Review Article | Palliative Medicine and Palliative Care for Incurable Cancer

A systematic review of the use of the then-test for evaluating health-related quality of life in cancer patients

Emma McCurdy-Franks[1], Victoria McCarvell[1], Samuel Finkelstein[1], Lauren Milton[1], Tara Behroozian[1], Charles B. Simone II[2], Edward Chow[1], Joel Finkelstein[1,3]

[1]Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Canada; [2]Department of Radiation Oncology, New York Proton Center, New York, NY, USA; [3]Division of Orthopaedic Surgery, Sunnybrook Health Sciences Centre, Toronto, Canada

Contributions: (I) Conception and design: E Chow, J Finkelstein; (II) Administrative support: J Finkelstein, S Finkelstein, CB Simone 2nd; (III) Provision of study materials or patients: E McCurdy-Franks, L Milton, T Behroozian, E Chow; (IV) Collection and assembly of data: E McCurdy-Franks, V McCarvell, S Finkelstein; (V) Data analysis and interpretation: E McCurdy-Franks, V McCarvell, S Finkelstein; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Joel Finkelstein, MD, MSc, FRCS(C). Division of Orthopaedic Surgery, University of Toronto, Toronto, Canada; Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, Toronto, Ontario, M4N 3M5, Canada. Email: Joel.Finkelstein@sunnybrook.ca.

Background: The then-test, also known as the retrospective pre- and post-test design method, is a measurement used to evaluate response shift. The method requires patients to assess their previous health-related quality of life (HRQoL) and provide a retrospective judgement based on their current perspectives. The then-test, however, has been criticized for its reliability and validity. The objective of this systematic review is to summarize the current literature for the use of the then-test for evaluating HRQoL in cancer patients and account for potential response shift effects.

Methods: A literature search was conducted in May 2022 using MEDLINE, PubMed, and PsychINFO dated from 2005 to the time of the search. Studies were included if they (I) used the then-test and (II) involved a population of cancer patients (any cancer site).

Results: The literature search resulted in 16 studies, published between 2005 and 2020. All studies used the then-test to detect response shifts. The EORTC QLQ-C30 and EORTC QLQ-BR23 questionnaires were the most common assessments used to evaluate HRQoL. Of the 16 articles, 5 exclusively reported on breast cancer, 5 reported on prostate cancer, and the remaining included all cancer sites. Most studies looked at patients undergoing a combination of chemotherapy, radiotherapy, and hormonal therapy. The mean differences between the retrospective assessment at both 3 and 6 months were significant for various quality of life (QoL) dimensions. Patients in some studies recalled their pretreatment HRQoL (then-score) as better than the pretreatment baseline scores and others reported them as worse, both confirming the existence of a response shift.

Conclusions: Our review demonstrates that the then-test measurement tool has been commonly used for the detection of response shift. Newer measures for response shift have become more accepted; the then-test, if used, should be used with caution considering its limitations and the emergence of more advanced methods.

Keywords: Cancer; then-test; response shift; outcome assessment; quality of life (QoL)

Submitted Jun 24, 2023. Accepted for publication Sep 11, 2023. Published online Oct 19, 2023.
doi: 10.21037/apm-23-462
View this article at: https://dx.doi.org/10.21037/apm-23-462

Introduction

The health-related quality of life (HRQoL) measurement is important for evaluating outcomes of treatment from a patient's perspective. This outcome takes into account both potential toxicities from the treatment regimens and disease progression in cancer studies (1,2). The HRQoL self-assessments are often subjective and evluation-based, hence dependent on patients' individual standards and experiences.

In longitudinal studies, a patient's status is measured over the course of the disease or treatment. Over this time, one's expectations or how one assesses their HRQoL may change (1). The change in meaning of one's self-evaluation of quality of life (QoL) is known as a response shift. This can occur based on changes in "internal standards (scale recalibration), values (reprioritization), or a redefinition of the target construct (reconceptualization)" (3). Response shift is a natural process but can alter the interpretation of true change in HRQoL (4,5). Previously response shift effects were interpreted as "bias", "noise", "nuisance", or otherwise warranting removal from the results rather than as information that matters (6). Currently many investigators consider response shift as information, not bias in the interpretation of outcome and may, therefore, allow us to improve understanding of outcomes with and without response shift effects (6). If bias is accepted as causing measurement error, the effects of HRQoL can differ whether an assessment is measured prospectively (before and after treatment) or just retrospectively (only after treatment). Prospective evaluations may be subject to scale recalibration, whereas retrospective evaluations can be influenced by recall bias (incorrect assessment of former QoL).

One measure of response shift that has been the most common and least labor-intensive approach is the then-test, also known as the retrospective pre- and post-test design method (4). This method requires patients to retrospectively assess their previous health state based on their current perspective (7). The then-test specifically quantifies recalibration. Individuals evaluate their HRQoL before intervention (pre-test) and after intervention (post-test). Immediately following the post-test, the then-test is administered whereby patients are asked to reassess their pre-test HRQoL. They are not instructed to remember their pre-test rating but, instead, to retrospectively give a judgement of their previous HRQoL based on their current perceptions. Since the then-test is completed at the same time as the post-test, participants likely use the same internal standards (4). Therefore, the difference between the then-test and pre-test indicates the degree of response shift and the difference between the then-test and post-test provides evidence of true change in QoL (2). The then-test is used as it is easy to analyze and interpret the results; however, it is subject to recall bias as the test assumes the patient has an accurate recall of his/her health and performance status from prior to treatment. It has been vastly documented that the validity of the then-test is impaired by recall bias (8-11). When there lies a difference between the then-test and pre-test score, it is difficult to interpret how much of that difference is due to recall bias, recalibration, or something else. Schwartz & Sprangers laid out steps authors should take when employing the then-test acknowledging the issues in interpreting the data (12). Schwartz & Rapkin reported that the then-test represents many health and self-care concerns in addition to recalibration (13). It is also theorized that implicit theory of change and social desirability can impact the results of then-test scores (14-16). When assessing the HRQoL of cancer patients, it is important to evaluate the presence and magnitude of response shift, as this may be more representative of "true change". Chemotherapy toxicities may improve health status in physical ways, but the side effects may have undesirable effects on one's evaluative HRQoL. The objective of this review is to summarize the current literature for the use of the then-test for evaluating HRQoL in cancer patients and account for potential

Highlight box

Key findings

- This study found that results of the then-test in the assessment of response shift for quality of life (QoL) in cancer patients show mixed results in several domains and many results may be explained by factors beyond response shift.

What is known and what is new?

- The then-test has been commonly used to assess response shift for QoL in cancer patients, but results of the then-test may be explained by other phenomena such as recall bias, social desirability, and implicit theory of change.
- Most studies in this systematic review reported that recall bias is a potential limitation to their results.

What is the implication, and what should change now?

- The then-test may not be a reliable tool to assess response shift in this population and other tools to assess response shift are recommended.

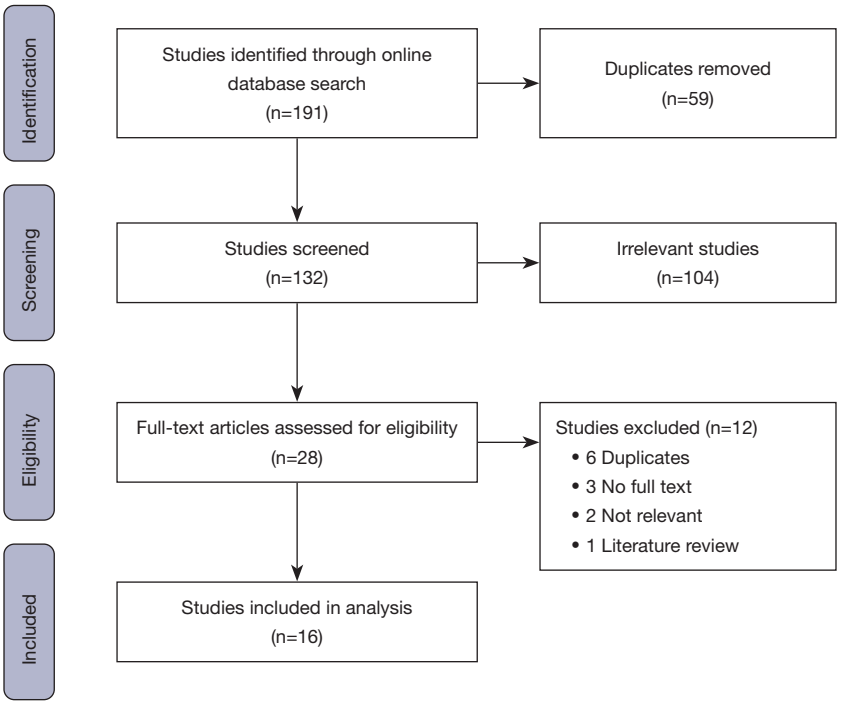

Figure 1 Article screening according to PRISMA guidelines. PRISMA, Preferred Reporting Items for Systematic Review and Meta-Analysis.

response shift effects. We present this article in accordance with the PRISMA reporting checklist (available at https://apm.amegroups.com/article/view/10.21037/apm-23-462/rc) (17).

Methods

Search strategy

In May 2022, a literature search for articles in English was conducted using MEDLINE, PubMed, and PsychINFO. The following keywords were used: "cancer", "then-test", "response shift", "outcome assessment", and "quality of life".

Article selection

Articles were included if (I) the study used the then-test and (II) the study was with a population of cancer patients (any cancer site). Articles were excluded if (I) the then-test was not used; (II) there were no cancer patients in the study population; and (III) the study was a literature or systematic review. Abstracts and posters were also excluded. Titles and abstracts were independently screened by two authors (McCurdy-Franks E, McCarvell V) according to the eligibility criteria. Articles that fit the criteria were then assessed through a full text screening (McCurdy-Franks E, McCarvell V). Differences in results were discussed, and a consensus was reached between the two authors. The screening process is depicted in *Figure 1*.

Data collection

The following information was collected for each study: lead author, country, study type, aim of study, sample size, cancer type, mean age, sex, race, marital status, education level, household income, type of therapy, survey type, cancer stage, and mean days between pre-treatment and post-treatment. The inclusion/exclusion criteria, type and timing of assessments, QoL results, response shift effect sizes with standard deviation, and conclusions were additionally recorded for each study. Response shift effect size using the then-test is defined as the difference between the then-test and the pre-test scores.

Results

Search results

The initial literature search resulted in 191 studies. After removing 59 duplicates, 132 results remained and were screened based on the eligibility criteria. From the title and abstract screening, 104 articles were deemed irrelevant and excluded. Twenty-eight full-text studies were then screened and an additional 12 were excluded, leaving 16 articles for the final analysis and extraction (*Figure 1*). Authors (McCurdy-Franks E and McCarvell V) used the data extraction tool, Covidence, to individually collect information from the 16 articles. Once the extraction was complete, both authors reviewed for discrepancies and came to a consensus.

Patient demographics

A summary of the patient demographics and included studies can be found in *Table 1* (1,2,4,5,7,18-28). The studies in this review were published since 2005 in a variety of countries, including the United States, France, Netherlands, Canada, Iran, Japan, Norway, UK, and Spain. Of the 16 articles, 5 exclusively reported on breast cancer, 5 exclusively on prostate cancer, and the remaining included all types of cancer. The sample size of patients ranged from 13 to 3,161. The mean age of patients was 61.9 years and ranged between 49 and 79.9 years old. Race was mentioned in two of the studies, both of which had a majority of white patients. Six studies looked at patients undergoing chemotherapy or adjuvant chemotherapy, 10 at radiotherapy, and 7 at hormonal therapy. Most of the studies also looked at a combination of these therapies. Three studies did not mention the type of therapy used. Four articles included patients that had undergone a mastectomy, one a lumpectomy, and another a combination of the two. Additional surgery types were also investigated, including prostatectomy, axillary lymph node dissection, APR, Hartmann's operation, U-LAR resulting in permanent diverting colostomy, and TPE. Eight of the 16 articles did not mention the cancer stage; however, the other eight included patients with cancer stages ranging from 0–4.

Measurement tools

The then-test was used to detect response shifts in all the studies. The EORTC QLQ-C30 and EORTC QLQ-BR23 questionnaires were the most common assessments used to evaluate QoL. The following assessments were additionally used: the 36-Item Short Form Health Survey (SF-36), EQ-5D, EuroQol self-rating of health, Functional Assessment of Chronic Illness Therapy (FACIT)-Fatigue scale, Fatigue Symptom Inventory (FSI), International Prostate Symptom Score (IPSS), Symptom Problem Index (SPI), Prostate Cancer Patient & Partner questionnaire (PPP), and the University of California Los Angeles (UCLA) Prostate Cancer Index (PCI) HRQoL measures. The assessments were usually conducted at baseline, post-treatment, and at 3- and 6-month follow-ups.

Evidence of response shift

Tables 2,3 show the response shift values for studies that used the QLQ-C30 and SF-36 questionnaires, respectively.

Response shift in prostate cancer patients was measured in a study done by Ten Ham *et al.* (n=3,161) using the then-test (25). Participants were asked to retrospectively evaluate their HRQoL using the SF-36 and UCLA PCI (25). Patients consistently recalled their pretreatment HRQoL (then-score) as better than the actual pretreatment baseline scores (pre-score), which confirms the existence of response shift ($P<0.05$). The PCI response shift score was 7.4 (range, 4.3 to 14.7) and the mean SF-36 response shift score was 15.1 (range, 4.0 to 30.2) (25). An additional study by Korfage *et al.* looked at response shift in men with prostate cancer (n=52) using the EuroQol self-rating of health, SF-36 mental health and vitality assessments to evaluate HRQoL (22). The results from the study indicated a positive judgement in retrospect, as the pre-diagnosis scores were significantly higher than the original scores. For example, in pre-diagnosis, the mental health score was 83.2, but the then-test score 1-month post-diagnosis was 84.5 (22). Furthermore, Rees *et al.* used the IPSS and SPI to look at response shifts in prostate cancer patients. Results showed that IPSS and SPI significantly decreased over 6 months from pre-test to post-test. The then-test scores were consistently higher than the pre-test scores with statistical significance for both these scales, indicating response shift possibly occurred (23).

Two studies investigated response shift effects in breast cancer patients. Hamidou *et al.* assessed QoL using the EORTC-QLQ-C30 and BR-23 questionnaires at inclusion (T0), at the end of the first hospitalization (T1), 3 months (T2), and 6 months after the first hospitalization (T3) (7). Interestingly, the mean differences (MDs) between the retrospective assessment at T2 (3 months) and T3

Table 1 Patient characteristics

Study	Sample size	Cancer type	Therapy type	QoL measure	Mean age in years ± SD	Test administration
Andrykowski *et al.*, 2009	102	Breast	CT + RT (41%), CT (6%), RT (53%)	FSI	54.7±10.6	P1, P2 (70.2± 40.9 days), post-treatment two (193.8± 57.3 days)
Anota *et al.*, 2014	381	Breast	CT (n=155), RT (n=254), HT (n=170)	EORTC QLQ-C30, EORTC QLQ-BR23	58.4±11	P1, P2, P3 (3 and 6 months after)
Serdà I Ferrer *et al.*, 2014	66	Prostate	HT (n=30), RT + HT (n=2), prostatectomy + HT (n=4), surgery (n=30)	FACIT-Fatigue scale	71.78±7.22	P1, P2, P3
Brinksma *et al.*, 2014	NS	NS	NS	Health status assessed by PPS, MSAS	Child report, 14 [8–17]*; parent report, 9 [2–17]*	P1, P2 (3 months after)
Chow *et al.*, 2007	217	Lung (n=59), prostate (n=50), breast (n=49), colorectal (n=13)	RT (n=217)	NS	66 [28–88]*	P1, P2, P3 (2 months after)
Dabakuyo *et al.*, 2013	381	Breast	Adjuvant CT (n=133), neoadjuvant CT (n=30), RT (n=254), HT (n=170)	EORTC-QLQ-C30, EORTC-QLQ-BR23, EuroQoL-EQ-5D	58.4±11	P1, P2
Hamidou *et al.*, 2011	381	Breast	Adjuvant CT (n=155), mastectomy (n=124), sentinel lymph node biopsy (n=131)	EORTC-QLQ-C30, EORTC QLQ-BR23	58±11.1	P1, P2, P3 (3 and 6 months after)
Hinz *et al.*, 2011	275	Urogenital	CT (n=12), RT (n=17), HT (n=33)	Generalized Anxiety Disorder Questionnaire, Patient Health and Life Satisfaction Questionnaire	63.7±8.3	P1, P2 (2 weeks later), P3 (3 months after)
Hosseini *et al.*, 2017	211	Breast (n=85), GI (n=63), reproductive organs (n=22), lung (n=7), bladder (n=8), prostate (n=6), head and neck (n=16), sarcoma (n=2)	NS	EORTC QLQ-C30	51.3±13.9	P1, P2 (3 months after)
Ito *et al.*, 2010	13	Rectum (n=10), hemorrhoid (n=1), rectum and anus (n=1), anus and hemorrhoid (n=1)	NS	SF-36 version 2 (Japanese version)	66.9±11.3	P1, P2 (2 months after)
Jakola *et al.*, 2017	73	Gliomas	NS	EQ-5D 3L questionnaire	49±15	P1, P2 (6 months after)
Korfage *et al.*, 2007	52	Prostate	Radical prostatectomy (n=25), brachytherapy (n=12), active surveillance (n=10), external RT (n=3), HT (n=1)	EQ-5D VAS, EuroQol self-rating of health, Short-Form 36 mental health and vitality	67.3±4.4	P1, P2 (1 month post diagnosis)

Table 1 (*continued*)

Table 1 (*continued*)

Study	Sample size	Cancer type	Therapy type	QoL measure	Mean age in years ± SD	Test administration
Rees *et al.*, 2005	55	Prostate	HT (n=44), radical RT + neoadjuvant HT (n=8)	PPP	72.9±8.5	P1, P2 (3 months and 6 months after)
Rees *et al.*, 2003	76	Prostate	HT (n=61), RT (n=11)	IPSS, SPI	Patients, 72.8±8.5; controls, 71.8±5.0	P1, P2 (3 and 6 months after)
Ten Ham *et al.*, 2020	3,161	Prostate	Radical prostatectomy (n=1,131), electron beam RT (n=243), brachytherapy (n=227), watchful waiting (n=76)	Medical Outcomes Study Questionnaire, SF-36 and the UCLA PCI HRQOL measures	SF-36, 63.91±7.78; PCI, 63.9±7.71	1–7 months after
Ousmen *et al.*, 2016	381	Breast	CT (n=155), RT (n=254), HT (n=170)	EORTC QLQ-C30 and QLQ-BR23	58.4±11	P1, P2 (3 and 6 months after)

*, median age [range]. NS, not specified; GI, gastrointestinal; CT, chemotherapy; RT, radiotherapy; HT, hormone therapy; FSI, Fatigue Symptom Inventory; FACIT, Functional Assessment of Chronic Illness Therapy; PPS, Play Performance Scale; MSAS, Memorial Symptom Assessment Scale; SF-36, 36-Item Short Form Health Survey; VAS, Visual Analog Scale; PPP, Prostate Cancer Patient & Partner questionnaire; IPSS, International Prostate Symptom Score; SPI, Symptom Problem Index; UCLA, University of California Los Angeles; PCI, Prostate Cancer Index; HRQOL, health-related quality of life; SD, standard deviation; P1, baseline; P2, post-treatment; P3, follow-up.

Table 2 Mean difference of response shift values for QLQ-C30

QoL measure	Anota *et al.*, 2014	Dabakuyo *et al.*, 2013	Hamidou *et al.*, 2011	Ousmen *et al.*, 2016	Hosseini *et al.*, 2017
Time of measurement	3 months	6 days (median)	3 months	3 months	3 months
Global QoL	–4.21 (18.45)	–	–3.7 (18.2)	3.82 (17.85)	–15.64 (42.10)
Physical functioning	–1.59 (13.26)	–	–1.6 (12.6)	1.54 (12.95)	–
Role functioning	–6.50 (23.72)	–	–6.3 (22.2)	6.06 (21.71)	–
Emotional functioning	6.97 (21.48)	5.3 (18.9)	6.8 (21.2)	–7.56 (21.03)	14.45 (21.74)
Cognitive functioning	2.37 (18.27)	2.9 (15.2)	3.0 (18.2)	–3.94 (16.85)	–
Social functioning	–5.01 (20.70)	–	–3.8 (19.3)	5.09 (19.43)	–
Fatigue	1.75 (20.92)	–1.3 (18.6)	1.37 (20.8)	–1.20 (19.42)	–11.73 (31.48)
Nausea and vomiting	1.77 (15.11)	–	1.3 (15.7)*	–1.37 (13.59)	–
Pain	3.24 (23.03)	–	2.5 (22.1)	–3.33 (22.51)	–10.80 (30.67)
Dyspnea	–1.08 (15.58)	–	–0.4 (16.2)	1.20 (15.16)	–
Insomnia	–6.93 (30.94)	–5.1 (26.5)	–7.2 (30.8)	6.83 (30.64)	–
Appetite loss	–1.19 (23.75)	–3.4 (19.5)	–2.7 (23.5)	2.28 (21.05)	–
Constipation	1.56 (24.93)	–	–	0 (22.61)	–
Diarrhea	–2.89 (17.25)	–3.1 (12.7)	–2.9 (16.8)	2.42 (14.99)	–
Financial difficulties	0.99 (16.51)	–	–	–1.11 (15.32)	–

Data are shown as mean response shift (SD). *, nausea only. QoL, quality of life; SD, standard deviation.

Table 3 Mean difference of response shift values for SF-36

QoL measure	Ito *et al.*, 2010	Korfage *et al.*, 2007	Ten Ham *et al.*, 2020
Vitality	–	−0.26	–
General state of physical health	0.1	–	–
Physical functioning	0	–	−4.1 (18.7)
Physical role	−0.37	–	−14.5 (40.9)
Emotional role	−0.09	–	−9.0 (32.8)
Energy	–	–	−12.5 (18.3)
Emotional well-being	–	–	−5.5 (13.9)
Social functioning	−0.04	–	−8.2 (21.5)
Bodily pain	−0.59	–	−5.3 (15.6)
General health	–	–	−3.6 (8.6)
Physical summary score	–	–	−4.6 (4.1)
Mental summary score	–	0.01	−1.9 (8.1)

Data are shown as mean response shift (SD). SF-36, 36-Item Short Form Health Survey; QoL, quality of life; SD, standard deviation.

(6 months) were statistically different for the majority of QoL dimensions. QoL scores were higher at the retrospective assessment T3 (6 months) compared to T2 (3 months) for physical-functioning (MD =4.3) and role-functioning (MD =6.95). However, QoL scores were higher at T2 (3 months) compared to retrospectively at T3 (6 months) for fatigue (MD =29.48), pain (MD =25.07), dyspnea (MD =23.25) and insomnia (MD =23.96) (7). The internal standards of QoL for breast cancer patients had an impact on time to deterioration (TTD), as TTD was shorter when not taken into account for global health, role-functioning, social-functioning, body-image and side effects of systemic therapy (7). The QoL of breast cancer patients in the study by Dabakuyo *et al.* was evaluated using the QLQ-C30, BR23, and EuroQOL-EQ-5D. The recalibration effect was statistically significant after first hospitalization for 6/15 dimensions of QLQ-C30 (emotional, cognitive, fatigue, insomnia, appetite loss, diarrhea) and 2/8 of BR-23 (future perspective, systemic therapy side effects) (19). It was also mentioned that it was clinically significant for the emotional dimension (MD =5.3), insomnia (MD =−5.1), and future perspectives (MD =7.9) (19). Furthermore, health and QoL expectancies changed between the baseline and the end of the first hospitalization, a larger number of people reported self-care (P=0.029), and usual activities (P=0.049) as "not important" (19).

The following studies evaluated various types of cancer patients. For example, using the EORTC QLQ-C30, Hosseini *et al.* found that response shift is significant in 4 scales: fatigue, pain, emotional functioning, and general QoL (P<0.001) (27). Emotional functioning significantly rated better in retrospect using the then-test approach with a response shift value of 14.45 (95% CI: 11.22 to 17.67), while fatigue, pain, and global QoL significantly deteriorated with response shift values of −11.73 (95% CI: −16.37 to −7.08) for fatigue, −10.80 (95% CI: −15.32 to −6.27) for pain and −15.64 (95% CI: −21.84 to −9.45) for QoL (27). Ousmen *et al.* found that the magnitude of the response shift effect increased over time in patients whose QoL deteriorated and decreased in patients who reported improvement of QoL (26). These changes were most notable in patients whose QoL deteriorated showing that the magnitude of response shift increased between 3 and 6 months in 13/15 dimensions of QLQ-C30 questionnaire and 4/7 dimensions of QLQ-BR23 questionnaires thereby providing evidence that response shift may have a greater impact if patients report declining QoL (26).

Two articles included additional groups in their study. Brinksma *et al.* used the then-test to evaluate response shift for HRQoL in pediatric cancer patients (n=37) and their parents, using the PedsQL and Cantril's ladder assessments (4). Cantril's ladder measures QoL on a visual analog scale from 0–10, with 10 being the best possible QoL, and 0 being the worst possible QoL (4). The results

found that the Cantril then-test ratings were lower than the pre-test ratings for both the child report (P<0.001) and parent report (P<0.05), indicating a response shift in HRQoL. There was no difference in the then- and pre-tests for the PedsQL assessment (Z=–0.57, P=0.572 for child report, Z=–0.08, P=0.935 for parent report) (4).

Rees *et al.* evaluated advanced prostate cancer patients and their partners (n=55) using the PPP. They were assessed at diagnosis and again at 3 and 6 months. The results show that partners have greater psychological morbidity on General Cancer Distress subscale (P<0.001, paired *t*-test) and worries about treatment (P=0.01) (24). A paired *t*-test suggested that the magnitude of response shift between patients and partners was not statistically significant during the first 3 months (P=0.3), but became significantly different between the second and third assessments (P=0.02). The use of the then-test resulted in a significant 'actual change' for both patients and partners between visits 1 and 2 (P=0.03 for patients, P=0.02 for partners, paired *t*-test). The only significant change using the then-test occurs in the patient group, between visits 2 and 3 (P=0.01) (24).

Discussion

Our study provides an overview of existing literature using the then-test to investigate response shift effects. Patients in some studies recalled their pretreatment QoL as better than the pretreatment baseline scores, and others reported them as worse, both confirming the existence of a response shift. Overall, 9/16 studies showed response shift impacting QoL in both directions. For studies that measured overall health using a specific scale such as "global QoL", "general cancer distress", or "general state of physical health", the results were varied. Three studies scored overall health as higher retrospectively compared to the baseline score, one study showed no change, and six studies scored overall health as lower retrospectively.

There has been controversy among leaders in this field about whether response shift effects represent a bias or whether it provides further valuable information (6). It is important to consider response shift in a patient's disease course because it may have implications for patient reported outcomes (29). This concept can be applied when considering how a treatment may alter aspects of a patient's life and thus allow more personalized treatment. The use of the then-test has been criticized by some as not being robust enough for detecting response shift due to recall bias and because it is difficult to use in longitudinal secondary data analysis (3,29).

The most common methods used to measure response shift are the then-test and Oort's structural equation models (SEMs) (3). These both detect recalibration. The then-test is used for studies with primary data, and SEM is used to assess both primary and secondary data. The then-test is easy to use, yet it is prone to random errors. SEM is more versatile and can detect a larger combination of response shift types: recalibration, reprioritization, and reconceptualization (3).

Despite its limitations, the then-test has been commonly used to assess response shift in cancer patients. A systematic review in 2019 of response shift in studies assessing QoL in cancer patients showed 21/35 studies used the then-test (30). Other methods used to detect response shift include, but are not limited to SEM, vignettes, appraisal, semi-structured interview, schedule for the evaluation of individual QoL, random forest regression, mixed models and growth mixture models, item response theory, classification and regression tree, and relative importance analysis (31). Since the then-test may be prone to bias, it may be more suited to studies with smaller sample sizes as other response shift methods based on statistical modelling require larger sample sizes (32). However, newer methods of measuring response shift can be effectively applied toward smaller samples. Schwartz *et al.* recently devised a method to detect response shifts that included recalibration, reprioritization, and reconceptualization in small samples of clinical trials using random-effects modeling, which is derived from Oort's structural-equation modeling (33).

Of the articles included in this review, 10 used the then-test only to detect response shift. Using only this test allows the researcher to explore the "recalibration" aspect of response shift solely, leaving out "reprioritization" and "reconceptualization". Twelve articles from this review mentioned recall bias as a limitation to their studies. These studies lacked methodology to assess recall bias. If a second method is used to detect response shift without the possibility of recall bias, the results can be compared and help control for recall bias and make the results more reliable (12,34). Timing of follow-up assessments may influence results. Assessments conducted too soon to baseline, patients are still adjusting to treatment and its effects, whereas too long in the future and issues with memory may arise. A study included in our review by Hamidou *et al.* assessed how TTD of QoL changed with or without accounting for then-test results (7). Several other studies in our review used multiple timepoints to

assess response shift. Sébille *et al.* [2021] have suggested that response shift detection methods such as SEM, mixed models, Rasch Measurement Theory, and Item Response Theory are accommodating to multiple timepoints (31).

It may be relevant to look at the findings of response shifts in specific cancer patient populations. For example, Friedrich *et al.* found that applying SEM to evaluate response shift in breast cancer patients detected a recalibration effect for social functioning (35). When comparing the two tests, it is useful to use the SEM approach to retrospective judgements, as it might show effects that the then-test will overlook (35). A study by Preiß *et al.* used both the vignettes and then-test methods to examine response shift effects (3). It was found that patients with physical problems were assessed as healthier using the vignette, and patients reported their QoL and mental burden as worse than pre-test using the then-test (3). This shows that using different methods can uncover more response shift effects. In addition, only one of all included studies was conducted in pediatric oncology. Thus, determining how different patient populations may experience response shift is another aspect that should be investigated. Future studies should consider either employing other methods than the then-test to uncover response shift effects or alternatively use an additional response shift method to the then-test in their research so the effect sizes of response shift can be compared, since the then-test is prone to recall bias and other shortcomings (29,36).

Conclusions

It is probable that the then-test will be used to a lesser degree in the future owing to the numerous other methods that have emerged to assess response shift. Other methods have become more popular in part due to concerns of recall bias, among other difficulties with the then-test, which is a limitation to this study. Another significant limitation is that the then-test only putatively measures recalibration. Ten articles used the then-test only therefore suggesting these studies missed response shift effects for reprioritization and reconceptualization. While the then-test has contributed to some patient perspective of cancer patient outcomes, this study has shown that the then-test, if used, should be used with caution considering its limitations and the emergence of more advanced methods.

Acknowledgments

This abstract was presented at the MASCC/JASCC/ISOO 2023 Annual Meeting in Nara, Japan.

Funding: This work was supported by the Feldberg Chair in Spinal Research, Sunnybrook Health Sciences Centre.

Footnote

Reporting Checklist: The authors have completed the PRISMA reporting checklist. Available at https://apm.amegroups.com/article/view/10.21037/apm-23-462/rc

Peer Review File: Available at https://apm.amegroups.com/article/view/10.21037/apm-23-462/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://apm.amegroups.com/article/view/10.21037/apm-23-462/coif). C.B.S. is serving as Editor-in-Chief of *Annals of Palliative Medicine*. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Anota A, Bascoul-Mollevi C, Conroy T, et al. Item response theory and factor analysis as a mean to characterize occurrence of response shift in a longitudinal quality of life study in breast cancer patients. Health Qual Life Outcomes 2014;12:32.

2. Hinz A, Finck Barboza C, Zenger M, et al. Response shift in the assessment of anxiety, depression and perceived health in urologic cancer patients: an individual perspective. Eur J Cancer Care (Engl) 2011;20:601-9.
3. Preiß M, Friedrich M, Stolzenburg JU, et al. Response shift effects in the assessment of urologic cancer patients' quality of life. Eur J Cancer Care (Engl) 2019;28:e13027.
4. Brinksma A, Tissing WJ, Sulkers E, et al. Exploring the response shift phenomenon in childhood patients with cancer and its effect on health-related quality of life. Oncol Nurs Forum 2014;41:48-56.
5. Chow E, Chiu H, Doyle M, et al. Patient expectation of the partial response and response shift in pain score. Support Cancer Ther 2007;4:110-8.
6. Schwartz CE, Rohde G, Biletch E, et al. If it's information, it's not "bias": a scoping review and proposed nomenclature for future response-shift research. Qual Life Res 2022;31:2247-57.
7. Hamidou Z, Dabakuyo TS, Mercier M, et al. Time to deterioration in quality of life score as a modality of longitudinal analysis in patients with breast cancer. Oncologist 2011;16:1458-68.
8. 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.
9. Schwartz CE, Sprangers MAG, Carey A, et al. Exploring response shift in longitudinal data. Psychology & Health 2004;19:51-69.
10. Taminiau-Bloem EF, Schwartz CE, van Zuuren FJ, et al. Using a retrospective pretest instead of a conventional pretest is replacing biases: a qualitative study of cognitive processes underlying responses to thentest items. Qual Life Res 2016;25:1327-37.
11. McPhail S, Haines T. Response shift, recall bias and their effect on measuring change in health-related quality of life amongst older hospital patients. Health Qual Life Outcomes 2010;8:65.
12. Schwartz CE, Sprangers MA. Guidelines for improving the stringency of response shift research using the thentest. Qual Life Res 2010;19:455-64.
13. Schwartz CE, Rapkin BD. Understanding appraisal processes underlying the thentest: a mixed methods investigation. Qual Life Res 2012;21:381-8.
14. Finkelstein JA, Quaranto BR, Schwartz CE. Threats to the Internal Validity of Spinal Surgery Outcome Assessment: Recalibration Response Shift or Implicit Theories of Change? Applied Research Quality Life 2014;9:215-32.
15. Norman G. Hi! How are you? Response shift, implicit theories and differing epistemologies. Qual Life Res 2003;12:239-49.
16. Nolte S, Elsworth GR, Sinclair AJ, et al. Tests of measurement invariance failed to support the application of the "then-test". J Clin Epidemiol 2009;62:1173-80.
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
18. Andrykowski MA, Donovan KA, Jacobsen PB. Magnitude and correlates of response shift in fatigue ratings in women undergoing adjuvant therapy for breast cancer. J Pain Symptom Manage 2009;37:341-51.
19. Dabakuyo TS, Guillemin F, Conroy T, et al. Response shift effects on measuring post-operative quality of life among breast cancer patients: a multicenter cohort study. Qual Life Res 2013;22:1-11.
20. Ito N, Ishiguro M, Tanaka M, et al. Response shift in quality-of-life assessment in patients undergoing curative surgery with permanent colostomy: a preliminary study. Gastroenterol Nurs 2010;33:408-12.
21. Jakola AS, Solheim O, Gulati S, et al. Is there a response shift in generic health-related quality of life 6 months after glioma surgery? Acta Neurochir (Wien) 2017;159:377-84.
22. Korfage IJ, de Koning HJ, Essink-Bot ML. Response shift due to diagnosis and primary treatment of localized prostate cancer: a then-test and a vignette study. Qual Life Res 2007;16:1627-34.
23. Rees J, Waldron D, O'Boyle C, et al. Prospective vs retrospective assessment of lower urinary tract symptoms in patients with advanced prostate cancer: the effect of 'response shift'. BJU Int 2003;92:703-6.
24. Rees J, Clarke MG, Waldron D, et al. The measurement of response shift in patients with advanced prostate cancer and their partners. Health Qual Life Outcomes 2005;3:21.
25. Ten Ham RMT, Broering JM, Cooperberg MR, et al. Understanding the Major Factors Affecting Response Shift Effects on Health-Related Quality of Life: What the Then-Test Measures in a Longitudinal Prostate Cancer Registry. Clin Genitourin Cancer 2020;18:e21-7.
26. Ousmen A, Conroy T, Guillemin F, et al. Impact of the occurrence of a response shift on the determination of the minimal important difference in a health-related quality of life score over time. Health Qual Life Outcomes 2016;14:167.
27. Hosseini B, Nedjat S, Zendehdel K, et al. Response shift in quality of life assessment among cancer patients: A study from Iran. Med J Islam Repub Iran 2017;31:120.

28. Serdà I Ferrer BC, Valle AD, Marcos-Gragera R. Prostate cancer and quality of life: analysis of response shift using triangulation between methods. J Gerontol Nurs 2014;40:32-41.
29. Schwartz CE, Bode R, Repucci N, et al. The clinical significance of adaptation to changing health: a meta-analysis of response shift. Qual Life Res 2006;15:1533-50.
30. Ilie G, Bradfield J, Moodie L, et al. The Role of Response-Shift in Studies Assessing Quality of Life Outcomes Among Cancer Patients: A Systematic Review. Front Oncol 2019;9:783.
31. Sébille V, Lix LM, Ayilara OF, et al. Critical examination of current response shift methods and proposal for advancing new methods. Qual Life Res 2021;30:3325-42.
32. Ortega-Gómez E, Vicente-Galindo P, Martín-Rodero H, et al. Detection of response shift in health-related quality of life studies: a systematic review. Health Qual Life Outcomes 2022;20:20.
33. Schwartz CE, Stark RB, Stucky BD. Response-shift effects in neuromyelitis optica spectrum disorder: a secondary analysis of clinical trial data. Qual Life Res 2021;30:1267-82.
34. Visser MR, Oort FJ, Sprangers MA. Methods to detect response shift in quality of life data: a convergent validity study. Qual Life Res 2005;14:629-39.
35. Friedrich M, Zenger M, Hinz A. Response shift effects of quality of life assessments in breast cancer survivors. Eur J Cancer Care (Engl) 2019;28:e12979.
36. Schwartz CE, Zhang J, Rapkin BD, et al. Reconsidering the minimally important difference: evidence of instability over time and across groups. Spine J 2019;19:726-34.

Cite this article as: McCurdy-Franks E, McCarvell V, Finkelstein S, Milton L, Behroozian T, Simone CB 2nd, Chow E, Finkelstein J. A systematic review of the use of the then-test for evaluating health-related quality of life in cancer patients. Ann Palliat Med 2023;12(6):1187-1197. doi: 10.21037/apm-23-462

