EQ-5D

EuroQol Working Paper Series

Number 17002
October 2017

ORIGINAL RESEARCH

Conceptual issues with Health Valuation and its Underlying Assumptions

Mark Oppe[1]
Benjamin Craig[2]

[1] EuroQol Research Foundation, Rotterdam, the Netherlands
[2] University of South Florida, Tampa, USA

www.euroqol.org

Abstract

The last decade has seen a wealth of conceptual issues emerge in pursuit of methodological innovations regarding health valuation. These issues are sparsely documented, yet form the basis for an array of jargon-filled discussions. Without a proper reference, it is difficult for less experienced researchers to participate in health preference research or for these concepts to mature. In this manuscript we describe 6 of these conceptual issues relating to different aspects of health preference research.

Keywords

Discrete Choice Experiments, Time Trade-Off, QALY, EQ-5D, Health Preference Research

Acknowledgements

We'd like to thank our colleagues Jennifer Jelsma, Kim Rand-Hendrikson, Juan Manuel Ramos Goñi, Axel Mühlbacher and Henry Bailey for the interesting and pleasant conversations that led to this paper.

Mark Oppe
EuroQol Research Foundation
Marten Meesweg 107
3068 AV Rotterdam
E: oppe@euroqol.org

Disclaimer: The views expressed are those of the individual author and do not necessarily reflect the views of the EuroQol Group.

Table of Contents

Introduction

The last decade has seen a wealth of conceptual issues emerge in pursuit of methodological innovations regarding health valuation. These issues are sparsely documented, yet form the basis for an array of jargon-filled discussions. Without a proper reference, it is difficult for less experienced researchers to participate in health preference research or for these concepts to mature. The aim of this manuscript was to describe six of the conceptual issues with health valuation and its underlying assumptions. The six issues originated from discussions that involved one or both of the authors of this manuscript and other scientists at various scientific meetings and conferences, and are presented in no particular order. To make them more accessible to less experienced researchers, we framed the conceptual issues as stories, each broken down into 4 components: Concept, Origin, Implication for health preference research, and the Moral of the story. The six issues attempt to explain each concept using practical examples and laymen terms; however, some understanding of discrete choice experiments (DCE), time trade-off (TTO) and health related quality of life is required. Each section includes recommendations on how one might deal with these issues and/or potential directions for future research focused on trying to solve these issues (i.e., Implementation). A draft manuscript was subsequently circulated to the scientists who also took part in the discussions. Those who commented are also included in the acknowledgement section of the paper.

Rotten apples

Rotten apples is the concept of a person's unwillingness to accept or forgo a specific attribute under any circumstances.

At the 2011 EuroQol group meeting in Oxford, Jennifer Jelsma completed a TTO task in front of the attendees of the meeting. The TTO task attempts to identify a person's indifference point between two health descriptions by iteratively increasing or decreasing the lifespan of one alternative. Her presentation demonstrated common difficulties and reactions that an educated, yet inexperienced participant could have during the task, particularly problems with valuing poor health. In response, Kim Rand-Hendrikson applied a "rotten apple" analogy to describe this difficulty with finding an indifference point for poor health and good health, which is paraphrased here.

Imagine the case of a choice between one good apple and one rotten apple and that you are asked "Which do you prefer, knowing that you must eat what you receive?" Obviously, you would prefer the good apple. In response, the iterative process in a valuation task might then offer you the choice of "What if I gave you one good apple and one rotten apple or just one good apple, which do you prefer?" You would probably prefer the good apple, because you don't want to eat the rotten apple. "Well, what if I gave you TWO good apples and the rotten apple, versus a single good apple?" You might reply that "There's no way you're going to make me eat that rotten apple. I'm not going to choose the rotten apple ever." Basically, this analogy shows that finding an indifference point may be impossible to achieve (even with infinite apples).

In terms of health preference research, this concept of an infinite positive or negative utility is also known as lexicographic preferences, where your preferences between two objects are determined by a decision rule. In this particular case, it is a rule based on a single attribute ("a rotten apple"). In economics, this concept is often applied to explain what is called inelastic demand for basic needs (infinite willingness-to-pay; e.g., insulin among diabetics) or objects with no demand (zero willingness-to-pay; e.g., hair spray among adults who are bald).

When a task applies an iterative process to find an indifference point, lexicographic preferences can be prohibitively detrimental, because the task may lead to infinite positive or negative responses. With just one infinite response, the mean of a response distribution is no longer defined.

For the EQ-5D, in terms of the EuroQol Group's health valuation, negative infinite responses for severe health states for the classic TTO are restricted to a value of -39 in the iterative procedure, followed by an arbitrary transformation which rescales the negative values such that the lower limit is negative one. When the composite TTO is used, negative infinite responses are restricted to negative one in the iterative procedure, followed by treating these negative one responses as censored observations. The Group recognizes that some respondents may never want to experience poor health ever; however, it remains unclear what the most appropriate way is to include these extreme responses in the analyses.

The moral of the story is that lexographic preferences exist. If you are going to apply an adaptative approach to the identification of indifferent points, a researcher should also include a protocol for how to deal with lexicographic preferences.

Chocolates versus cars

Chocolates versus cars refers to the modeling of the relative value of alternatives in a DCE. Specifically, whether the relative value of two alternative should be expressed as an additive difference (i.e., A-B) or as a ratio (i.e., A/B).

At the 2012 ISPOR meeting in Berlin, Benjamin was discussing his problems modelling health preferences with Mark and Juan Manuel Ramos Goñi. He had tried multiple alternative cumulative density functions, but found that the scaling issues in typical additive functions (e.g., logits) were not well suited for health valuation. Instead of an additive function, he proposed a function based on ratios due to the fact that the scaling term cancels in a ratio, which led to the following analogy.

Imagine that you had a group of individuals and asked their preference between a milk chocolate and a dark chocolate. Let's suppose that 75% prefer the milk chocolate over the dark chocolate. Then suppose you asked them about their preference between a sports car and a sports utility vehicle (SUV) and 75% prefer the sports car over the SUV. Does this mean that the difference in value between the 2 chocolates is the same as the difference in value between the 2 cars? The answer is: no, that can't be true. The difference in value between the 2 chocolates is around, say, one dollar (i.e., Milk-Dark=$1), but the difference in value between the 2 cars is probably much higher, say, around 1,000 dollars (i.e., Sport-SUV=$1000). Instead, suppose relative value is expressed as a ratio, might the ratio of milk to dark chocolates equal the ratio of sports car to SUV (i.e., Milk/Dark=Sports car/SUV)? The answer is: yes, that might be true. Maybe if the 2 ratios are equal, the 2 choice probabilities are equal (i.e. 75%)?

For health preference research, this is a fundamental problem for the modelling of relative values if and only if the alternatives are on different scales (i.e. chocolates vs. cars). If all alternatives are on the same scale and similar in value, additive models will produce nearly identical results as ratio-based models (i.e., log(A/B)=log(A)-log(B)). Additive models are easier to estimate, because ratios can misbehave (e.g., A/B approaches infinity as B approaches zero). For the EuroQol Group, their health valuation studies attempt to determine the relative value of mild losses in health-related quality of life (i.e., chocolates) and severe losses in health-related quality of life (i.e., cars). By limiting the study to just one or the other, the results would no longer be fit for purpose. In the recent DCE studies, some mild pairs (e.g., 11121 vs 11211) and some severe pairs (e.g. 55545 vs 55455) were difficult to predict, likely due to scaling issues in the additive models, which can inflate the "slight" parameters and deflate the "extreme" parameters.

The moral of the story is that a health preference study that includes alternatives with a wide range in values may consider using ratios to express relative value instead of additive differences.

Stars within galaxies

Stars within galaxies refers to the inherent trade-off between understanding individual and aggregate values.

At the 2014 ISPOR meeting in Montreal, Mark was comparing the TTO approach in health valuation to the non-adaptive paired comparison approaches. Specifically, he had conducted a cluster analysis of the TTO data to classify respondents according to their responses. To explain this to Axel Mühlbacher, he used an analogy from astronomy. If you're interested in determining the average brightness for a galaxy, you can measure the average brightness of the stars in the galaxy and multiply that by the number stars in the galaxy. Alternatively, you can measure the brightness of each of the 7 main types of stars in the galaxy, and take a weighted average of those based on the number of stars of each type present in the galaxy. Although the first might be less expensive (i.e. you're able to use a cheaper telescope), it does not tell you anything about the composition of the galaxy, which impedes inference.

The relation to health preference is the following: if a pair probability is 50%, the interpretation is unclear: (1) were the responses completely random; (2) was each respondent indifferent between

the two alternatives; (3) did half the people choose A because they want A and half the people choose B because they want B; or (4) did half the people choose A because they did not want B, and the other half choose B because they did not want A. All these interpretations are possible and may affect the study conclusions. By using an adaptive approach (e.g., TTO) and identifying the values of each person (i.e., stars within galaxies), you know the composition of the values, not just the general preference. Furthermore, understanding the composition may facilitate the interpretation of future studies and aid quality control initiatives. For example, you can count the number of respondents in a particular cluster and compare their frequency across samples.

The moral of the story is that if you don't have sufficient budget or time, you can measure general preferences and forget the composition. However, it may be better to design your task in such a way that you can understand the composition of values within the population, not just determine the average.

The spiral staircase

The spiral staircase refers to the fact that near unanimous preferences express relative value up to a censoring threshold.

This analogy was developed based on an angular approach to preference responses. Imagine that you are walking up a spiral staircase with a friend, where the difference in steps represents relative value. If you are on the same step, this difference in relative value is zero. In a DCE, a choice between two states with equal value (i.e., A=B) means that the probability of choosing A is 50% (P=0.5). If your friend is a few steps ahead of you and you can still see her, you can accurately assess how far ahead she is (i.e., $0.5<P<1$). However, if your friend is ahead of you beyond the horizon, you can no longer see her (P=1). You know she is ahead of you, but you do not know how far ahead she is (i.e., the distance is censored). She may be just ahead of your horizon (the censoring threshold) or well beyond, at the top of the stairs.

A comparison of two dumbbells provides a similar analogy to a spiral staircase. In this dumbbell analogy, you can judge the difference in weight if you can lift both dumbbells; however, it is really hard to discern the difference in weight, if you can't lift one or both dumbbells. Again, the relative difference is censored by a threshold (your ability to lift).

Multiple factors may influence the location of the censoring threshold. For the spiral staircase, the curature of the stairs determines the number of steps to the horizon. For the dumbbells, stronger people are better at assessing differences in weights. Like differences between weightlifters, persons walking closer to the center of a spiral staircase will have a shorter horizon than a person walking closer to the edge, hugging the outer wall, due to the opaque center of the spiral. It is often important to recognize in preference research that threshold location may vary between persons (i.e., heterogeneity) due to differential expertise.

Heterogeneity and its control can have other effects. For example, a very strong weightlifter may have difficulty discerning the difference between light dumbbells (scale effect). Likewise, if the staircase is extremely wide, walking close to its edge may make it difficult for you to discern if your

friend is on the same step as you. Efforts to reduce heterogeneity may extend the distance to the censoring threshold and induce other biases.

In health preference research, non-unanimous preferences (0<P<1) express relative value, but unanimous preferences (P=0 or 1) express only the lower bound of the relative value. The differential location of the horizon (i.e., heterogeneity) may be incorporated into an angular model along with the concept of censoring. This issue is of increasing importance for the interpretation of probabilities near 0 and 1.

The moral of the story is not to choose pairs at the extremes, near 0 or 1. If you do, it is also important to incorporate some form of control of censoring. This censoring argument is separate from typical distributional arguments against the inclusion of unanimous pairs.

Blue cat versus red dog

Blue cat versus red dog relates to the choice of stimulus in DCEs, specifically the merits of using hypothetical vignettes or respondent experiences.

Imagine if you showed a respondent a framed painting of a blue cat and a framed painting of a red dog and asked "which do you prefer?" Alternative, imagine if you showed a respondent a framed text with the words "blue cat" and a framed text with the words "red dog" and asked "which do you prefer?" Each choice includes the respondent's preferences, but the latter also depends on their perception, imagination, and comprehension. Seeing the actual paintings may better represent their preferences. Alternatively, you could show a respondent a framed painting of a blue cat and a blank canvas and asked them "which do you prefer?" Or, you could show the respondent a framed painting of a red dog and a blank canvas and asking them again "which do you prefer?" This may also better represent preferences than text alone, because a blank canvas (hypothetical vignette of no painting) is more easily described and imagined.

Likewise, this analogy is relating to the use of hypothetical vignettes or respondent experiences. The example with the text is like a hypothetical description of the paintings. The example with the painting is like experience-based stimuli. If you have a respondent who reports experiencing moderate depression and moderate pain, you could ask "If you could have relief of either for the next 30 days, which do you prefer?" This approach does not require any vignettes, except for the hypothetical relief of each (i.e., blank canvas).

Experience-based preference assessment is commonly used in decision aids to customize interventions from the patient's perspective. In health valuation, it is particularly useful to identify problems that patients experience and value. However, in health preference research, the approach may greatly limit generalizability, because persons who experience health problems may be incapable of expressing preferences when (s)he is experiencing them. Also, the persons who experience the problems may be rare and not representative of the stakeholder population.

The moral of the story is that different forms of presentation exist for health preference research, and that each form of preference evidence may provide useful information. Future research should

attempt to improve our understanding of when they differ and how to merge both forms of preference evidence.

The Trinidad Snake

The Trinidad snake refers to the emotional distress caused when respondents are asked to imagine poor health.

In January 2016, Benjamin and Henry Bailey met in Trinidad to review the comments received from respondents during the timing, duration, and lifespan study. In this study respondents were asked to choose between quality of life and lifespan. Some said that they really disliked the task of thinking about poor health and that they were "jinxed." Henry related this to an example from Boy Scouts in Trinidad: while walking through the forest, a scout does not say the word "snake," because he would be kicked by the scout behind them. Stating the word "snake" is equated to calling it forth (mentally or physically) and frowned upon among the scouts. By saying the word "snake," you jinxed yourself and your fellow scouts, because the likelihood of its appearance has increased. Other cultures have similar notions (e.g., knock on wood).

Whether or not you believe in such superstitions, it brings to mind that researchers may take into account the emotional burden of imagining poor health on their respondents. Drawing on the "rotten apple" analogy, if a respondent is given a choice between a rotten apple and a rotten pear, (s)he may express a preference, but greatly dislike the task. Emotional distress may cause individuals to drop out of the study thereby biasing the results.

The moral of the story is that consideration of the respondent's well being is needed in health preference research, particularly when the elicitation of preferences may impose emotional distress.