International Journal of Technology Assessment in Health Care, **19:2** (2003), 347–361.

EFFECT OF INCLUDING (VERSUS EXCLUDING) FATES WORSE THAN DEATH ON UTILITY MEASUREMENT

Duska M. Franic
University of Georgia

Dev S. Pathak
Ohio State University

Abstract

Objectives: Most studies typically measure health preferences excluding health states perceived as worse than death. The objective of this study is to test the impact of including (versus excluding) health states perceived to be worse than death on utility measurement using standard gamble (SG) and visual analogue scale (VAS) methods.
Methods: By means of a cross-sectional descriptive study design, women were asked to rate the utility of three hypothetical breast cancer health states: cure, treatment, and recurrence ($n = 119$). Preference weights were estimated, allowing for negative utilities with death (perfect health) scaled at zero (1.0).
Results: Unpaired t-test analysis showed significantly greater change in SG and VAS weights for individuals perceiving cancer recurrence as worse than death than those perceiving death as least desirable state. Excluding negative utilities from the study resulted in significantly smaller changes in utility. Study results show that preference elicitation methods can be successfully adapted to acquire negative utilities.
Conclusions: Changes in utility were greater when negative preferences were permitted. Addressing negative preference scores could significantly affect quality adjusted life year estimates in economic analyses.

Keywords: Breast cancer, Visual analogue scale, Patient preference, Death, QALY

How best to measure outcomes in cost utility analysis (CUA) has been an active area of debate and research (18). Two approaches have been used in anchoring holistic utility values. First, the most widely used method reported in the literature requires respondents to assume the best health state is perfect health—typically anchored at 1.0—and the worst is immediate death anchored at zero (16;40). This method defines alternative health states as intermediate to the anchors and, therefore, excludes health states worse than death, despite evidence to the contrary (2;21;22;27;28;31;32;34–36;38;41).

The second approach to utility assessment allows respondents to label anchors with health states they regard as the most and least desirable (as opposed to dictating anchors to

Study was supported by a grant awarded by The Columbus Medical Research Foundation, Columbus, Ohio.
This paper was presented at International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 6th Annual International Meeting, Hyatt Regency Crystal City, Arlington, Virginia, May 22, 2001.

respondents). Therefore, patients are permitted to consider health states worse than death if this is consistent with their preferences (34). Scales can still be standardized such that perfect health is anchored at 1.0, immediate death is zero, and health states worse than death are assigned negative weights.

This standardization results in all individuals being equally presented in the health domain. Utility for full health is set equal for all individuals at 1.0 (33). There is nothing in the expected utility theory (EUT) that states utility values cannot be negative, nor is there a theoretical lower boundary for preference weights (35). As a result, death is not necessarily the bottom of the scale but a convenient reference point, that is, ideal for this purpose due to its absolute nature—"the big zero whence we do not return" (31). This concept requires clarification on two issues: first, it must also be noted that this quote was made in reference to health states only considered preferable to death; however, in the case of health states considered equivalent to death, improvement or deterioration is possible. Second, if the intervention in question is a health "good" to be maximized, then perfect health is defined as 1.0 and death as zero, for example, the estimation of quality adjusted life years (QALYs) for use in CUA. However, if the research question is to estimate burden of disease, that is, a "bad" to be minimized, such as with the use of disability adjusted life years (DALYs) then, perfect health is assigned a value of zero and death is assigned as 1.0 (1;20). Reference to the use of anchor definitions in this study specifically refer to QALY estimations.

The significance of assigning anchors versus allowing patient selection of anchors is twofold. If classic welfare economics is adopted as the theoretical framework by CUA advocates (17;33), it would follow that the method used to acquire weights should be consistent with the requirements of the welfarist approach. One of the central tenets used to describe welfare economics—consumer sovereignty—is that consumers are the best judges of their own welfare (4;17). Theoretically, asking patients to select scale anchors concordant with their preferences is more consistent with meeting requirements of consumer sovereignty than anchor assignment by a researcher or clinician.

Practically, choice of how anchors are selected can have a significant impact on the change in preference score (ΔQ) estimate, and, therefore, the change in quality adjusted life years (ΔQALYs) attributed to a given health care intervention. Figure 1 provides examples

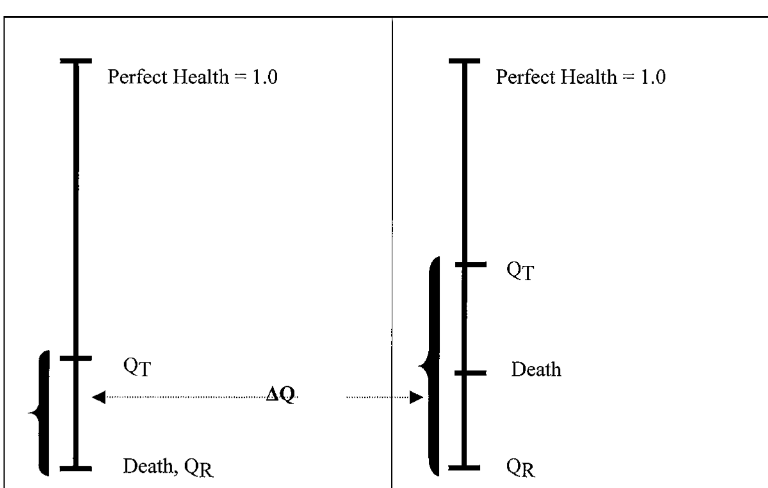

Figure 1. Left: Implications of anchor assignment on change in weights (ΔQ) by researcher. Right: Implications of patient-selected anchors on change in weights (ΔQ), where Q_T and Q_R are the preference weights for breast cancer treatment and recurrence, respectively. Perfect health and death are scaled at 1.0 and zero.

of patient-assigned and patient-selected approaches for utility estimation, respectively, for a hypothetical breast cancer therapy. Figures are scaled with perfect health at 1.0 and death at zero. Two intermediate conditions are breast cancer treatment (Q_T) and recurrence (Q_R). Example provided for patient-assigned anchors (Figure 1, left) shows a floor effect for breast cancer therapy; therefore, Q weight for Q_R is equal to that of death (zero), and perfect health is scaled at 1.0 in both (Figure 1). The patient-selected anchor approach (Figure 1, right) shows a larger ΔQ than the patient-assigned anchor approach (Figure 1, left) due to allowance of negative preference weights and, therefore, removal of the floor effect.

Although literature on health measurement preference is extensive, addressing health states worse than death in utility assessment is limited (6;12–15;21;25–27;32;34;42). Because the objective of CUA is to estimate incremental cost per health care benefit, this study will evaluate the impact of fates worse than death on ΔQ as opposed to Q values. The main objective of this study, thus, was to test the impact of including versus excluding health states perceived as worse than death by respondents on utility measurement using standard gamble (SG) and visual analogue scale (VAS) methods in breast cancer. Breast cancer was the chronic condition selected because it was the most common form of cancer in women and the second leading cause of death in women after lung cancer in the United States. It is expected that one in nine American women will develop breast cancer and every woman is at risk. In 2001, 192,000 new cases are expected with 40,000 women anticipated to die from the disease (29).

In this article, three analyses were conducted to compare ΔQ for three hypothetical treatments, that is, breast cancer cure versus do nothing (situation C versus R), treatment versus do nothing (situation T versus R), and cure versus treatment (situation C versus T). First, we compared mean ΔQ estimates for respondents perceiving fates worse than death with those that do not. We anticipated that there would be substantive differences in mean ΔQ between the two groups of respondents. Second, mean ΔQ was compared for all respondents versus those not perceiving health states worse than death. The rationale for the second test is twofold. First, excluding fates worse than death is typical of studies included in the literature. Thus, comparing their exclusion versus inclusion provides an estimate of ΔQ affect in a setting more consistent with that used in the literature. Second, because the second analysis is anticipated to show a significant ΔQ resulting from excluding respondents perceiving fates worse than death, this analysis provides an estimate of the proportion of study respondents perceiving fates worse than death resulting in significant ΔQ. The third test, threshold analysis, is an extension of the previous analysis and estimates the smallest proportion of respondents required to perceive fates worse than death to result in significantly different mean ΔQ compared with the group of respondents not perceiving fates worse than death.

METHODS

A cross-sectional descriptive study design was used (39). The study presented is part of a larger study described in detail elsewhere (10;11). The study sample consisted of women 22 to 50 years of age with affiliations to a midwestern university. This frame is consistent with a highly educated population, regarded as advantageous due to the cognitively demanding nature of study interviews. Because the health states discussed during the interview process were breast cancer, respondents with a history of breast cancer or cancer requiring chemotherapy were excluded from the study to avoid possible emotional discomfort or distress.

Respondents were recruited using e-mails, word-of-mouth, and fliers. Respondents completing interviews were offered $30 as an incentive. Potential study respondents were told that all information provided would remain strictly confidential. Before study participation, respondents completed a consent form stating they agreed to participate in the study voluntarily. Respondents were free to discontinue participation at any time. Study was reviewed and approved by Internal Review Board at The Ohio State University.

As part of a larger study, each respondent included in this study participated in two face-to-face interviews. One interview focused on utility elicitation for postchemotherapy nausea and vomiting (PCNV) scenarios and the other on breast cancer. To account for ordering and testing effects in the larger project, half of the respondents were randomly assigned to receive PCNV surveys first with the remainder assigned breast cancer surveys first. All interviews were fully structured and interviewer administered (10;11). This study concentrates on VAS and SG preference elicitations for breast cancer scenarios only; therefore, only the protocol used during this interview will be addressed here.

Breast cancer survey consisted of seven sections: One section provided background material regarding breast cancer. The second section asked respondents to read and rank six health states. States were rated using VAS and SG methods, assuming a time horizon of "the rest of your life". The third section recorded sociodemographic information. The fourth section assessed willingness to pay (WTP). The fifth section provided respondents with an opportunity to assess the three different methods used to assess preferences (VAS, SG, and WTP). The sixth section asked respondents to rate the health state scenarios used in the interview. The final section was interviewer completed, assessing the status of the interview.

Ranking

After reading background information on breast cancer, respondents were asked to rank a set of six shuffled cards representing six health states from best to worst. Six health states included three breast cancer scenarios (Figure 2) and three noncancer health states: perfect health, respondent current health, and death (Figure 3). Ranking six health states hypothetically can result in 720 ($6 \times 5 \times 4 \times 3 \times 2 \times 1$) different possible combinations; however, health states were designed to meet the following preference order: perfect health

PERFECT HEALTH
Physical
In the best state of health for your age.
You have not had cancer.
Emotional
Not depressed or anxious.
You feel confident and in control of your life.
Occasionally concerned by the possibility of developing cancer.
Social
You are able to go to work and find it rewarding.
Friends and family enjoy visiting and being visited by you.
Have interests and hobbies.
There is mutual support with partner.
Sexual relations are good.

DEATH
Imagine that you will die, without pain, in your sleep.

YOUR HEALTH
Your current health state.

Figure 2. Three noncancer health states: perfect health (situation P), death (situation D), and your health (situation Y).

Situation C (CANCER CURE)
Physical
- Apart from a small scar - your breast looks pretty much the same as before the surgery.
- As a side effect of therapy you have arthritis – chronic periodic mild joint pain.
- Your mobility is the slightly impaired, e.g., some tasks may involve mild discomfort. You sleep well.

Emotional
- Not anxious or depressed.
- You feel confident and in control in your life.
- You do not fear breast loss or dying of cancer.
- You have occasional concerns that breast cancer will come back.

Social
- You are able to work.
- Friends and family enjoy visiting and being visited by you.
- Interests and hobbies are slightly affected by periods of mild joint pain.
- Partner is supportive.
- Sexual relations are good.

Situation T (CANCER TREATMENT)
Physical
- Apart from small scars on the side of both breasts –your breasts look pretty much the same as before the surgery.
- Your mobility is the same as before cancer diagnosis.
- You sometimes have trouble sleeping.

Emotional
- Not anxious or depressed.
- Your confidence is shaken.
- You have lost feeling of control in your life.
- You worry about dying of cancer.
- You are relieved that you did not lose your breast.

Social
- You are able to work.
- Friends and family are able to visit you.
- Interests and hobbies have declined.
- Partner is supportive.
- Sexual relations have declined.

Situation R (CANCER RECURRENCE)

Physical
- Apart from a small scar - your breast looks pretty much the same as before the surgery.
- You felt a lump in your other breast.
- Your mobility is the same as before cancer diagnosis.
- You do not sleep well.

Emotional
- Anxious and depressed.
- You feel that you no longer have control of your life.
- Overwhelmed by fears of early death.
- Your worst fear of breast cancer recurrence has come true.
- Worried about losing your breast.

Social
- Work is difficult because of concerns about cancer.
- Not able to go out and see people.
- Your interest and hobbies have ceased.
- Partner is not supportive because also feels overwhelmed.
- Sexual relations are nonexistent.

Figure 3. Three cancer health states: breast cancer cure (situation C), breast cancer treatment (situation T), and breast cancer recurrence or "do nothing" (situation R).

https://doi.org/10.1017/S026646230300031X Published online by Cambridge University Press

(situation P) > your health (situation Y) > breast cancer cure (situation C) > breast cancer treatment (situation T) > breast cancer recurrence (situation R) > death (situation D), where A > B means state A is preferred over state B. Respondent's own health—situation Y—nonhypothetical health state was included in study design primarily to familiarize respondents with VAS and SG techniques. Consistent with welfare rule of consumer sovereignty, different preference orders were permitted; however, to ensure respondents understood tasks assigned, respondents providing different preference orders were asked to explain their choices. Each health state was presented on separate colored laminated cards (pale pink).

During ranking of VAS and SG procedures, respondents were asked to assume the following (10):

- You will be in the condition the rest of your life, until you are 74 years old.
- The scenarios apply to you at your current age.
- You have complete medical insurance so that you will not have out-of-pocket expenses for any treatments.
- Your affairs are in order and you have made all preparations for the possibility of your death.

At the completion of ranking, respondents were asked whether they were sure rankings were in order from best to worst situation, and if not, they were provided with an opportunity to change their rankings.

Visual Analogue Scale

After ranking, with cards still laid out in rank order in front of the respondent, VAS was administered. VAS is a vertical, calibrated interval scale labeled with anchors zero and 100 for the least and most desirable health states, respectively (16). Respondents were asked to designate their most desirable and least desirable health states as zero and 100 using marked arrows. The remaining four health states were assigned values by respondents in chronological order—from second to the fifth most preferred health state using the remaining four arrows.

As a validity check, the ordering of the arrows for the six different health states was compared with the ranking responses for consistency. If preference ordering was reversed, this was considered irrational and the discrepancy was brought to the respondent's attention. At this time, the respondent was free to change either the VAS or the ranking to ensure consistency of responses. If the respondent was unable to make the change to provide consistent answers, the interview was terminated.

Because the VAS is designed to be an interval scale, the magnitude of the distances between the arrows placed on the thermometer scale can be compared. Therefore, to be consistent with an interval scale, respondents were also reminded to place the arrows on the scale such that the distances between the arrows reflected the differences perceived between the health states.

Preferences derived from the 0 to 100 point VAS (feeling thermometer) were converted to a zero to 1.0 scale by dividing the number by 100 to facilitate comparisons between preferences acquired using VAS and SG. Because it is expected that not all respondents will rate death as the worst health state, and there is no lower boundary for scores worse than death (34), a linear transformation will be used allowing for negative preference values for health states worse than death. Linear transformation to be used when death is not regarded as the worst health state is described by equations 1 and 2 (2;34), where Q_{raw} is the original value provided by VAS (divided by 100 to transform to 0 to 1.0 scale) or SG, and "D" is the value assigned to death using feeling thermometer or SG methods.

$$Q = (Q_{raw} - D)/(1 - D) \qquad \textbf{(1)}$$

Because the distribution provided using this translation is highly skewed, Patrick and colleagues (34) recommend the following additional transformation for negative values.

$$\begin{aligned} Q' &= Q & \text{if } Q > 0 \quad \text{or} \\ &= Q/(1 - Q) & \text{if } Q < 0 \end{aligned} \qquad \textbf{(2)}$$

This additional transformation results in preferences that range from −1.0 to 1.0, with equal distances in both the positive and negative direction for death. Without the transformation, negative utilities exceeding |−1.0| can significantly impact analyses when comparing mean preference values. For example, if a respondent strongly preferred death (situation D) to breast cancer recurrence (situation R), where, $Q_D = 0.98$ and $Q_R = 0$, then the negative Q weight for breast cancer recurrence using equation 1 would equal —49.0 (i.e., 0–.98/1–.98), where death and perfect health are scaled to zero and 1.0, respectively. The transformation produced from equation 2 results in values that are much less skewed $Q' = -0.98\,(-49.0/(1-[-49.0]) = -0.98)$; however, transformed values (resulting from equations 1 and 2) can no longer be regarded as utilities (34). Both transformations will be used for VAS and SG preference estimates in this study.

Standard Gamble

The top-down titration method was used to elicit SG weights due to its greater efficiency and precision compared with the more commonly used ping-pong approach. Greater standard deviations reported with the ping-pong approach, up to four times larger, would have required a substantive increase in sample size to detect differences in SG weight (7;23;24).

The SG method was presented graphically on computer using Microsoft PowerPoint 2000. The SG question used the same six health state descriptions presented in the ranking and feeling thermometer. The most and least desirable health states, as defined by the respondent, were used as the anchor health states, similar to VAS. Respondents were presented with a series of four sets of slides (four intermediate health states descriptions) in chronological preference order from most to least desirable.

The SG method was formatted to provide respondents with two choices as described by Furlong et al. (16) and is described in more detail elsewhere (10;11). Choice A involved risk, a gamble between respondent's best and worst health states. Choice B was an intermediate health state that occurred with 100% certainty and, therefore, did not involve risk.

Rating of health state scenarios using VAS and SG methods and initial ranking were consistent for all respondents included in study (n = 119).Instrument development, evaluation and validation are described in detail elsewhere (10).Average time required for survey completion ranged from 45 to 60 minutes with a mean of 57.5 minutes.

Analysis

SPSS 10.1 for Windows statistical package was used for all analyses. Means and standard deviations or standard errors of the mean were reported for interval data. The decision to transform all negative Q weights for statistical tests was made a priori consistent with the recommendations of Patrick and colleagues (34) due to the expected skewed nature of the data. Unpaired and one-group t-tests (two-tailed) were used to test for differences in change in utility estimates. The second analysis in this study used one-group t-tests to compare mean ΔSG (ΔVAS) for all respondents to a test value—mean ΔSG (ΔVAS) excluding respondents perceiving fates worse than death (test value)-for three interventions.

The third analysis—threshold analysis—also used one-group t-tests. The objective of the final analysis was to determine the smallest percentage of respondents required to perceive fates worse than death to result in statistically significant ΔQ when compared with ΔQ for respondents perceiving death as the least desirable health state (i.e., test value was ΔQ for respondents perceiving death as worst health state). Mean ΔQ for threshold analysis was obtained by randomly and incrementally dropping respondents perceiving fates worse than death from data set until no significant difference was reported. Based on convention, all tests conducted assumed α level of .05 significance (two-tailed).

RESULTS

Response Rate

Data were collected from March 31 to June 28, 2000. Of the 146 interviewees scheduled, 20 cancelled their interviews, claiming unavailability due to other commitments. Of the remaining 126 respondents interviewed, 7 were dropped: one due to distress (her mother recently had died of breast cancer), two failed to keep scheduled interviews, and four respondents were dropped due to lack of understanding of health utility measures used and belief in scenarios. The remaining 119 responses were included in study for analysis.

Sociodemographics

The majority of respondents were young, single, American, full-time graduate students (97 of 119) with gross incomes less than $20,000 per year (66 of 119). This salary range was expected, given that most respondents were graduate teaching assistants and full-time students (82%). In addition, the majority of respondents were single (61%), did not have children (84%), and were nonsmokers (88%). Nonstudent participants were also highly educated, with over 95% (19 of 22) completing a college education or higher.

Fates Worse Than Death

First analysis, unpaired t-test, tested for differences in ΔVAS and ΔSG for respondents perceiving fates worse than death (i.e., negative Q values for situation R after rescaling) versus those that perceived death as the least desirable health state (i.e., no negative Q values after rescaling). Situation R was perceived as a fate worse than death for 15% of respondents (18 of 119, Table 1). Differences in ΔSG weights estimated for three hypothetical interventions between the two groups were statistically significant: situation T versus R, $t(19.02) = -11.915$, $p < .001$; situation C versus T, $t(18.82) = -2.302$, $p = .03$; and situation C versus R, $t(20.14) = -13.68$, $p < .001$. Unpaired t-test results for ΔSG shows

Table 1. Unpaired t-Test: ΔSG and ΔVAS Situation R Rated as Worse than Death (D ≻ R, N = 18) Versus Death Least Desirable (D ≻ R, N = 101)

Δ Therapy	R ≻ D D ≻ R	Unpaired t-test ΔSG (SEM)	Unpaired t-test ΔVAS(SEM)
Situation T	N = 18	1.225 (.083)	.540 (.074)
versus R	N = 101	.211 (.020)**	.303 (.016)*
Situation C	N = 18	.126 (.030)	.293 (.049)
versus T	N = 101	.056 (.007)*	.214 (.014)
Situation C	N = 18	1.350 (.076)	.832 (.069)
versus R	N = 101	.266 (.023)**	.517 (.017)**

*p value < .05.
**p value < .001.

Table 2. One-Group t-Test: Comparing Mean ΔSG for All Respondents to a Test Value—ΔSG Excluding Respondents Perceiving Fates Worse Than Death—for Three Interventions

Δ Therapy	ΔSG(SEM) N = 119	ΔSG test value D ≻ R, N = 101
Situation T versus R	.364 (.039)	.211**
Situation C versus T	.066 (.008)	.056
Situation C versus R	.430 (.042)	.266**

**p value < .001.

Table 3. One-Group t-Test: Comparing Mean ΔVAS for All Respondents to a Test Value—ΔVAS Excluding Respondents Perceiving Fates Worse Than Death—for Three Interventions

Δ Therapy	ΔVAS(SEM) N = 119	ΔVAS test value D ≻ R, N = 101
Situation T versus R	.339 (.019)	.303
Situation C versus T	.226 (.014)	.214
Situation C versus R	.565 (.020)	.517*

*p value < .05.

respondents perceiving fates worse than death resulted in statistically significantly greater ΔQ weights, ranging up to five times more than respondents who did not perceive fates worse than death. Furthermore, allowance of negative utilities resulted in ΔSG differences greater than 1.0 for two of three analyses-situation T versus R and C versus R—the former being an incremental analysis. This finding is possible because allowing for negative utilities resulted in a utility range of 2 units, not 1.

Differences in ΔVAS weights estimated for three hypothetical interventions between respondents perceiving fates worse than death and those who did not were statistically significant: situation T versus R, $t(18.69) = -3.127$, $p = .006$ and situation C versus R, $t(19.11) = -4.454$, $p < .001$. Unpaired t-test results for ΔVAS estimated for the two groups shows respondents perceiving fates worse than death resulted in statistically significantly greater ΔQ weights, in two of three analyses, ranging up to two times more than respondents who did not perceive fates worse than death. Unlike ΔSG differences, no ΔVAS exceeded 1.0 for the three analyses (Table 1).

The second analysis, a one-group t-test, where the test value was mean ΔQ for respondents perceiving death as the worst health state—excluding respondents perceiving fates worse than death (n = 101). Test value was compared with mean ΔQ for all respondents in sample (n = 119). Results are presented in Table 2 (ΔSG) and Table 3 (ΔVAS). One-group t-test for mean ΔSG showed excluding respondents perceiving fates worse than death from the sample resulted in a statistically significant smaller changes in utility estimates for two of the three therapy interventions: situation T versus R, $t(118) = 3.881$, $p < .001$; and situation C versus R, $t(118) = 3.904$, $p < .001$ (Table 2). One-group t-test for mean ΔVAS showed excluding respondents perceiving fates worse than death from the sample resulted in a statistically significant smaller changes in utility estimates for one of the three therapy interventions: situation C versus R, $t(118) = 2.347$, $p = .021$ (Table 3).

The third analysis, an extension of the previous, was only performed for statistically significant differences for one-group t-tests (Tables 4 and 5). Threshold analysis (one-group t-tests) performed for mean ΔSG for situation T versus R, $t(107) = 2.096$, $p = .039$, and situation C versus R, $t(107) = 2.145$, $p = .034$, showed in both cases the smallest proportion of respondents required in the sample to perceive fates worse than death to make a statistical difference in ΔSG was 6.5% (7 of 108). The result was, when only

https://doi.org/10.1017/S026646230300031X Published online by Cambridge University Press

Table 4. Effect Size (ES) Estimations for Three Interventions for Utility Estimations Using Standard Gamble (SG)[a]

Δ Therapy	ΔSG (SD range) N = 119 Effect size index: *d*	ΔSG (SD range) D ≻ R, N = 101 Effect size index: *d*
Situation T versus R	.364 (.164−.495) $d = 1.113$	.211 (.120 − .251) $d = 1.137$
Situation C versus T	.066 (.013 − .164) $d = 0.505$	.056 (.078 − .120) $d = 0.565$
Situation C versus R	.430 (.113 − .494) $d = 1.405$	.266 (.078 − .251) $d = 0.618$

[a]Median SD value was used from SD range to estimate *d*.

Table 5. Effect Size (ES) Estimations for Three Interventions for Utility Estimations Using VAS[a]

Δ Therapy	ΔVAS (SD range) N = 119 Effect size index: *d*	ΔVAS (SD range) D ≻ R, N = 101 Effect size index: *d*
Situation T versus R	.339 (.207−.201) $d = 1.660$	.303 (.192 −.153) $d = 1.756$
Situation C versus T	.226 (.171 − .207) $d = 1.196$	.214 (.159 − .192) $d = 1.219$
Situation C versus R	.565 (.171 − .201) $d = 3.034$	.517 (.159 − .153) $d = 3.314$

[a]Median SD value was used from SD range to estimate *d*.

6.5% of respondents perceived fates worse than death, ΔSG weights remained statistically significant and resulted in mean ΔQ = 0.07 (SEM = .032) and 0.08 (SEM = .035) for situation T versus R and situation C versus R, respectively.

Threshold analysis for situation C versus R ΔVAS preference weights showed the minimal proportion of respondents required to perceive fates worse than death to result in statistically significantly different ΔQ was 12.9% (15 of 116), $t(115) = 2.016$, $p = .046$. When just over 10% of respondents perceived fates worse than death, ΔVAS weights remained statistically significant and resulted in mean ΔQ = 0.039 (SEM = .019).

The decision to transform or not to transform VAS weights made no difference to the conclusions drawn, because only one data point was less than −1.0 (−1.3). However, raw SG weights were highly skewed, situation R SG-weights ranged from +1.0 with one score equal to −49.0. As a result, raw data was not presented for VAS weights because it would not make any difference in the study conclusions or for SG weights because parametric tests would no longer be suitable given the highly skewed nature of the raw data.

DISCUSSION

In this study, we investigate the impact of including (versus excluding) fates worse than death on utility measurement for a chronic condition—breast cancer. Utility elicitation required respondents to read and rank six health state scenarios, including death and perfect health using the VAS and SG top-down titration method. Consistent with consumer sovereignty, respondents were assumed to be the best judges of their own welfare; therefore, fates worse than death were permitted. Death and perfect health are typically scaled to zero and 1, respectively. Consequently, linear transformations were performed, allowing for negative utilities (34).

Similar to previous study findings (34), results of this study show that SG and VAS methods can be successfully adapted to acquire preference values for health states considered worse than death by healthy respondents. Respondents perceiving fates worse than death were required using the SG elicitation method to choose one of two options: a gamble between perfect health and breast cancer recurrence as described by health state scenarios or certain death. The majority of respondents included in this study were healthy educated young women, who had no apparent difficulty when asked to imagine this scenario.

Although cancer health state scenarios used in this study were formulated to be preferable to death (with input from oncology clinicians and subsequent changes made after pilot testing), 15% of respondents perceived situation R to be worse than death. Their inclusion resulted in larger ΔQs, independent of the utility elicitation method used. Larger ΔQs observed were explained by removal of a potential floor effect by allowance of negative utilities. Thus, the utility scale was expanded from the traditional 0 to 1.0 scale to a −1.0 to +1.0 scale. The doubling of scale breadth has additional implications when considering minimally important clinical differences (MICD). Based on Cohen's (8) conventional benchmark of $\frac{1}{2}$SD equating to a moderate effect, the two scales would require different criteria to determine whether statistical differences equated to clinically meaningful changes in patient quality of life or utility. For example, using $SD = range/6$, for a 1 unit scale is $SD = 1/6\,([1.0 - 0]/6)$, then for a two-unit scale, the $SD = 1/3([1 + 1]/6)$. Therefore, generic MICD for one-unit and two-unit utility scales are approximately 0.083 and 0.166 utils (i.e., 2 × MICD for one-unit scale), respectively, assuming no end scale aversion. Therefore, when discussing MICD specification of breadth of utility scale used or allowance of negative utilities is important. An independent of rule of thumb used for MICD, as the computation above suggests, is that MICD for a two-unit scale is twice that of the traditional 0 to 1.0 utility scale.

In this study, when accounting for different SD resulting from including versus excluding respondents perceiving fates worse than death, larger ΔQs did not result necessarily in a larger effect size index (d): unitless number calculated by dividing mean ΔQ by standard deviation from their respective populations, assuming equal standard deviations (Table 4). Because standard deviations varied substantively for mean Q values in this study, the median SD values were used from the ranges listed in Tables 4 and 5 to estimate d. This expanded utility scale had no apparent effect on d for incremental mean ΔQ (Tables 4 and 5). For situation C versus R (Table 4), as anticipated because ΔQ estimates used entire utility scale, the blunting of ΔQ seen on the 0 to 1.0 scale was not present when negative utilities are permitted. The result was a "large" (>0.8) versus moderate (>0.5) effect size index. Furthermore, d was twice as large when negative utilities were permitted for SG. For VAS data, the same effect on d was not observed. This finding may be attributed to the lack of risk involved with VAS preference scores.

Furthermore, allowing for health states worse than death results in a complete range of utility values consistent with respondent preferences and consumer sovereignty. In this study, on an individual level, all respondents interviewed were able to provide consistent preference rankings for SG and VAS methods. This result is contrary with previous findings reporting consistent findings for group level data only (30;34). Possible explanations for this discrepancy could be explained first by sample differences. Respondents in this study constituted well-educated women. Second, in this study, SG weights were acquired using a computerized format (Microsoft PowerPoint 2000) and the top-down titration method as opposed to the ping-pong approach traditionally reported (16). Greater efficiency and precision reported with the top-down titration method (7;23;24) suggest that utility estimation with this approach would reduce cognitive burden of survey completion. Third, during face-to-face interviews, health state card scenario ranking and VAS ratings were evaluated for rank consistency by interviewer. Any discrepancies were brought to the attention of the respondent. In these cases, subjects appeared to understand the inconsistency in their responses and were asked to either change their rankings or the values assigned on the feeling thermometer. Respondents did not appear to have any difficulty doing this. Although this step added to the duration of the interview, we believe that the benefits gained in improving the validity of responses was certainly worth the cost.

The face-to-face interview process, although helpful to ensure respondent understanding of utility methods, is time-consuming. On average, each interview required one hour to complete, however, the results presented in this article were part of a larger study, which included elicitation of respondent willingness to pay for a health improvement. Also, as part of a larger study each respondent participated in two face-to-face interviews, the second eliciting utility preferences for an acute condition. First interviews were substantially longer than the second. This was attributed to the time required to learn the new method. For some respondents, SG elicitation in the second interview required a fraction of time from that of the first.

The present study examined the effect of fates worse than death on ΔQs estimations in breast cancer. As expected, the impact was an increase in ΔQ by removal of floor effect for fates perceived as worse than death, and in some cases, ΔQ exceeded 1.0 and was as high as 1.98 (Tables 2 and 3). For example, for one study respondent:

Subject 1:

$$Q_C = 0.2,\ Q_R = -0.75, \text{ where death} = 0, \text{ and perfect health} = 1.0, \text{ then,}$$
$$\Delta Q_{C\text{–}R} = 0.2 - (-0.75) = 0.95 \text{ utils.}$$

For second study, respondent not perceiving fates worse than death:

Subject 2:

$$Q_C = 1.0,\ Q_R = 0.97, \text{ then,}$$
$$\Delta Q_{C\text{–}R} = 1.0 - 0.97 = 0.03 \text{ utils.}$$

We also expected a subsequent increase in ΔQALY with inclusion of negative utilities. However, contrary to expectations, further analysis (not presented) showed that a large ΔQ did not necessarily result in large ΔQALY. This finding can be explained by the predominant impact of time on QALY computation: assuming the same patient data as above and assuming that the chronic condition is for the rest of the respondent's life (mean duration of life in the Midwest is 74 years) and breast cancer recurrence results in an additional 2 years of life:

Subject 1: Age = 59 years

$$\begin{aligned}\Delta QALY_{C\text{–}R} &= Q_C(74 - \text{age}) - Q_C(2 \text{ years})\\ &= (0.2)(74 - 59) - (-0.75)(2) = 4.5 \text{ QALYs}\end{aligned}$$

Subject 2: Age = 33 years

$$\begin{aligned}\Delta QALY_{C\text{–}R} &= Q_C(74 - \text{age}) - Q_C(2 \text{ years})\\ &= (1.0)(74 - 33) - (0.97)(2) = 39.06 \text{ QALYs}\end{aligned}$$

However, if breast cancer recurrence resulted in a condition enduring the rest of the respondent's life, then:

Subject 1:

$$\Delta QALY_{C-R} = \Delta Q_{C-R}(74 - \text{age}) = (0.95)(15) = 14.95\,\text{QALYs}$$

Subject 2:

$$\Delta QALY_{C-R} = \Delta Q_{C-R}(74 - \text{age}) = (0.03)(41) = 1.23\,\text{QALYs}$$

Large ΔQs do not necessarily result in substantial ΔQALY estimations, especially if respondents reporting negative utilities were older and, therefore, the number of additional life years to be gained is less. In this study, respondents reporting fates worse than death tended to be older; however, the difference was not statistically significant, $t(18.62) = -1.24$, $p = .227$. Furthermore, if negative utilities are estimated for a chronic condition that is short-lived (e.g., 2 years), then the impact of a large ΔQ is blunted by the short duration of the condition.

Sensitivity of QALY estimation to effect of time, for example treatment of acute versus chronic conditions, has been reported previously (3;33). As ΔQALY calculations elude to, allowance of negative utilities for treatment of acute conditions, such as post chemotherapy nausea and vomiting, may result in substantive changes in ΔQ; however, the effect on ΔQALY is expected to be nominal due to its short duration—average of 0.008 years (3 days). The results of this study also support that, within the category of chronic conditions, sensitivity of QALYs to time can result in inclusion of negative utilities having no significant effect on ΔQALY if duration of fates worse than death is of short duration. For example, in this study, mean $\Delta QALY_{C-R} = 40.5$ QALY (all respondents); however, when excluding respondents perceiving fates worse than death, $\Delta QALY_{C-R} = 41.15$ QALY ($n = 101$), $t(118) = -0.856$, $p = .394$. Moreover, effect of time on ΔQALY would be further compounded by the selected discount rate (18;33).

Limitations

The primary limitation of this study was that the sample consisted of young healthy well-educated female graduate students; therefore, the ΔQ weights and WTP bids acquired cannot be generalized to other populations or patients. However, the goal of this article was to evaluate the impact of fates worse than death on utility estimates, not the provision of population utility estimates. The study also used the SG top-down titration method; other methods, for example, the SG ping-pong method, may result in different Q weights. Other limitations are similar as for other studies using hypothetical scenarios such as information bias-responses can be affected by how information is presented. A panel of researchers reviewed surveys, and changes were also made to incorporate respondent comments from the pilot study. Hypothetical bias-respondents perceive scenarios as unrealistic. The study presented part of a larger study. Respondents who were unable to accept health state scenarios provided (7 of 126) were excluded from analysis.

CONCLUSION

If welfare economics is assumed to be the underlying theory of CUA, then it would follow that the method used to acquire Q weights should be consistent with the requirements of the welfarist approach—theory of utility. Assuming consumers are the best judges of their own welfare, respondents were asked to rate and rank all health state scenarios, including traditional anchor states: perfect health and death. Cancer scenarios used in this study were designed to be preferred over death. Using VAS and SG top-down titration methods, 15%

of respondents perceived breast cancer scenario to be worse than death. As anticipated, inclusion of these respondents resulted in significant increases in ΔQ independent of elicitation method used. Furthermore, an ΔQ exceeding 1.0 was observed for the SG method. Due to the sensitivity of QALY methodology to time, inclusion of negative utilities did not necessarily translate to significant ΔQALY. This study showed that the successful adaptation of current utility elicitation methods was possible for acquiring negative utilities for fates perceived as worse than death. The effect of including negative utilities can have a significant impact on ΔQ. A large ΔQ is expected to translate to a large ΔQALY for chronic conditions such as rheumatoid arthritis or renal dialysis.

Policy Implications

CUA is recommended for use in allocation resource decision-making in health care (19). Typically, perfect health and death are assumed to be the best and worst health states and are assigned values of 1.0 and zero, respectively. However, the results of this study support previous findings in the literature, indicating that some individuals do perceive some health states as worse than death (21;22;27;28;34;38). Exclusion of fates worse than death can have a significant impact on ΔQ weight estimates, resulting ΔQALY estimates, and subsequent cost per QALY calculations. In summary, excluding fates worse than death from CUA can underestimate the benefits of a health care intervention even if only a small proportion of individuals (15%) perceive it as such. Given the increased use of CUA in decision-making, policy makers need to familiarize themselves with anchors permitted in ΔQALY estimates, particularly with chronic conditions, because failure to include fates worse than death can result in higher (i.e., less favorable) cost per QALY ratios.

REFERENCES

1. Arnesen T, Nord E. The value of DALY life: Problems with ethics and validity of disability adjusted life years. *BMJ*. 1999;319:1423-1425.
2. Badia X, Monserrat R, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade off values for EQ-5D health states. *Med Decis Making.* 2001;21:7-16.
3. Bala MV, Zarkin GA. Are QALYs an appropriate measure for valuing morbidity in acute diseases? *Health Econ Lett*. 1999;3:10-13.
4. Birch S, Gafni A. Cost effectiveness and/utility analysis. Do current decision rules lead us to where we want to be? *J Health Econ.* 1992;11:279-296.
5. Brazier J, Deverill M. A checklist for judging preference based measures of health related quality of life: Learning from psychometrics. *Health Econ*. 1999;8:41-51.
6. Ciampi A, Silberfield M, Till JE. Measurement of individual preferences: The importance of situation specific variables. *Med Decis Making.* 1982;2:483-95.
7. Clarke AE, Goldstein MK, Garber AM, Lenert LA. The effect of assessment method and respondent population on utilities elicited for Gaucher disease. *Qual Life Res.* 1997;6:169-184.
8. Cohen J. *Statistical Power Analysis For The Behavioral Sciences*. 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers; 1988.
9. Ditto PH, Druley JA, Moore KA, Danks JH, Smucker WD. Fates worse than death: The role of valued life activities in health state evaluations. *Health Psychol.* 1996;15:332-343.
10. Franic DM. *Relating Quality Adjusted Life Years To Contingent Valuation: Acute Versus Chronic Conditions* [dissertation]. Columbus (OH): The Ohio State University; 2000.
11. Franic DM, Pathak DS, Gafni A. Are health states “timeless”: A case study of an acute condition-post chemotherapy nausea and vomiting. *J Eval Clin Pract*. 2003;69-82.
12. Froberg DG, Kane RL. Methodology for measuring health-state preferences-1: Measurement strategies *J Clin Epidemiol.* 1989;42:345-354.
13. Froberg DG, Kane RL. Methodology for measuring health-state preferences-2: Scaling methods. *J Clin Epidemiol.* 1989;42:459-471.
14. Froberg DG, Kane RL. Methodology for measuring health-state preferences-3: Population and context effects. *J Clin Epidemiol.* 1989;42:585-592.

15. Froberg DG, Kane RL. Methodology for measuring health-state preferences-4: Progress and research agenda. *J Clin Epidemiol.* 1989;42:675-685.
16. Furlong W, Feeney D, Torrance GW, et al. *Guide to design and development of health state utility instrumentation*. Center Health Economics Policy Analysis Working Paper Series. Paper No. 90-9. McMaster University: Hamilton, Ontario; 1990.
17. Gafni A. Willingness to pay in the context of an economic evaluation of healthcare programs: Theory and practice. *Am J Manag Care.* 1997;3(suppl):S21-S32.
18. Gold MR, Franks P, McCoy KI, Fryback DG. Toward consistency in cost-utility analysis. *Med Care.* 1998;36:778-792.
19. Gold MR, Siegel JE, Russell LB, Weinstein MC.*Cost-Effectiveness in Health and Medicine*. Oxford: Oxford University Press; 1996.
20. Gold MR, Stevenson D, Fryback DG. HALYs, QALYs and DALYs, oh my: Similarities, differences in summary measures of population health. *Annu Rev Public Health.* 2002;23:115-134.
21. Haigh THB, Scott DA, Wickett LI. The rational zero point for an illness index with ratio properties. *Med Care.* 1986;24:113-124.
22. Imbasciati E, Surian M. When dialysis becomes worse than death. *Nephrol Dial Transplant*. 1996;11:1484-1485. Letter.
23. Lenert LA, Cher DJ, Goldstein MK, et al. The effect of search procedures on utility elicitations. *Med Decis Making.* 1998;18:76-83.
24. Lenert L, Morss S, Goldstein M, et al. Measurement of the validity of utility elicitations. *Med Care.* 1997;35:915-920.
25. Llewellyn-Thomas H, Sutherland HJ, Tibishirani R, et al. The measurement of patients' values in medicine. *Med Decis Making.* 1982;2:449-462.
26. Llewellyn-Thomas HA, Sutherland HJ, Tibshirani R, et al. Describing health states methodologic issues in obtaining values for health states. *Med Care.* 1984;22:543-552.
27. Macran S, Kind P. Death and the valuation of health related quality of life. *Med Care.* 2001;39:217-227.
28. Mitchell C. When living is a fate worse than death. *Newsweek.* 2000;136:12.
29. National Alliance of Breast Cancer Organizations (NABCO). *Facts about breast cancer in the USA*. Available at: http://www.nabco.org/resources/facts/usafacts.html. Accessed February 20, 2000.
30. Nord E. Methods for quality adjustment of life years. *Soc Sci Med.* 1992;34:559-569.
31. O'Brien BJ. Health state utility anchors: Being clear on what "1" means. *Med Decis Making*. 1997;17:352-353.
32. O'Hanlon M, Fox-Rushby J, Buxton MJ. A qualitative and quantitative comparison of the EuroQol and time-trade off techniques. *Int J Health Sci.* 1984;5:85-97.
33. Pathak DS. QALYs in health outcomes research: Representation of real preferences or another numerical abstraction. *J Res Pharm Econ.* 1995;6:3-27.
34. Patrick DL, Starks HE, Cain KC, et al. Measuring preferences for health states worse than death. *Med Decis Making*. 1994;14:9-18.
35. Pearlman RA, Cain KC, Patrick DL, et al. Insights pertaining to patient assessments of states worse than death. *J Clin Ethics*. 1993;4:33-41.
36. Post PN, Stigglebout AM, Wakker PP. The utility of health states after stroke a systematic review of the literature. *Stroke.* 200l;32:1425-1429.
37. Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preferences for health outcomes: Comparison of assessment methods. *Med Decis Making*. 1984;4:315-329.
38. Sessa A. When dialysis becomes worse than death. *Nephrol Dial Transplant.* 1995;10:1128-30.
39. Spector PE. *Research Designs*. Sullivan JL, Niemi RG, eds. Beverly Hills, CA: Sage Publications; 1987.
40. Torrance GW. Social preferences for health states: An empirical evaluation of three measurement techniques? *Socio-Econ Plan Sci.* 1976;10:129-136.
41. Torrance GW. *Health States Worse Than Death.* In: von Eimeren W, Engelbrecht R, Flagle CD, eds. Third International Conference on System Science in Health Care, Berlin, Germany: Springler-Verlag; 1984:1085-1089.
42. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ*. 1986;5:1-30.