Valuing the Q in QALYs: Does Providing Patients' Ratings Affect Population Values?

Robert P. Murphy

University of Stirling, and Department of Health, Dublin, Ireland

Christopher J. Boyce

University of Stirling

Paul Dolan

London School of Economics and Political Science

Alex M. Wood

London School of Economics and Political Science

Reference: Murphy, R. P., Boyce, C. J., Dolan, P., & Wood, A. M. (accepted). Valuing the Q in QALYs: Does Providing Patients' Ratings Affect Population Values? *Health Psychology*

© 2019, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/hea0000806

Abstract

Objective: QALYs are used to measure the health benefits associated with treatments. QALYs are derived from objective mortality data weighted by assessments made by the general population of the impact on health-related quality of life associated with particular health states. In this study a simple change is introduced to improve the validity of QALYs by giving raters information about how people living in the health states rate the health states.

Methods: Participants from the general population (N = 155) judged three health states using a standard valuation technique after being randomly allocated to one of two groups. The intervention group was given patients' mean ratings of their own health states from worst to best imaginable health (0-100 scale) before providing their valuations, whilst the control group was given this information only after providing their valuations. The participants in both groups also indicated whether patients' mean ratings were higher, broadly similar, or lower than they previously expected.

Results: When the mean ratings given by patients were higher (lower) than expected, participants in the intervention group provided significantly higher (lower) valuations than participants in the control group. These findings show that participants adjust their valuations of a health state in the direction of the appraisals of those experiencing that state.

Conclusions: Insofar as policy-makers are committed to valuing health states using valuations given by people from the general population, it is desirable to elicit more informed values by providing people with information on how patients rate those states.

Keywords: Quality-Adjusted Life Years, judgment, decision making, resource allocation

Valuing the Q in QALYs: Does Providing Patients' Ratings Affect Population Values?

Quality-Adjusted Life Years (QALYs) combine two attributes of value – quality of life and length of life - into a single number. Each health state is assigned a value between zero (for death) and one (for full health) and that value is then multiplied by how long the state lasts. For example, one year of life in full health is equivalent to one QALY. When combined with the costs of intervention, QALYs allow policy-makers to determine an efficient allocation of resources (i.e., the one that generates the most QALYs given the resources available to healthcare). The QALYs method is accepted internationally as the gold standard in comparing the impact of medical treatments (NICE, 2013) and it is the preferred measure of benefit for health technology assessments in England, Wales, Scotland and the Netherlands (Karimi, Brazier, & Paisley, 2017b). Concerns have been raised, however, about how best to capture the Q in the QALY; that is, how best to value quality of life. It is widely accepted that the preferences of the general public should be taken into account, but there are doubts about whether they take sufficient account of patients' experiences of specific health states (Dolan & Kahneman, 2008; Drummond et al., 2009). This paper is the first to test the impact of a simple, quick, and low-cost way of briefing participants in health valuation surveys about the experiences of patients by providing patients' mean ratings of their conditions. It is also the first study to test whether the effect of giving patients' mean ratings of their conditions to participants from the general population interacts with participants' prior beliefs about how such patients appraise their conditions.

When computing QALYs for the purposes of health technology assessment, expected length of life (the LY in the QALY) is established by epidemiological research, whereas quality of life (the Q in the QALY) is estimated by asking people what value they ascribe to specific health states. Most countries elicit valuations from members of the general population, who are asked their preferences over different states of health that vary according to their severity (e.g., in terms of physical functioning and mental health). The use of population values is consistent with the idea that those who pay for publicly funded healthcare, and who are also the prospective beneficiaries of it, should have some say in how the benefits afforded by that expenditure are valued.

Valuations of the Imagined Compared to the Experienced

A potential problem is that many members of the general population have no direct experience and only limited knowledge of the health states being judged (Kind & Dolan, 1995). Furthermore, this lack of knowledge is not addressed in health valuation studies, where the descriptions of health states are typically limited to information on salient dimensions of health (Stiggelbout & De Vogel-Voogt, 2008). Information on patients' experience of living in the health states is never provided. As a result, it has been argued that the valuations produced using the current approach are largely uninformed by what it is like to be in the health states so described (Brazier et al., 2005; Karimi, Brazier, & Paisley, 2017a).

Those with knowledge of what it is like to live in a particular health state may value that state very differently to people without knowledge of it. A review of 38 studies by De Wit, Busschbach, and De Charro (2000) found that patients provided higher values than the public in 22 studies (58%); the public provided higher values than patients in two (5%); there was no difference in 11 (29%), and there were contradictory findings in three (8%). A recent meta-analysis by Peeters and Stiggelbout (2010) also found that, across all studies examined, patients on average gave higher values. Differences in nonpatients' exposure to health states (through family, friends, work or volunteering) is associated with differences in perceptions of and valuations of health states (Schkade & Kahneman, 1998; Ubel et al., 2001).

Stiggelbout and De Vogel-Voogt (2008) argued that interpretation and judgment are two key steps in the process of health state appraisal. According to the stress-appraisal model of Lazarus and Folkman (1984), a person's primary appraisal upon encountering a health-related

VALUING THE Q in QALYs

stimulus (in this instance a written description of a health state) will be whether the health state represents harm, a threat, or a challenge. The primary appraisals of people with and without experience of what it is like to live in the health state are likely to differ. In the secondary appraisal, according to Lazarus and Folkman, people consider the resources (internal and external) available to them to cope with the health state, and again there are likely to be different perceptions. For example, nonpatients may be unaware of coping strategies that patients use and of the possibilities of adapting to a health state. As a result, for many health states, people with knowledge of what it is like to live in a given health state are likely to have less negative primary and secondary appraisals of that state than people without such knowledge, and hence to value that health state less severely.

Obtaining Experience-Informed General Population Valuations

One way to overcome a lack of knowledge among the general population is to provide respondents with information on patients' appraisals of their experience of living with a health state before respondents judge that health state. Indeed, studies have called for patients' judgments of what it is like to live in particular health states to be provided to the general population when they judge those health states (e.g., Fryback, 2003; Menzel, Dolan, Richardson, & Olsen, 2002). This position is referred to as a "third position" that implies the use of "informed general population values" (Brazier et al., 2005). The approach is consistent with recommendations of the influential Washington Panel on Cost-effectiveness in Health and Medicine (Gold, Siegel, Russell, & Weinstein, 1996) which are, firstly, that fully informed members of the community would provide the best articulation of society's preferences for particular health states (Brazier et al., 2005), and secondly, that there is a need to create a better understanding in the general population of the experiences of patients living in different health states (Stiggelbout & De Vogel-Voogt, 2008). Indeed, developing ways of briefing people from the general population on the experiences of those in particular health states is identified as one of thirteen research priorities for the future of the QALY by the Consensus Development Group (Drummond et al., 2009). Despite this, little effort has gone into making people from the general population more informed when undertaking health valuation studies (Brazier et al., 2005). In one notable exception, McTaggart-Cowan, Tsuchiya, O'Cathain, and Brazier (2011) examined the effect of providing the general population with audio recordings of three patients discussing adapting to life with rheumatoid arthritis. They found that provision of this information positively influences the valuation of this health condition. Audio recordings of patients discussing their experiences provide one way of briefing members of the general population on the experiences of those in particular health states, but this method is time and resource intensive (Dolan, Kavetsos, & Tsuchiya, 2013).

The Current Study

This study explored a novel way of briefing people from the general population on the experiences of people with a health condition, through providing patients' mean ratings of their condition, on a 0-100 scale, from worst to best health imaginable. For each of the three conditions that meet the study selection criteria – arthritis, colostomy and tinnitus – participants were provided with patients' mean rating of the health condition and asked whether the patients' mean rating of a given condition was higher, about the same, or lower than expected. The control group was briefed after rating the conditions to establish whether their prior belief of patients' mean rating was accurate, whereas the intervention group was briefed before rating the conditions to establish whether their prior belief of a condition would be worse than it is, we expected the mean valuation of the condition to be higher in the intervention group than in the control group because the intervention group's misconceptions

would be corrected in an upward direction. For participants who believed patients' mean rating of a condition would be better than it is, we expected the opposite pattern, and we did not expect an impact for participants whose prior beliefs of patients' mean rating were accurate. Thus, we expected to find a cross-over interaction, depending on the expectations about the patient ratings.

Methods

Participants

This study was conducted online in the United States using Amazon Mechanical Turk (MTurk). MTurk has been judged to be a source of high quality data (Crump, McDonnell, & Gureckis, 2013) and online health valuation surveys have been deemed reliable (Bansback, Tsuchiya, Brazier, & Anis, 2012). Eligible participants were adult (aged 18 and over) members of MTurk, had not undertaken pilot surveys for this study, and had completed an informed consent form. To support the quality of data collection two additional criteria, often used by MTurk requesters, were applied: (a) participants needed to have previously completed at least 100 surveys and (b) they were required to have a 95% approval rating. Participants who completed the survey were compensated for their time, in line with MTurk norms. Ethics approval was provided by the University of Stirling Management School's Ethics Committee.

Consenting participants (n = 227) accessed the survey between 18th August to 4th September 2016. Of these, six declined to participate after reading the introduction, three in the intervention group discontinued the survey before receiving the intervention, 22 did not complete the survey (split evenly between the control and intervention groups), and 196 completed the survey. Two a priori criteria for exclusion from analysis were set: (a) not giving sufficient consideration to the task, defined as completion in less than 50% of the median time; and (b) responding with logically incoherent judgments, defined as rating one health condition at the anchor 100 (*best health you can imagine*). Of the 196 people who completed the study 155 met the inclusion criteria (67 control group; 88 intervention group). The Consolidated Standards of Reporting Trials (CONSORT) participant flow chart is shown in Figure 1.

[Insert Figure 1 around here]

To check that the study results were not dependent on differences in exclusions on criterion (b) between the control and intervention groups (23.40% versus 7.78%), the tests of the study hypothesis were re-run treating responses from these participants as missing data and imputing their valuations (differences did not seem substantial for noncompletion rates, 10.48% versus 12.07%, nor completion in less than 50% of median time, 5.32% versus 5.89%). For these participants the valuation was estimated through a multiple imputation technique (Rubin, 1987) which imputes a series of missing values based on estimates from other observed variables and which more appropriately accounts for the statistical uncertainty in the imputations than many other commonly used imputation techniques, such as sample wide or item averages. Specifically, multiple imputation using chained equations (MICE: White, Royston, & Wood, 2011) was used. Details on the specific MICE applied are presented in the Supplementary File. As discussed in the Results, this did not have implications for the study conclusions.

Measures and Materials

Valuation technique. This study used the EQ VAS valuation technique, a standard method used in health-related quality of life research (Williams, 1990 p. 200). Participants "warmed-up" by describing their current health using the two parts of the EQ-5D-3L, namely (a) its descriptive system (a generic health state descriptive system) and (b) its visual analogue scale (EQ VAS) – this is a vertical rating scale (thermometer) where the anchors are 0 (*worst health you can imagine*) and 100 (*best health you can imagine*). After this, participants used the EQ VAS scale to judge the hypothetical health conditions. A staff member from the EuroQol Research Foundation checked this study's presentation of the instrument in a draft of the questionnaire and approved it subject to changing the format of one word (to ensure consistency with the instrument).

Health conditions. This study used the health conditions rheumatoid arthritis, colostomy, and tinnitus because they were the only health conditions which met the five study inclusion criteria. As international inventories of patients' appraisals of their actual health conditions do not exist, two meta-analyses (Dolders, Zeegers, Groot, & Ament, 2006; Peeters & Stiggelbout, 2010) and a review (De Wit, Busschbach, & De Charro, 2000) of patients' and nonpatients' judgments of ill health were searched to identify conditions which met the five selection criteria. The criteria were:

1. The health scenario was of a specific condition (rather than a generic health state) that was not predominantly specific to one gender.

2. The condition was appraised by patients of the condition rather than patients of a different condition or proxies such as patients' family members or health professionals.

3. The condition was rated using the EQ VAS.

4. Studies showed that patients rate the condition higher than people from the general population do.

5. The condition was rated by a minimum of 50 patients. The intention was for participants to take the information provided into account, and not to discount it as unreliable. A sample size of at least 50 was a priori selected as a "rule of thumb" minimum sample which participants may require to regard the information as reliable. It also turned out to be a natural break point, as after criteria 1 to 4 were applied only four conditions remained; laryngeal-cancer which was rated by only 20 patients and the three conditions used in this study which were rated by more than 50 patients.

Questionnaire. Participants completed an online survey with seven sections as summarized below. The order in which participants completed the sections differed between the control and intervention groups.

1. Welcome. This explained why the study was being undertaken, stressed that participation was voluntary and that participants could withdraw at any stage, explained what partaking in the survey involved, confirmed that all information provided would be kept confidential, stated who was funding the survey, and provided an informed consent form.

2. Your health. As is standard practice in health valuation studies, participants described their own health using the EQ-5D-3L's generic health state descriptive system and assessed their general health using its EQ VAS rating scale.

3. Description of health conditions and how to value them. Participants were provided with a description of each health condition along with an explanation of how to complete the valuation question. In line with established practice, participants were instructed to use the EQ VAS scale to rate the health state, from 0 for worst health imaginable to 100 for best health imaginable. This was followed by an instruction on where to enter their answer.

The description of each health condition consisted of (a) a summary description and (b) a detailed functional description (provided in box format, as is standard practice in health valuation studies). The summary descriptions are presented below.

Arthritis is an illness that leads to inflammation in the joints; a joint is where two bones meet, such as your elbow or knee. This inflammation can lead to stiffness, swelling, redness, tenderness, and pain.

A colostomy is an operation involving the surgical redirection of your bowels through a hole created in your gut, called a stoma.

Tinnitus is the perception of noise or ringing in the ears.

The detailed description of each health condition consisted of five sentences: Each sentence described a salient feature of the condition. The descriptions of arthritis and colostomy were taken from published health valuation studies in which people from the general population were provided with the descriptions of the health conditions and asked to value the conditions

(McTaggart-Cowan, Tsuchiya, O'Cathain, & Brazier, 2011; Smith, Sherriff, Damschroder, Loewenstein, & Ubel, 2006). The description of tinnitus used in the one study found in which people from the general population rated tinnitus (Happich & von Lengerke, 2005) was not used: The survey in that study was undertaken in German and an English translation of the description of tinnitus was not available. Therefore, a description of tinnitus from the website of the nonprofit Mayo Clinic, which is regarded as a leading hospital in the US, was used (Mayo Foundation for Medical Education and Research, n.d.).

4. Valuation of health conditions. Participants rated the three conditions in random order.
 Each condition was rated on a 0-100 scale from worst to best health imaginable using the EQ
 VAS.

5. *Ease of completion.* Participants were asked how easy or difficult they found the specific task of rating the health conditions, and how easy or difficult they experienced the survey to be overall.

6. Presentation of patients' mean rating of a condition. In this section participants (a) viewed information on patients' mean ratings on the 0-100 scale; and (b) were asked whether the mean rating of people with each health condition was higher, about the same, or lower than expected. A respondent could in principle have provided different answers for different health conditions (e.g., indicated that they expected patients would have rated one condition higher, but indicated that they expected patients would have rated another condition lower). Patients' mean assessments of the three health conditions were provided in text and in a graphic. The text read: "When asked to choose a number between 0 and 100 that best represents their current health the average answer was: 63 for people with a colostomy, 60 for people with arthritis, 54 for people with tinnitus" (the last three items were presented as bullet points; the rating and the health condition text was bolded). The graphic showed an image of the EQ VAS 0-100 scale with patients' mean rating of each condition superimposed to the right of the scale. Information on

patients' appraisal of a condition was also available to the intervention group when valuing a condition. The patient sample size in each survey was 184 for arthritis (McTaggart-Cowan et al., 2011), 95 for colostomy (Smith et al., 2006) and 208 for tinnitus (Happich & von Lengerke, 2005).

7. *Personal characteristics.* Participants answered questions on their socio-economic status, namely: gender, age, highest educational attainment, and principal economic status.

Experimental Manipulation

Both groups saw the same questionnaire. The control group saw sections 1 to 7 of the questionnaire in order. That is, the process for the control group followed standard practice for health valuation studies; the only information provided at the point of judging each health condition was the description of the condition, and all information regarding how patients rated their own conditions was only presented after the participants had completed the valuation task so that it could not have affected their responses. The intervention group judged the health conditions after being informed of patients' ratings. That is, for the intervention group Section 6: Presentation of patients' ratings came before Section 4: Valuation of health conditions and Section 5: Ease of completion. Participants were randomly assigned to the control group or the intervention group by a computerized random number generator.

Statistical Analysis

Analysis was conducted with a 2 (between: Group [control, intervention]) \times 3 (within: Health Condition [arthritis, colostomy, and tinnitus]) \times 3 (between: Prior Belief of Patients' Mean Rating ([imagined to be worse than is experienced, imagined to be similar to as is experienced, imagined to be better than is experienced]) mixed analysis of variance (ANOVA). The study prediction was that participants with inaccurate prior beliefs of patients' mean rating of a given condition would adjust their own valuation of a given condition in the direction of the mean rating of those living with a given condition. Therefore, the test of the study hypothesis was whether there was a significant interaction between prior belief of patients' mean rating and group. No other result of the ANOVA has importance for the test of the hypothesis.

Results

Descriptive Analyses

Participant characteristics. See Table 1 for participant characteristics.

[Insert Table 1 around here]

Ease of completion. Respondents found the survey "Very easy" (29.03%), "Easy" (48.39%), or "Neither easy nor difficult" (16.77%); only 5.81% found it "Difficult" and none indicated that it was "Very difficult". Participants were also asked whether they found the specific judgment task easy (rating a health condition on a continuum from worst to best health imaginable). Only 2.58% "Disagreed" and 1.29% "Completely disagreed" that the rating task was easy, while 45.81% "Completely agreed", 40.00% "Agreed" and 10.32% "Neither agreed nor disagreed". These results suggest that the difficulty of the task had not posed a problem for the study.

Prior belief of patients' mean rating. Participants who indicated that patients' mean rating was a "higher rating than expected" for a given health condition were classified as having "imagined the condition to be worse than is experienced". Participants who indicated that patients' mean rating was "about the same rating as expected" were classified as having "imagined the condition to be similar to as is experienced". Finally, if the respondent indicated that patients' mean rating was a "lower rating than expected" they were classified as having "imagined the condition to be better than is experienced". Response bias could have affected answers to this question because the intervention group answered the question about prior beliefs after being informed but prior to rating the health conditions, whereas the control group answered the question about prior beliefs after rating the conditions. For instance, a form of social desirability bias could have led participants in the control group to be more likely to indicate that

VALUING THE Q in QALYs

patients' mean rating of a condition (which was provided to them after they rated the condition) was broadly similar to as they thought it would be. Nevertheless, the answers to this question by the intervention group was not different to the answers by the control group, $\chi^2(2, N = 465) = 0.09$, p = 0.958. There is no evidence of a response bias.

There was, as expected, variation in participants' prior beliefs about how patients value health conditions (see Table 2). Participants indicated that their prior beliefs were "broadly similar" to patients' mean rating in less than half of the cases (around four in ten). In a substantial proportion of cases, health conditions were imagined to be worse than is reported by patients (around four in ten cases) or were imagined to be better than is reported by patients (around two in ten cases). This finding is consistent with the observation that using the current approach to obtain valuations from the general population yields values that are largely uninformed by what it is actually like to be in the health states (Brazier et al., 2005).

[Insert Table 2 around here]

Valuations. The mean valuations are illustrated in Figure 2 (the 95% confidence intervals for a group that do not bound the mean of the other group indicate statistically significant differences). For both the intervention and control group, there appeared to be a positive relationship between prior beliefs of patients' mean rating of their health conditions and valuations of the conditions by the participants. This simply shows that the valuations participants gave of a given condition were strongly related to their prior impressions of how people with a given condition would rate it. The pattern in Figure 2 is consistent with the study prediction of a cross-over interaction between prior beliefs and provision of patients' information. Intervention-group participants who underestimated patients' mean ratings gave valuations that were higher than the valuations provided by control-group participants who underestimated patients' mean ratings (effect sizes averaged d = 0.76). For participants with broadly accurate prior beliefs there appeared to be no difference in the valuations given by the intervention group and the control

group. While, intervention-group participants who overestimated patients' mean ratings gave valuations that were lower than the valuations provided by control-group participants who overestimated patients' mean ratings (effect sizes averaged d = 0.38). The next section formally tests for an interaction between prior belief of patients' mean rating and group.

[Insert Figure 2 around here]

Test of Study Prediction

Analysis of the study prediction of a cross-over interaction between participants' valuation of a health condition and their expectations about patients' mean ratings of a condition was undertaken with a 2 (between: Group) \times 3 (within: Health Condition) \times 3 (between: Prior Belief of Patients' Mean Rating) mixed ANOVA (see Table 3). There was a main effect of prior belief about patients' mean rating, F(2, 447) = 46.95, p < .001. This means that participants own valuation of a given health condition was influenced by whether participants believed patients typically rate a given condition as worse, similar or better than the mean rating of that condition by patients. So, both groups' judgments were influenced by prior beliefs of patients' mean rating of the health conditions. There was a main effect for group, F(1, 447) = 4.11, p = .043, with higher mean valuations by the intervention group, 57.00 (SD = 19.01), than the control group, 52.37 (SD = 23.81), d = 0.22. However, this finding should be interpreted with caution given the study prediction (confirmed below) that prior beliefs will interact with group. The positive intervention effect on valuations pooled across the three health conditions occurred because the subgroup of participants who underestimated patients' mean ratings was larger than the subgroup of participants who overestimated patients' mean ratings, and there was a larger intervention effect size for the former subgroup.

Indeed, there was, as predicted, an interaction between prior belief about patients' mean rating and group, F(2, 447) = 10.41, p < .001, $\omega^2 = .16$. This means that the effect of prior expectations about health conditions on valuations of conditions made by participants from the

general population was reduced when the descriptions of conditions were supplemented with patients' experience-based ratings of health conditions. This interaction between prior belief of patients' mean rating and group, F(2, 555) = 9.67, p < .001 remained when the ANOVA was rerun treating responses from participants with logically incoherent judgments (i.e., who rated one health condition as 100 or best health imaginable) as missing data and imputing the VAS values using MICE (rather than excluding responses from such participants). Details are in the Supplementary File. The 3-way Group × Health Condition × Prior Belief of Patients' Mean Rating interaction was not significant, F(8, 447) = 1.00, p = .437. This finding was not consistent with the interaction between group and prior belief of patients' mean rating differing across the three health conditions (i.e., it was not consistent with the interaction being driven by one health condition only). The results lent support to the study hypothesis. Overall, the results showed that for both groups prior beliefs about patients' ratings of the health conditions influenced valuations of the conditions on a worst to best health imaginable continuum, but the effect of prior beliefs was reduced for the intervention group who had seen patients' mean rating of the conditions.

[Insert Table 3 around here]

Discussion

Strengths and Limitations

This study evaluates the effect of making a simple change to the current process through which the information used to estimate the Q in the QALY is obtained. This approach is likely to be broadly acceptable among stakeholders on the basis that more faith can be placed in values obtained from the general population when those values are better informed. The approach used in this paper can be incorporated into existing representative surveys of people from the general population as it is simple for participants to understand, does not add substantial labor or processing costs, and can be applied to a wide range of health conditions and generic health states. This study shows that the provision of patients' mean rating of their condition influences the valuations of people from the general population with inaccurate perceptions of how patients typically rate a condition. The judgments of people with inaccurate perceptions of patients' mean rating of a given health condition change to align more with patients' reported experiences.

This study was designed to provide a proof of concept that it is possible to elicit more informed values of health states by providing people with information on how patients rate those states and it includes several limitations. There is the possibility that the intervention created a focusing effect (Schkade & Kahneman, 1998) or an anchoring effect (Tversky & Kahneman, 1992). In the case of a focusing effect the intervention group may have placed too much importance on patients' reported experience. With an anchoring effect, on the other hand, the intervention group's valuations may have stayed close to patients' mean rating even if that information was not relevant to the judgment task. Notwithstanding these concerns, providing patients' mean ratings seems warranted; doing so is arguably a requirement if people from the general population are to provide informed preferences. The informed preference theory of utility (Harsanyi, 1997) contends that "in welfare economics and in ethics a person's utility function should be defined in terms of his hypothetical informed preferences" (p. 134) and that preferences are informed only when they are "freed from the distorting effects of factual errors" (p. 140). Other research shows that when valuing hypothetical health states people combine their interpretation of the health state description with their own experiences of ill health and their imagination (Karimi et al., 2017b), but what they imagine as the consequences of ill health is not accurate (Karimi et al., 2017a).

Directions for Future Research

It will be important to examine the degree to which effects found in this study apply to a wider range of health conditions, for generic health sates, for choice-based judgment methods and for a random sample of the general population. It would be useful to examine the direct mechanisms that lead valuations of health states obtained from the general population to be

VALUING THE Q in QALYs

influenced by providing patients appraisals of their health states. Future studies adopting this study's approach should be powered to test not only for interactions between prior beliefs and information provision, but also to detect the effect of information provision on the mean valuations of individual health conditions. This study focused on three conditions that patients value higher than people from the general population do, but this study's approach should also be applied to conditions, such as mental health conditions, that patients do not value higher than nonpatients (Revicki, Shakespeare, & Kind, 1996; Rosser & Kind, 1978). It would also be useful to investigate whether there is an asymmetry in the size of the effect of providing patients' experience to people from the general population. Providing information on the experiences of patients may be more likely to reduce cognitive biases (such as loss-framing, failure to anticipate adaptation and focusing illusion) among respondents who misconceive an ill health state as worse (as opposed to better) than is typically experienced by patients.

More discussion between stakeholders on the possible approaches of obtaining QALY weights (i.e., using patient valuations, general population valuations or experience-informed general population valuations) would be helpful. Whether is it appropriate to use patients' judgments that are possibly influenced by adaptation to their conditions to inform resource allocation decisions is a topic that is likely to arise. The answer to this vexing question, in part, depends on whether the major elements of adaptation are perceived as admirable achievements or are perceived as deficiencies. Menzel, Dolan, Richardson, and Olsen (2002) have argued that only when adaptation is comprised primarily of cognitive denial and suppressed recognition of full health should patients' actual valuations not be provided to people from the general population when judging health states. A better understanding of patients' experience of living with different forms of ill health (such as which major elements of adaptation or other factors have the most influence), and a fuller appreciation of the cognitive mechanism that influence judgment of ill health states by patients and by nonpatients is needed. While a number of helpful

frameworks have been proposed to support the study of these issues (Brazier et al., 2005; Menzel et al., 2002; Stiggelbout & De Vogel-Voogt, 2008; Ubel, Loewenstein, & Jepson, 2003) it will be some time before a fuller picture emerges. In the meantime, it would be informative to obtain the views of the general population on the appropriateness of using experience-informed valuations to value the Q in the QALY.

Application

This study illustrates that it is possible to brief people from the general population of patients' experience of ill health in a way that can be incorporated into the current approach of obtaining QALY weights for health conditions. If this study's findings hold for additional health conditions, generic health states, and using choice-based judgments, then the provision of patients' experiences could be incorporated into the current approach to produce experience-informed general population QALY weights to capture the impact of medical treatments.

References

- Bansback, N., Tsuchiya, A., Brazier, J., & Anis, A. (2012). Canadian valuation of EQ-5D health states: Preliminary value set and considerations for future valuation studies. *PLoS One*, 7(2), e31115. doi:10.1371/journal.pone.0031115
- Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., . . . Tsuchyia, A. (2005). Should patients have a greater role in valuing health states? *Applied Health Economics and Health Policy*, *4*, 201-8. doi:10.2165/00148365-200504040-00002
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One*, 8(3), e57410. doi:10.1371/journal.pone.0057410
- De Wit, G. A., Busschbach, J. J. V., & De Charro, F. T. (2000). Sensitivity and perspective in the valuation of health status: Whose values count? *Health Economics*, 9, 109-126. doi:10.1002/(SICI)1099-1050(200003)9:23.0.CO;2-L
- Dolan, P., & Kahneman, D. (2008). Interpretations of utility and their implications for the valuation of health. *Economic Journal*, *118*(525), 215-234.
- Dolan, P., Kavetsos, G., & Tsuchiya, A. (2013). Sick but satisfied: The impact of life and health satisfaction on choice between health scenarios. *Journal of Health Economics*, *32*, 708-714.
- Dolders, M. G. T., Zeegers, M. P. A., Groot, W., & Ament, A. (2006). A meta-analysis demonstrates no significant differences between patient and population preferences. *Journal* of Clinical Epidemiology, 59, 653-664.

- Drummond, M., Brixner, D., Gold, M., Kind, P., McGuire, A., Nord, E., & Consensus Dev Grp. (2009). Toward a consensus on the QALY. *Value in Health, 12*, S31-S35. doi:10.1111/j.1524-4733.2009.00522.x
- Fryback, D. G. (2003). Whose quality of life? Or whose decision? *Quality of Life Research*, 12, 609-610.
- Gold, M. R., Siegel, J. E., Russell, L. B., & Weinstein, M. C. (Eds.). (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.
- Happich, M., & von Lengerke, T. (2005). Valuing the health state 'tinnitus': Differences between patients and the general public. *Hearing Research*, 207(1-2), 50-58.
 doi:10.1016/j.heares.2005.04.002
- Harsanyi, J. (1997). Utilities, preferences, and substantive goods. *Social Choice and Welfare, 14*, 129-145.
- Karimi, M., Brazier, J., & Paisley, S. (2017a). Are preferences over health states informed?*Health and Quality of Life Outcomes*, 15, 105. doi:10.1186/s12955-017-0678-9
- Karimi, M., Brazier, J., & Paisley, S. (2017b). How do individuals value health states? A qualitative investigation. *Social Science & Medicine*, 172, 80-88. doi:10.1016/j.socscimed.2016.11.027
- Kind, P., & Dolan, P. (1995). The effect of past and present illness experience on the valuations of health states. *Medical Care, 33*, AS255-AS263.
- Lazarus, R. S., & Folkman, S. (1984). Stress appraisal and coping. New York: Springer.

- Mayo Foundation for Medical Education and Research. (n.d.). Tinnitus. Retrieved from https://www.mayoclinic.org/diseases-conditions/tinnitus/symptoms-causes/syc-20350156
- McTaggart-Cowan, H., M., Tsuchiya, A., O'Cathain, A., & Brazier, J. (2011). Understanding the effect of disease adaptation information on general population values for hypothetical health states. *Social Science & Medicine*, *72*, 1904-1912. doi:10.1016/j.socscimed.2011.03.036
- Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: A preliminary normative analysis. *Social Science & Medicine, 55*, 2149-2158.
- NICE. (2013). *Guide to the methods of technology appraisal 2013*. (No. PMG9). London, England: NICE. Retrieved from <u>https://www.nice.org.uk/process/pmg9/chapter/introduction</u>
- Peeters, Y., & Stiggelbout, A. M. (2010). Health state valuations of patients and the general public analytically compared: A meta-analytical comparison of patient and population health state utilities. *Value in Health*, *13*, 306-309. doi:10.1111/j.1524-4733.2009.00610.x
- Revicki, D. A., Shakespeare, A., & Kind, P. (1996). Preferences for schizophrenia-related health states: A comparison of patients, caregivers and psychiatrists. *International Clinical Psychopharmacology*, 11, 101-108. doi:10.1097/00004850-199611020-00004
- Rosser, R., & Kind, P. (1978). Scale of valuations of states of illness is there a social consensus? *International Journal of Epidemiology*, 7, 347-358. doi:10.1093/ije/7.4.347
- Rubin, D. B. (1987). Frontmatter. *Multiple imputation for nonresponse in surveys* (pp. i-xxix) John Wiley. doi: <u>http://onlinelibrary.wiley.com/doi/10.1002/9780470316696.</u>

- Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, 9, 340-346. doi:10.1111/1467-9280.00066
- Smith, D. M., Sherriff, R. L., Damschroder, L., Loewenstein, G., & Ubel, P. A. (2006). Misremembering colostomies? Former patients give lower utility ratings than do current patients. *Health Psychology*, 25, 688-695.
- Stiggelbout, A. M., & De Vogel-Voogt, E. (2008). Health state utilities: A framework for studying the gap between the imagined and the real. *Value in Health*, *11*, 76-87.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect-theory cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297-323. doi:10.1007/BF00122574
- Ubel, P. A., Loewenstein, G., Hershey, J., Baron, J., Mohr, T., Asch, D. A., & Jepson, C. (2001).Do nonpatients underestimate the quality of life associated with chronic health conditions because of a focusing illusion? *Medical Decision Making*, 21, 190-199.
- Ubel, P. A., Loewenstein, G., & Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research*, 12, 599-607. doi:10.1023/A:1025119931010
- White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations:Issues and guidance for practice. *Statistics in Medicine*, *30*, 377-399. doi:10.1002/sim.4067
- Williams, A. (1990). EuroQol a new facility for the measurement of health-related quality-oflife. *Health Policy*, 16, 199-208.



Figure 1. CONSORT Participant Flow Diagram.

Table 1

Participant Characteristics

Variable	Intervention	Control
Age		
18-19	3%	3%
20-29	33%	37%
30-39	32%	24%
40-49	8%	18%
50-59	14%	16%
60-69	8%	1%
Over 70	2%	0%
Highest educational attainment		
Not a high school graduate	2%	1%
High school graduate	13%	9%
Some college, but no degree	33%	27%
Associate degree	11%	19%
Bachelor's degree	30%	36%
Advanced degree or more	11%	7%
Principal economic status		
Employed or self-employed	64%	63%
Retired	5%	1%
Taking care of home	9%	13%
Student	13%	7%
Unemployed	3%	9%
Unable to work due to long-term sickness/disability	7%	6%
Total	88	67

Table 2

	Prior belief of patients' mean rating, Number (%)					
Group	Imagined to be worse than is experienced	Imagined to be similar to as is experienced	Imagined to be better than is experienced			
Intervention	96 (36.36)	108 (40.91)	60 (22.73)			
Control	72 (35.82)	81 (40.3)	48 (23.88)			

Prior Belief by Group of Patients' Mean Rating of Health Conditions



Figure 2. Mean valuation (EQ VAS) of conditions by prior belief of patients' mean rating and group. Error bars represent the 95% confidence intervals.

Table 3

Results of Mixed ANOVA

Partial				
Sum of		Mean		
Squares	df	Square	F	р
33777.71	2	16888.86	46.95	0.000
1477.82	1	1477.82	4.11	0.043
479.17	2	239.58	0.67	0.514
101.33	2	50.67	0.14	0.869
7492.56	2	3746.28	10.41	0.000
2871.57	8	358.95	1.00	0.437
160805.90	447	359.74		
210827	464	454.37		
	Partial Sum of Squares 33777.71 1477.82 479.17 101.33 7492.56 2871.57 160805.90 210827	Partial Sum of Squares df 33777.71 2 1477.82 1 479.17 2 101.33 2 7492.56 2 2871.57 8 160805.90 447 210827 464	Partial Mean Sum of Mean Squares df Square 33777.71 2 16888.86 1477.82 1 1477.82 479.17 2 239.58 101.33 2 50.67 7492.56 2 3746.28 2871.57 8 358.95 160805.90 447 359.74 210827 464 454.37	Partial Mean Sum of Mean Squares df Square F 33777.71 2 16888.86 46.95 1477.82 1 1477.82 4.11 479.17 2 239.58 0.67 101.33 2 50.67 0.14 7492.56 2 3746.28 10.41 2871.57 8 358.95 1.00 160805.90 447 359.74 1.00 210827 464 454.37 1.00

Author Note

Robert P. Murphy, Stirling Management School, University of Stirling, and Department of Health, Dublin, Ireland; Christopher J. Boyce, Stirling Management School, University of Stirling; Paul Dolan, Department of Psychological and Behavioural Science, London School of Economics and Political Science; Alex M. Wood, Department of Psychological and Behavioural Science, London School of Economics and Political Science.

This study is part of the doctoral dissertation of Robert P. Murphy, his doctoral dissertation received funding from and was undertaken when an employee of the Department of Health, Dublin, Ireland. The Department of Health did not have any editorial role in the production of the research. This research was supported in part by a grant (55116) from the Scottish Institute for Research in Economics. The authors would like to thank Leonhard Lades and David Comerford for comments on an early draft of this paper.

Correspondence concerning this article should be addressed to Robert Murphy, Department of Health, Block 1, Miesian Plaza, 50-58 Lower Baggot Street, Dublin, D02 XW14, Ireland. E-mail: r.p.murphy@stir.ac.uk