-test	Reprioritization of QOL measured by parent-administered 6-item quality of life survey (OM-6). SRM provided in paper	N = 77 children with persistent otitis media under surgery [62.3% boys, 37.7% girls; Mean age = 24.6 (SD = 7.5)]. Baseline response rate not provided; 90% follow-up rate	3	
Visser et al. [24]	Longitudinal observational study	Then-test	Recalibration response shift in fatigue; Fatigue measured by a one-item rating scale	N = 199 cancer patients receiving radiotherapy [58% m, 42% f; Mean age = 64 yrs (SD = 13), 79% married] 81% response rate; 86% follow-up rate	4	
Visser et al. [8]	Longitudinal observational study	Anchor-recalibration, then-test, and SEM	Recalibration QOL measured by SF-36	N = 170 cancer patients undergoing invasive surgery [51.2% m, 48,8% f; Mean age: 57.5 yrs (SD = 14.1); range: >27–83 yrs]. Baseline response rate not provided; 100% follow-up rate	4	
Wyrwich and Tardino [39]	Longitudinal observational study	Qualitative method	QOL measured by interview	N = 41 outpatients with chronic disease [63% m, 37% f; median age >65.5 yrs]	3	No quantitative data available for ES computation

[a]If included in meta-analysis.

References

1. Mosteller F, Colditz GA. Understanding research synthesis (meta-analysis). Annu Rev Public Health 1996; 17: 1–23.
2. Cohen J. A power primer. Psychol Bull 1992; 112: 155–159.
3. 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(3): 599–609.
4. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates 1988.
5. Sloan JA, Cella D, Hays RD. Clinical significance of patient-reported questionnaire data: another step toward consensus. J Clin Epidemiol 2005; 58(12): 1217–1219.
6. Sprangers MA, Van Dam FS, Broersen J, et al. Revealing response shift in longitudinal research on fatigue – the use of the thentest approach. Acta Oncol 1999; 38(6): 709–718.
7. Oort FJ. Using structural equation modeling to detect response shifts and true change. Qual Life Res 2005; 14(3): 587–598.
8. Visser MR, Oort FJ, Sprangers MA. Methods to detect response shift in quality of life data: A convergent validity study. Qual Life Res 2005; 14(3): 629–639.
9. Adang EM, Kootstra G, Engel GL, van Hooff JP, Merckelbach HL. Do retrospective and prospective quality of life assessments differ for pancreas–kidney transplant recipients? Transpl Int 1998; 11(1): 11–15.
10. Ahmed S, Mayo NE, Wood-Dauphinee S, Hanley JA, Cohen SR. Response shift influenced estimates of change in health-related quality of life poststroke. J Clin Epidemiol 2004; 57(6): 561–570.
11. Ahmed S, Mayo NE, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Change in quality of life of people with stroke over time: True change or response shift? Qual Life Res 2005; 14(3): 611–627.
12. Ahmed S, Mayo NE, Wood-Dauphinee S, Hanley JA, Cohen SR. Using the Patient Generated Index to evaluate response shift post-stroke. Qual Life Res 2005; 14(10): 2247–2257.
13. Bernhard J, Hurny C, Maibach R, Herrmann R, Laffer U. Quality of life as subjective experience: Reframing of perception in patients with colon cancer undergoing radical resection with or without adjuvant chemotherapy. Swiss Group for Clinical Cancer Research (SAKK). Ann Oncol 1999; 10(7): 775–782.
14. Jansen SJ, Stiggelbout AM, Nooij MA, Noordijk EM, Kievit J. Response shift in quality of life measurement in early-stage breast cancer patients undergoing radiotherapy. Qual Life Res 2000; 9(6): 603–615.
15. Joore MA, Potjewijd J, Timmerman AA, Anteunis LJ. Response shift in the measurement of quality of life in hearing impaired adults after hearing aid fitting. Qual Life Res 2002; 11(4): 299–307.
16. Lepore SJ, Eton DT, Schwartz CE, Sprangers MAG. Response Shifts in Prostate Cancer Patients: An Evaluation of Suppressor and Buffer Models. 8395. Adaptation to Changing Health: Response Shift in Quality-of-Life Research. Washington, DC: American Psychological Association, 2000, pp. 37–51.
17. Rapkin BD. Personal goals and response shifts: Understanding the impact of illness and events on the quality of life of people living with AIDS 8400. In: Schwartz CE, Sprangers MAG (eds.), Adaptation to Changing Health: Response Shift in Quality-of-Life Research. Washington DC: American Psychological Association, 2000: 53–71.
18. Rees J, Clarke MG, Waldron D, O'Boyle C, Ewings P, MacDonagh RP. The measurement of response shift in patients with advanced prostate cancer and their partners. Health Qual Life Outcomes 2005; 3(1): 21.
19. Schwartz CE, Feinberg RG, Jilinskaia E, Applegate JC. An evaluation of a psychosocial intervention for survivors of childhood cancer: Paradoxical effects of response shift over time. Psychooncology 1999; 8(4): 344–354.
20. Schwartz CE, Merriman MP, Reed GW, Hammes BJ. Measuring patient treatment preferences in end-of-life care research: Applications for advance care planning interventions and response shift research. J Palliat Med 2004; 7(2): 233–245.
21. Schwartz CE, Sprangers MAG, Carey A, Reed G. Exploring response shift in longitudinal data. Psychol Health 2004; 19(1): 51–69.
22. Schwartz CE, Wheeler HB, Hammes B, et al. Early intervention in planning end-of-life care with ambulatory geriatric patients: Results of a pilot trial. Arch Intern Med 2002; 162(14): 1611–1618.
23. 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(9): 987–991.
24. Visser MR, Smets EM, Sprangers MA, de Haes HJ. How response shift may affect the measurement of change in fatigue. J Pain Symptom Manage 2000; 20(1): 12–18.
25. Carver CS, Scheier MF. Scaling back goals and recalibration of the affect system are processes in normal adaptive self-regulation: Understanding 'response shift' phenomena. Soc Sci Med 2000; 50(12): 1715–1722.
26. Kahneman D, Tversky A, Moser PK. Prospect Theory: An Analysis of Decision Under Risk. Reality in Action: Contemporary Approaches. New York NY: Cambridge University Press, 1990, pp. 140–170.
27. Norman G. Hi! How are you? Response shift, implicit theories and differing epistemologies. Qual Life Res 2003; 12(3): 239–249.
28. Obisesan O. The evaluation of upper respiratory tract infection symptoms to show the significance of developing a quality-of-life evaluation instrument for upper respiratory tract infections to assess respiratory disorder-related disability. Am J Ther 2005; 12(2): 142–150.
29. Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. Health Qual Life Outcomes 2004; 2: 16.
30. Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. Health Qual Life Outcomes 2004; 2: 14.
31. Bar-on D, Lazar A, Amir M. Quantitative assessment of response shift in QOL research. Soc Indicators Res 2000; 49(1): 37–49.

32. Bernhard J, Lowy A, Maibach R, Hurny C. Response shift in the perception of health for utility evaluation. An explorative investigation. Eur J Cancer 2001; 37(14): 1729–1735.
33. Bernhard J, Lowy A, Mathys N, Herrmann R, Hurny C. Health related quality of life: A changing construct? Qual Life Res 2004; 13(7): 1187–1197.
34. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: Differences between improvement and worsening. Qual Life Res 2002; 11(3): 207–221.
35. Hagedoorn M, Sneeuw KC, Aaronson NK. Changes in physical functioning and quality of life in patients with cancer: Response shift and relative evaluation of one's condition. J Clin Epidemiol 2002; 55(2): 176–183.
36. Jansen SJ, Stiggelbout AM, Wakker PP, Nooij MA, Noordijk EM, Kievit J. Unstable preferences: A shift in valuation or an effect of the elicitation procedure?. Med Decis Making 2000; 20(1): 62–71.
37. Postulart D, Adang EM. Response shift and adaptation in chronically ill patients. Med Decis Making 2000; 20(2): 186–193.
38. Rees J, Waldron D, O'Boyle C, Ewings P, MacDonagh R. Prospective vs. retrospective assessment of lower urinary tract symptoms in patients with advanced prostate cancer: The effect of 'response shift'. BJU Int 2003; 92(7): 703–706.
39. Wyrwich KW, Tardino VMS. Understanding global transition assessments. Qual Life Res 2006; (in press).

Address for correspondence: Carolyn Schwartz, Sc.D., DeltaQuest Foundation, Inc., 31 Mitchell Road, Concord, MA, 01742, USA
Phone: 978-318-7914;
E-mail: carolyn.schwartz@deltaquest.org