

Quality-adjusted life year weights and treatment bias: Theory and evidence from cognitive interviews

SAGE Open Medicine

Volume 7: 1–8

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/2050312119856986

journals.sagepub.com/home/smo



Bryan N Patenaude¹  and Till Bärnighausen^{2,3,4}

Abstract

Objectives: The purpose of this research is to understand the thought processes that underpin responses to stated preference approaches for eliciting quality of life, in particular the standard gamble.

Methods: We utilize standard gamble preference elicitation survey techniques to elicit quality-adjusted life year weights for two reduced health states: chronic severe depression and total blindness. After the survey, we conduct open-ended qualitative interviews with respondents to determine their thought processes while taking the surveys and to shed light on what their quality-adjusted life year weight is capturing. Survey responses were coded and analyzed for themes in NVivo, the results of which were then formalized in the terminology of decision sciences.

Results: The qualitative results of the cognitive interviews present systematic evidence for a type of cognitive bias present in standard gamble quality-adjusted life year weight elicitation, which has not been previously highlighted and which we call treatment bias. We define this treatment bias as the consideration of salient treatment alternatives correlated with a reduced health state, when these alternatives are not explicitly posed in the question. Our formalization of this cognitive behavior demonstrates that treatment bias will always bias the elicited health state utility of treating the illness in question downward.

Conclusion: The treatment bias highlighted in this study has implications for economic evaluation when comparing treatment for illnesses where alternative treatments are widely publicized versus those that are not. For example, comparing the effectiveness of treating depression versus arthritis may be biased against depression if advertisements for anti-depressants are more widely viewed by survey respondents than advertisements for arthritis treatments. We propose a statement to be imbedded in all questionnaires regarding stated preference elicitation of quality-adjusted life year weights in order to correct for this bias in future stated preference surveys.

Keywords

Pharmacoeconomics/health economics, mental health/psychiatry, epidemiology/public health, decision science, heuristics and biases

Date received: 24 January 2019; accepted: 23 May 2019

Introduction

The quality-adjusted life year (QALY) is one of the most widely utilized metrics in health economics and health policy prioritization. The QALY is used as a standardized measure of effectiveness in cost-effectiveness analysis or cost-utility analysis and often forms the denominator of an incremental cost-effectiveness ratio (ICER). ICERs are used to compare or rank different public health or medical interventions based on how much money is spent per QALY gained. These data are utilized by healthcare policy-makers in order to prioritize medical intervention areas and to channel limited resources

¹Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

²Heidelberg Institute of Public Health, University of Heidelberg, Heidelberg, Germany

³Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁴Africa Health Research Institute, Somkhele and Durban, South Africa

Corresponding author:

Bryan N Patenaude, Department of International Health, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA.

Email: bpatena1@jhu.edu



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons

Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

into interventions with the largest health gains per dollar spent. As such, the accurate measurement of a QALY is of prime importance for accurate decision-making.

Mechanically, a QALY accounts for both quantity and quality of life resulting from a reduced health state compared with a year in full health. To compute the QALY measure, the length of life lived in a certain health state is multiplied by the quality of life induced by this state, as a fraction of full health.¹ The quantity component of a QALY is straightforward to assess and is measured as the length of life spent in a specific health state. The quality of life component, or QALY weight, is comparatively difficult to measure. It is typically obtained through preference elicitation surveys, which require the respondents to provide information about their personal evaluations of hypothetical health states. One of the most common methods for assessing QALY weights is through stated preference surveys utilizing the standard gamble approach.²⁻⁴

In this article, we discuss and provide evidence for a particular type of cognitive bias that may arise in the standard gamble context known as treatment bias. Treatment bias is when an individual considers salient treatment alternatives when responding to a standard gamble question, despite only one treatment option being presented. This type of bias is often correlated with specific reduced health states. For example, if an individual has seen multiple advertisements for anti-depressants during their lifetime but no advertisements for treating a condition such as hypothyroidism they may subconsciously believe that there are more treatment options available for depression or that depression is much more easily treatable than hypothyroidism, regardless of the reality. This is due to the treatment alternatives for one condition being more salient to the individual than for other conditions, which may cause a bias when QALY weights from two different conditions with varying treatment salience are compared. The remainder of this article will describe the importance of treatment bias in terms of QALY elicitation, provide evidence of its existence from cognitive interviews, formalize the bias in economic terms, and propose one potential solution to this problem.

Qualitative methodology

The first author (B.N.P.) administered QALY elicitation surveys and conducted cognitive interviews with 20 graduate students recruited from the Harvard T.H. Chan School of Public Health. The sample size was chosen to meet adequate saturation and to fall within acceptable bounds for rigorous qualitative research, as described in the Cognitive Interviewing Reporting Framework.⁵ We purposefully selected students who were enrolled in coursework pertaining to global health, which broadly covered topics of global burden of disease, health program evaluation, cost-effectiveness analysis, and both direct and external impacts of health interventions. All students had completed at least one semester of coursework

prior to participation. The rationale for the purposeful selection of highly educated participants with specific training relevant to our specific research question was theoretically motivated so that the recruited participants would, on average, have a better understanding of the immediate and external consequences and benefits of common health states on both an individual and population level than most other populations. As such, major omissions of relevant consequences of reduced health states in this population will likely imply similar or more extensive omissions in the general population. Furthermore, the respondents had also completed required coursework in quantitative methods and statistics, which should have heightened their ability to interpret and understand probabilities, thus minimizing cognitive biases in our results.

We obtained exempt Institutional Review Board (IRB) status through the Harvard T.H. Chan School of Public Health IRB, given the anonymous and hypothetical nature of the study. As per requirements of the exempt status, participants were given a written description of what the study entailed, risks, benefits, as well as contact information of the principal investigator and told that they could opt out at any time. No individuals who were recruited for the study chose to opt out. Respondents in this study were in the age range from 22 to 39 years, with an average age of 27 years. The recruitment was conducted in two stages. Initially, 15 respondents were recruited, only 2 of whom were male. After these results were obtained there was some evidence that results may vary by gender and so five more males were purposefully recruited, bringing the total sample to 20. Of the 20 respondents, 6 were married or engaged and 2 had or were expecting children. Average annual income in the sample was between \$30,000 and \$39,000 with respondents ranging from the less than \$5000 level to the \$75,000–\$99,000 level.

Survey

The survey questionnaire was divided into four sections. The first section included nine brief baseline demographic questions including age, sex, marital status, parental status, number of children, degree program, previous degrees, current health status, and income level. The second section consists of two questions, a time trade-off between healthy life years and years spent with total blindness in both eyes followed by a standard gamble with blindness as the reduced health state (Figure 1). The third section consists of two questions, a time trade-off between healthy life years and years spent in a state of chronic severe depression, followed by a standard gamble with chronic severe depression as the reduced health state. The final section consists of a series of six payment card questions to elicit willingness-to-pay. Overall qualitative findings apart from treatment bias and specific results from the willingness-to-pay portion are discussed in detail in another forthcoming study. Each of the time trade-off, standard gamble, and

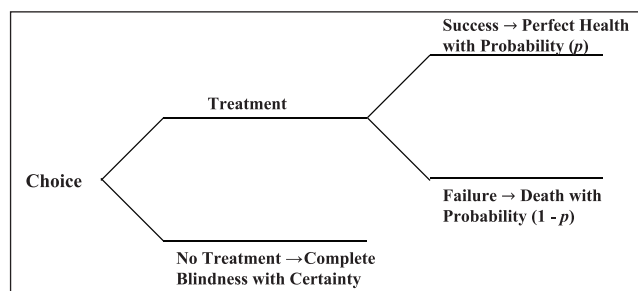


Figure 1. Standard gamble format.

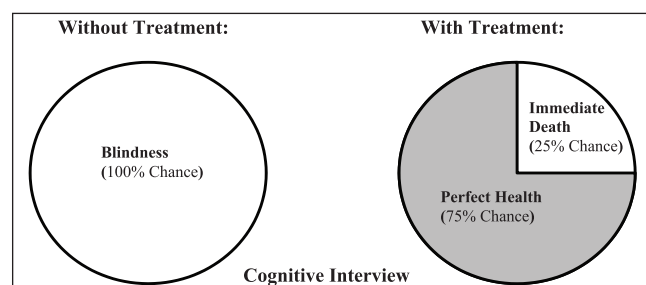


Figure 2. Standard gamble probability pie charts.

payment card willingness-to-pay questions was worded in a validated and standard format for the appropriate methodology and the standard gamble approach adopted is a theoretical gold standard for measuring QALY weights.^{2,6–8} The reduced health states of blindness and chronic severe depression were chosen based on the presumed ease of envisioning the reduced health states compared to many other potential scenarios and a desire to drive respondents to consider both physical and psychological factors in their valuation of reduced health states. In this questionnaire, the time trade-off questions serve as mental warm-up exercises to get respondents thinking about trade-offs and the utility value of life in the described health states, before answering the standard gamble questions related to the same health state. Similar warm-up tactics are often practiced in QALY elicitation surveys and have been demonstrated to improve the accuracy of acquired results.⁸ Pie graphs corresponding to the risk of death if treatment is selected were included in the standard gamble questions (Figure 2).

Immediately following the survey questionnaires, a cognitive interview was conducted in order to capture the participant's thought process when responding to the standard gamble survey questions. The framework of the cognitive interviews was organized prior to the interviews and structured around eight main questions with room for supporting questions to encourage more in-depth discussion or clarification. All interviews began with a series of open-ended questions asking the participants to recount what factors entered their mind when responding to the various types of elicitation questions and then to describe their decision-making process. This question was repeated four times in reference

to the standard gamble questions to elicit QALY weights for both blindness and for chronic severe depression. Participants were permitted to speak freely for as long as they desired. Clarifying questions were posed if responses were unclear to the interviewer. After the participants stopped responding to the open-ended questions, two additional probing questions were asked about whether the participants took into account specific aspects of a given health state, if it was not addressed in their open-ended portion, and if so how they did so.

If participants did not take into account a given aspect of the health state in question, they were further probed as to why, whether they perceived the excluded health state aspect to be of utility value to them, and whether they would think about this aspect if repeating the hypothetical exercise or in a real-life decision-making scenario. The remaining two questions sought to capture the respondent's ability to comprehend the scenarios of blindness and chronic severe depression and whether or not they faced any difficulty in answering the questions in terms of visualizing the probabilities or risks described, which could lead to cognitive bias. The interviews were concluded with any remaining open-ended comments the respondents had about the process.

Qualitative results

All cognitive interviews were recorded and assigned a randomly generated numeric code for identification. The interviews were then uploaded, transcribed, coded, and analyzed by the authors in Qualitative Survey Research (QSR) International's *NVivo* qualitative data analysis software.⁹ Mean survey time was 14 min and 58 s. Instead of delving into the breadth of considerations that came to participants' minds when responding to a QALY weight elicitation survey, this article focuses on the theme of cognitive bias and sub-theme of treatment bias, which we felt warranted its own separate discussion due to the implications it has for systematic QALY mismeasurement.

Comparing the results from blindness-related QALY questions to those related to chronic severe depression sheds light on the vast range of perceptions of severity. Generally, for blindness, most respondents considered it to be a devastating health state with a large post-onset learning curve required to rebuild, recover, and come to psychological terms with the state. For chronic severe depression, however, about half of all respondents considered it significantly less severe than blindness and as many as a third considered it trivial and believed the condition to be easily manageable given treatment options beyond the scope of the questionnaire. One respondent proclaimed:

It probably is a reflection of the world we live in or my opinion of it where I'm just like: 'whatever, medications will solve it all'. It [chronic severe depression] just didn't have the same negative consequence [compared to blindness].

A second respondent states, when asked further about the inclusion of treatment into their consideration that:

I know factually that chronic depression would have a large psychological impact, but I probably didn't give enough weight to that.

A third respondent shared similar views, stating:

I think I just basically didn't view chronic depression as badly as I should and it could be because, maybe, because I've never been close to anyone who has been depressed ... so in my head I'm just like: 'oh well, you'll take a pill and you'll be fine'.

This last statement is of particular interest and was expressed either directly or indirectly by about half of all participants. This statement is evidence that despite the question's explicit wording that you would be in a reduced health state (either blindness or chronic severe depression) with no treatment options apart from the proposed risky treatment, individuals still focused on available and salient treatment options from their external experience when making their decision. As such, it is clear that treatment bias seems to be somewhat persistent and may have large impacts on both elicited QALYs and cost-effectiveness evaluation using those QALYs.

A model of treatment bias

We take a grounded theory approach to formalize the treatment bias sub-theme uncovered in the cognitive interviews and to incorporate this phenomenon into existing decision-science and economic models. In general terms, this bias is the result of considering the reduced health state described in a standard gamble in the context of salient (known and readily memorable) treatment options, rather than the explicit terms of the question posed. This example is closely connected to the psychological and behavioral economics literature on heuristics and biases.^{10,11} As a result of differential salience, utility estimates for diseases with more salient treatment options may be undervalued in a utility loss sense compared with diseases with less salient treatment options. This occurs regardless of the treatment options that actually exist for each health state. As such, treatment bias occurring in questions attempting to elicit health state utility is a problem for both preference estimation and for corresponding cost-effectiveness analysis.

To begin isolating the treatment bias from the health state utility, we write out the general QALY model

$$U(Q, T) = H(Q)F(T)$$

where Q is a specific reduced health state, $H(Q)$ is the health-related quality of life attached to state Q , T is time spent in health state Q , and $F(T)$ is utility over duration of life. It is

commonly assumed that utility over duration of life is linear and represented by $F(T) = T$, implying risk neutrality with respect to life years.^{12,13} Thus

$$U(Q, T) = H(Q)T$$

Under the standard gamble, we estimate $U(Q, T)$ through the indifference point between living in health state Q for a duration of T years with certainty and taking a treatment for health state $\underline{Q}$, which has the probability p of restoring perfect health for T years and the probability $(1-p)$ of immediate death.¹⁴ Under the standard gamble setup, $\bar{Q}$ represents the state of perfect health, $\underline{Q}$ represents the state of death, and $U(\bar{Q}, T)$ is the utility associated with living in perfect health for T years and $U(\underline{Q}, T)$ is the utility associated with immediate death.¹⁵ Therefore, under this framework, the utility of living in reduced health state Q for time period T is equivalent to

$$U(Q, T) = pU(\bar{Q}, T) + (1-p)U(\underline{Q}, T)$$

This means that

$$H(Q)T = p[H(\bar{Q})T] + (1-p)[H(\underline{Q})T]$$

Standardizing to $T=1$ year

$$H(Q) = p[H(\bar{Q})] + (1-p)[H(\underline{Q})]$$

We assume in the standard gamble context that the health state utility associated with perfect health $H(\bar{Q}) = 1$ and the health state utility associated with death $H(\underline{Q}) = 0$. Thus, the model simplifies to

$$H(Q) = p$$

As a result, the indifference probability elicited from the standard gamble directly gives us our QALY weight, or the health state utility associated with a given reduced health state Q . However, in the context of our observed treatment bias, we are not accurately capturing health state utility with our elicited p resulting in a biased estimate of $H(Q)$. To see this, suppose that health state utility is composed of k different attributes representing utility-relevant qualities of that reduced health state. Examples of such attributes are listed in Figure 3.

Suppose $H(Q)$ can be decomposed into utility derived from these separate attributes as such

$$H(Q) = \sum_{k=1}^K h_k(Q)$$

1. Personal pain or suffering (physical)
2. Personal pain or suffering (psychological/emotional)
3. Difficulty of living daily life
4. Inability to complete future life goals
5. Loss of income or reduction in economic productivity
6. Hardships faced by family, friends, and loved ones as a result of the respondents worsened health state (economic) that impact individual utility
7. Hardships faced by family, friends, and loved ones as a result of the respondents worsened health state (physical and psychological) that impact individual utility.
8. Impact on community or those acquaintances, co-workers, and individuals a respondent comes in contact with on a day-to-day basis that impacts individual utility.

Figure 3. Health state aspects relevant to utility valuation. Aspects derived from Meltzer,¹⁶ World Health Organization,¹⁷ and Weinstein et al.²

where $h_k(Q)$ is the health state utility derived from the k th attribute associated with health state utility over health state Q . Now, assume that individuals consider, in addition to the K general attributes associated with health state Q , the availability of treatments outside of the treatment scenario proposed within the standard gamble question. Suppose that a given health state Q has n potential treatments. This framework draws from the literature on modeling salience, limited attention, and focusing in standard utility models.^{18,19} Thus, our estimated $\widehat{H}(Q)$ from the standard gamble can be represented as

$$\widehat{H}(Q) = \sum_{k=1}^K h_k(Q) + \sum_{n=1}^N g_n h_n(Q)$$

where g_n is a set of binary weights corresponding to treatment option n of health state Q . If the treatment option was salient (came to mind) then $g_n = 1$, and if it was not salient (was forgotten or unknown) then $g_n = 0$. Suppose that of the n potential treatments J treatment options are salient to the respondent and L elements are not salient such that $N = J + L$. Then, the equation can be rewritten as

$$\widehat{H}(Q) = \sum_{k=1}^K h_k(Q) + \sum_{j=1}^J g_j h_j(Q) + \sum_{l=1}^L g_l h_l(Q)$$

where $g_j = 1$ when the treatment attribute is salient and $g_l = 0$ if the attribute is not salient. Since the treatment attributes not considered have a weight of 0, this equation reduces to the following

$$\widehat{H}(Q) = \sum_{k=1}^K h_k(Q) + \sum_{j=1}^J h_j(Q)$$

Therefore, letting $\hat{p}$ be our estimated probability from the standard gamble. What $\hat{p}$ is actually representing is

$$\hat{p} = \widehat{H}(Q) = \sum_{k=1}^K h_k(Q) + \sum_{j=1}^J h_j(Q)$$

However, what we want to measure is

$$p = H(Q) = \sum_{k=1}^K h_k(Q)$$

Therefore, the treatment bias associated with this standard gamble is

$$\begin{aligned} p - \hat{p} &= H(Q) - \widehat{H}(Q) \\ &= \sum_{k=1}^K h_k(Q) - \left(\sum_{k=1}^K h_k(Q) + \sum_{j=1}^J h_j(Q) \right) \\ &= - \sum_{j=1}^J h_j(Q) \end{aligned}$$

Thus, depending on both the total number of considered alternative treatment attributes and the health-related utility associated with those salient treatment attributes, the treatment bias may be substantial. Formalizing the treatment bias demonstrates that for conditions where treatment options are more salient to consumers than others, QALY measurements will contain a negative treatment bias. The result is that this bias causes us to systematically undervalue, in utility terms, QALYs gained from treating those conditions that have more salient treatment alternatives relative to those with non-salient alternatives, regardless of the reality or scope of existing treatments. Such a bias is particularly problematic in comparing mental health conditions such as depression, where advertising presents numerous treatment options, to physical conditions where treatment options may be less salient unless the individual has encountered the situation herself.

Discussion

Utilizing a grounded theory approach and the qualitative results presented, there is evidence that treatment bias may be strong in QALY elicitation via standard gamble discrete choice experiments. This is particularly problematic for QALY weight measurement as the estimated QALY metric is no longer empirically valid for the reduced health state itself, but rather for the health state given a particular sociomedical setting. Such a result makes QALYs from different individuals in

different settings incomparable. Since QALYs were designed to be a standardized way to compare health gains across individuals and interventions, these findings are problematic for their validity. In addition to the health state utility being mis-measured, cost-effectiveness analysis utilizing QALYs elicited with treatment bias will result in the systematic undervaluation of the effectiveness of policies targeting diseases with a higher degree of treatment salience compared to those with a lower degree of treatment salience. This is true even if the treatment salience does not accurately represent the true and existing treatment landscape.

We provide a vignette, below, to illustrate the potential impact of treatment bias on policy prioritization using cost-effectiveness analysis. Suppose you have a fixed budget of \$1000 and that you are comparing the cost-effectiveness of treatment to the status quo. Therefore, set baseline cost (C_0) and baseline effectiveness (E_0) equal to 0. Now assume your health system is considering investing in a new treatment and has two options, treatment 1 and treatment 2, that each cost \$1000 (i.e. $C_1 = C_2 = \$1000$). Treatment 1 cures rheumatoid arthritis and treatment 2 cures depression. Assume that the two conditions are non-fatal and persist for 20 years. Also, assume that the QALY weight $H_1(Q)$ for rheumatoid arthritis is 0.7 and the QALY weight $H_2(Q)$ for clinical depression is 0.65. Therefore, the QALYs associated with each condition are

$$\begin{aligned} QALY_1 &= \text{Rheumatoid Arthritis QALYs} \\ &= H_1(Q)T_1 = 0.7(20) = 14 \text{ QALYs} \\ QALY_2 &= \text{Clinical Depression QALYs} \\ &= H_2(Q)T_2 = 0.65(20) = 13 \text{ QALYs} \end{aligned}$$

Since both treatments are 100% effective and the effectiveness of each treatment (E_1 and E_2) is measured by QALYs gained, we can measure treatment effectiveness by computing the difference between the QALYs associated with perfect health and the QALYs associated with each condition. Therefore

$$\begin{aligned} E_1 &= H(Q_{\text{Perfect}})T_1 - H_1(Q)T_1 = 20 - 14 = 6 \text{ QALYs} \\ E_2 &= H(Q_{\text{Perfect}})T_2 - H_2(Q)T_2 = 20 - 13 = 7 \text{ QALYs} \end{aligned}$$

As such, comparing treatment 1 to baseline and treatment 2 to baseline using ICERs we get

$$\begin{aligned} ICER_1 &= \frac{C_1 - C_0}{E_1 - E_0} = \frac{\$1000 - \$0}{6 - 0} \approx \$166.67 \text{ per QALY} \\ ICER_2 &= \frac{C_2 - C_0}{E_2 - E_0} = \frac{\$1000 - \$0}{7 - 0} \approx \$142.85 \text{ per QALY} \end{aligned}$$

Thus, $ICER_2 < ICER_1$, and assuming the value of a statistical life year (VSLY) is above both values, then the decision-maker should invest in treatment 2, treating

clinical depression because it has a lower cost per QALY gained.

Now suppose that we can observe the years of life lived in the reduced health state, which is still 20 years, but instead of knowing the true QALY weights for treatments 1 and 2 ($H_1(Q)$ and $H_2(Q)$), we acquire them through surveying individuals through some type of preference elicitation technique, such as the standard gamble. Suppose there is no salient treatment bias for rheumatoid arthritis but there is salient treatment bias for clinical depression. This is due to the prevalence of advertisements for anti-depressants, whereas there is not such a salient understanding of readily available treatments for rheumatoid arthritis. Assume that this knowledge increases elicited QALY weights by 10 percentage points. As such, we estimate

$$\begin{aligned} \widehat{QALY}_1 &= \text{Rheumatoid Arthritis QALYs} \\ &= H_1(Q)T_1 = 0.7(20) = 14 \text{ QALYs} \\ \widehat{QALY}_2 &= \text{Clinical Depression QALYs} \\ &= H_1(Q)T_1 = \widehat{H_1(Q)}T_1 \\ &= (H_1(Q) + 0.10)T_1 \\ &= (0.65 + 0.10)T_1 \\ &= 0.75(20) \\ &= 15 \text{ QALYs} \end{aligned}$$

Now when we estimate the corresponding effectiveness measures ($\hat{E}_1$ and $\hat{E}_2$)

$$\begin{aligned} \hat{E}_1 &= H(Q_{\text{Perfect}})T_1 - \widehat{H_1(Q)}T_1 = 20 - 14 = 6 \text{ QALYs} \\ \hat{E}_2 &= H(Q_{\text{Perfect}})T_2 - \widehat{H_2(Q)}T_2 = 20 - 15 = 5 \text{ QALYs} \end{aligned}$$

Thus, comparing treatment 1 to baseline and treatment 2 to baseline using the estimated ICER we get

$$\begin{aligned} \widehat{ICER}_1 &= \frac{C_1 - C_0}{\hat{E}_1 - E_0} = \frac{\$1000 - \$0}{6 - 0} \approx \$166.67 \text{ per QALY} \\ \widehat{ICER}_2 &= \frac{C_2 - C_0}{\hat{E}_2 - E_0} = \frac{\$1000 - \$0}{5 - 0} = \$200.00 \text{ per QALY} \end{aligned}$$

Thus, $ICER_2 > ICER_1$, and assuming the VSLY is above both values, then the decision-maker should invest in treatment 1, treating rheumatoid arthritis because it has a lower cost per QALY gained. The prioritization decision was completely reversed in this scenario due to only a 10-percentage point bias. This demonstrates that even a small degree of salient treatment bias can have meaningful implications for effectiveness measurement and corresponding policy prioritization.

It is important to note two potential limitations of this study and what we have done to control them. First,

cognitive bias may exist in terms of responses to the standard gamble questions. Cognitive bias, or that the respondent cannot visualize or appropriately comprehend the probabilities or values associated with the question, is often associated with standard gambles due to the inability of respondent to visualize and conceptualize the probabilities associated with their choices.² Other forms of bias are also associated with time trade-off questions including time discounting bias, or systematically valuing future health gains different from present health gains, and scale bias, or how the order and distribution of options on a rating scale can influence decision-making. Since each possible approach contains the potential for bias, a decision was made to use the standard gamble approach, which is largely regarded to be the most theoretically sound approach and one whose bias may be most easily overcome with properly designed questions.²⁻⁴ In addition, to examine the extent of potential cognitive bias within this approach, respondents were asked during the cognitive interviews if they faced difficulties envisioning any of the scenarios, probabilities, question types, or monetary values. This was done in order to assess the possible extent to which cognitive bias came into play. Results show that respondents actually found the standard gamble questions easy to visualize and to answer, especially when compared to other types of questions such as willingness-to-pay and time-tradeoff type questions.

A second potential limitation is that interviewer bias may have come into play if the probing questions led interviewees to respond in a way that is different to their actual beliefs in order to conform to some norm or standard expressed by the interviewer. While this is certainly a possibility in any interview scenario, our format was designed to begin with an open-ended response and then dig deeper into the foundations of that response. As such, respondents were given time and space to provide detailed responses about why they felt a certain way or had the realizations that they did, which made it much more difficult to artificially construct a response than would be the case from a simple yes or no question. In addition, many respondents were able to give examples of how certain aspects would and would not factor into their thought process, making the obtained results more believable. For instance, one individual directly stated that they forgot to take into account future employment prospects, but that this would not matter so much to their utility, while other aspects that they left out concerning family would matter:

I don't think that future income would be [important] because I think both of these issues [blindness and chronic severe depression] would be surmountable. Like I could overcome these issues enough to not have major offset in my future potential earnings. But I would definitely think more about the impact on other.

Several respondents were able to define aspects in a similar manner, and such a response provides evidence that the interviewer bias was likely minimal.

One potential solution to the issue of treatment bias may be to repeat and reinforce the phrase "in the absence of external treatment" both within the introductory portion of the standard gamble question and alongside the second portion of the question, immediately before the respondent is making her decision. Such an additional clarifying emphasis could improve the consistency and comparability of elicited QALYs and allow them to satisfy the assumption that "a QALY is a QALY is a QALY."² In addition to reducing the potential for treatment bias to occur, this change in phrasing could also help to reduce the likelihood that incorrect prioritization decisions are made when policy-makers are comparing the cost-effectiveness of interventions using QALYs as the effectiveness measure.

Conclusion

This study provides qualitative evidence that it is difficult for individuals to take into account a reduced health state in the absence of treatment, particularly for conditions where treatment is highly salient. This was especially true for chronic severe depression, when compared to blindness. Specifically, blindness seems to be viewed as permanent and absent of any treatment even though a risky treatment was presented. Conversely, chronic severe depression was viewed in the context of alternative external treatments or as something that could easily be resolved even though the health state was clearly defined as permanent and without treatment apart from the risky treatment option posed in the standard gamble question. This means that medical, social, or even geographic setting could impact how the same individual perceives chronic severe depression due to the salience of available treatment options. Thus, the true utility of chronic severe depression is not being evaluated. Instead, the utility of chronic severe depression in the context of salient treatment is being evaluated, which was not the case for blindness and results in the utility measures for each being incomparable. Incorporating repeated reminders into the phrasing of standard gamble questions could help to reduce the likelihood of treatment bias occurring in QALY weight elicitation and improve the usability of the resulting QALYs for economic healthcare decision-making.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval

Ethical approval for this study was granted by the Harvard T.H. Chan School of Public Health IRB as Exempt Human Research because no personally identifiable information was collected and all questions posed to individuals were hypothetical. The protocol number assigned to this approved research was IRB14-3915.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Informed consent

Verbal informed consent was obtained from all subjects before the study as required under Exempt Human Research under the Harvard T.H. Chan School of Public Health IRB. The IRB-approved exempt consent form for this study is attached as a supplementary document.

Supplemental material

Supplemental material for this article is available online.

ORCID iD

Bryan N Patenaude  <https://orcid.org/0000-0002-5180-3626>

References

1. Torrance GW. Measurement of health state utilities for economic appraisal: a review. *J Health Econ* 1986; 5(1): 1–30.
2. Weinstein MC, Torrance G and McGuire A. QALYs: the basics. *Value Health* 2009; 12(Suppl. 1): S5–S9.
3. Gafni A. The standard gamble method: what is being measured and how it is interpreted. *Health Serv Res* 1994; 29(2): 207–224.
4. Smith RD. Use, option and externality values: are contingent valuation studies in health care mis-specified? *Health Econ* 2007; 16(8): 861–869.
5. Boeije H and Willis G. The cognitive interviewing reporting framework (CIRF): towards the harmonization of cognitive testing reports. *Methodology* 2013; 9(3): 87–95.
6. Cropper M, Hammitt JK and Robinson LA. Valuing mortality risk reductions: progress and challenges. *Annu Rev Resour Econ* 2011; 3(1): 313–336.
7. Hammit JK. QALYs versus WTP. *Risk Anal* 2002; 22(5): 985–1001.
8. Adler MD. Contingent valuation studies and health policy. *Health Econ Policy Law* 2010; 5(Pt. 1): 123–131.
9. NVivo qualitative data analysis software, version 10. QSR International Pty Ltd, <https://www.qsrinternational.com/nvivo/support-overview/faqs/how-do-i-cite-qsr-software-in-my-work>
10. Tversky A and Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* 1974; 185(4157): 1124–1131.
11. Kahneman D. *Thinking, fast and slow*. Reprint edition. New York: Farrar, Straus and Giroux, 2013, p. 512.
12. Bleichrodt H, van Rijn J and Johannesson M. Probability weighting and utility curvature in QALY-based decision making. *J Math Psychol* 1999; 43(2): 238–260.
13. Mehrez A and Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making* 1989; 9(2): 142–149.
14. Manski CF. Measuring expectations. *Econometrica* 2004; 72(5): 1329–1376.
15. Abellán-Perpiñán JM, Pinto-Prades JL, Méndez-Martínez I, et al. Towards a better QALY model. *Health Econ* 2006; 15(7): 665–676.
16. Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ* 1997; 16(1): 33–64.
17. World Health Organization (WHO). Macroeconomics and health: investing in health for economic development: report of the commission on macroeconomics and health, <http://whqlibdoc.who.int/publications/2001/924154550x.pdf> (accessed 5 May 2017).
18. Bordalo P, Gennaioli N and Shleifer A. Salience theory of choice under risk. *Quart J Econ* 2012; 127(3): 1243–1285.
19. Köszegi B and Szeidl A. A model of focusing in economic choice. *Quart J Econ* 2013; 128(1): 53–104.