VALUE IN HEALTH ▮ (2016) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Preference Weighting of Health State Values: What Difference Does It Make, and Why?

Admassu N. Lamu, MA, MPhil[1,*], *Thor Gamst-Klaussen, MA*[1], *Jan Abel Olsen, PhD*[1,2]

[1]*Department of Community Medicine, University of Tromsø, Tromsø, Norway;* [2]*Centre for Health Economics, Monash University, Clayton, VIC 3800, Australia*

ABSTRACT

Background: Most patient-reported outcome measures apply a simple summary score to assess health-related quality of life, whereby equal weight is normally assigned to each item. In the generic preference-based instruments, utility weighting is essential whereby health state values are estimated through preference elicitation and complex algorithms. **Objectives:** To examine the extent to which preference-weighted value sets differ from unweighted values in the five-level EuroQol five-dimensional questionnaire and the 15D instrument, on the basis of a comprehensive data set from six member countries of the Organisation for Economic Co-operation and Development, each with a representative healthy sample and seven disease groups (N = 7933). **Methods:** Construct validities were examined. The level of agreement between preference-weighted and unweighted values was also assessed using intraclass correlation coefficient (ICC), Bland-Altman plots, and reduced major axis regression. **Results:** The performances of preference-weighted and unweighted measures were comparable with regard to convergent and known-group validities for each instrument. Although unweighted values in the five-level EuroQol five-dimensional questionnaire differ considerably from the preference-weighted values at the individual level, the discrepancy is minimal at the group level with a mean difference of 0.023. The ICC (0.96) and the Bland-Altman plot also suggest strong overall agreement. For the 15D, both the ICC (0.99) and the Bland-Altman plot revealed almost perfect agreement, with a negligible mean difference of −0.001. Results from the reduced major axis regression also showed small bias. **Conclusions:** Overall, preference weighting has minimal effect if the unweighted values are anchored on the same scale as the preference-weighted value sets.

Keywords: EQ-5D-5L, 15D, health-related quality of life, preference weighting.

Introduction

A wide range of instruments has been developed to measure patient-reported outcomes, often by use of a summary score to indicate the degree of disease severity [1]. Most of these instruments assign equal weight to each dimension or item included, that is, every health dimension and each level change are assumed to have equal importance. Furthermore, these instruments do not account for how people value a health state improvement relative to how they value lifetime gains.

Generic preference-based instruments are different. They were designed to enable comparisons of the effectiveness of competing health care programs in economic evaluations [2,3]. Because effectiveness can be in terms of both improved health and prolonged life, the health-related quality-of-life gains are made commensurable with lifetime gains, using a scale that accounts for people's trade-offs between quality and quantity of life. Furthermore, reflecting economists' attention to the preferences of affected parties, these instruments also seek to account for importance weighting of the included health dimensions. The distinct features of these preference-based instruments are that they 1) use a generic health state descriptive system designed to apply across all health conditions and 2) provide an indirect means of obtaining preference weights. Hence, respondents are assigned a health state value on the basis of their responses to a health state questionnaire, and prespecified preference weights obtained from other populations are then applied [4]. The focus on utility represents a key element, in that the class of cost-effectiveness analyses on the basis of these instruments is referred to by a specific term—cost-utility analyses.

The most widely used health state utility instrument is the EuroQol five-dimensional questionnaire (EQ-5D), followed by the six-dimensional health state short form, the health utilities index, and the 15D. Together, these four instruments are found in around 95% of applied cost-utility studies [5]. Furthermore, a

* *Address correspondence to*: Admassu N. Lamu, Department of Community Medicine, University of Tromsø, PO Box 6050 Langnes, Tromsø 9037, Norway.

E-mail: admassu.n.lamu@uit.no.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2016.10.002

review of 1663 studies using preference-based instruments published between 2005 and 2010 found that the EQ-5D had been applied in 63% of these studies [6]. In addition to their different descriptive systems, these instruments apply different preference elicitation methods: the visual analogue scale (VAS) or the choice-based methods of time trade-off (TTO), standard gamble, and discrete-choice experiments (DCEs). Furthermore, different scoring algorithms are used. Consequently, different instruments produce different preference weights [7,8].

Several researchers have questioned the complex algorithms used to create preference weights [9–11]. Richardson et al. [12] suggest that differences in preference weights are primarily via their effect on the measurement scales. Although each preference-based measure was developed on a unit scale of 0 to 1, their actual scales differ: the original English value set for the EQ-5D has a scale length of 1.594 (i.e., from −0.594 to 1), whereas the six-dimensional health state short form has a scale length of 0.699 (i.e., +0.301 to 1). The aim of this study was to examine what difference it makes to assign preference-weighted values to health states, as compared with the unweighted values obtained when summary scores are converted onto a 0 to 1 scale. Given that some preference-based instruments include negative values, reflecting that the most inferior health states are considered worse than being dead, parts of the discrepancy between preference-weighted and unweighted values are explained by scale length differences. Hence, a key issue is to make scale-adjusted comparisons to determine how much of the observed discrepancy is due to scale length differences, and how much is attributable to the importance weighting of health dimensions.

This article examines two preference-based instruments, the five-level EQ-5D (EQ-5D-5L) and the 15D, that are contrasting in terms of both their descriptive system and their valuation methods. The EQ-5D-5L has the most condensed descriptive system, including only five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression [13]. In the construction of the new EQ-5D-5L, the original dimensional structure was retained, but it now includes five levels of severity (no problems, slight problems, moderate problems, severe problems, and unable to/extreme problems) [14]. The 15D describes health along 15 dimensions (mobility, vision, hearing, breathing, sleep, eating, speech, bladder/bowel function, usual activities, mental function, discomfort/pain, depression, distress, vitality, and sexual activity), each with five levels, giving a combination of more than 30.5 billion ($=5^{15}$) possible health states [15].

As for valuation methods, in the 15D, subjects were asked to rank the dimensions and the levels within each dimension according to their relative importance using a 0 to 100 VAS scale, in which 100 was assigned to the most important dimension or level, and 0 was assigned if a dimension or level was not considered important at all [15]. The EQ-5D-5L tariff considered here is the latest version, which is based on an English population sample. It applies a combination of TTO and DCE tasks, which makes explicit trade-offs between quality and quantity of life, with scales that go below 0 [16].

Data and Methods

Data

Data were obtained from the Multi-Instrument Comparison (MIC) study, which is based on an online survey administered in Australia, Canada, Germany, Norway, the United Kingdom, and the United States by a global panel company, CINT Australia Pty Ltd. [17]. The personal and medical details recorded by the panel company were used to recruit individuals from a "healthy group" (N = 1760) and from seven major chronic disease groups (N = 6173). Quotas on age, sex and education were used to obtain a demographically representative sample of "healthy" respondents, defined by the absence of chronic disease and a VAS score of at least 70 on overall health. Quotas were also applied to obtain a target number of respondents in each disease group: arthritis, asthma, cancer, depression, diabetes, hearing loss and heart problems.

In addition to the MIC data set, the full set of the EQ-5D-5L health states ($N = 5^5 = 3125$) was used to explore the degree of agreement between preference-weighted and unweighted values. For the 15D, however, all analyses were based on the MIC data set because it is problematic to use the 30.5 billion full set of 15D health states. For the purpose of comparing preference-weighted and unweighted values in both the EQ-5D-5L and the 15D in terms of construct validity, four variables were considered: two variables (VAS and standard of living) correspond to the full sample (N = 7933) and the other two (diabetes 39 [D-39] and the Kessler Psychological Distress Scale [K10]) were taken from the included "disease groups." The D-39 and K10 were chosen because they were relatively more inter-related with both the EQ-5D-5L and the 15D dimensions.

Preference-Weighted Scoring Approach for the EQ-5D-5L and the 15D

The EQ-5D-5L

Health states defined by the EQ-5D-5L may eventually be converted to a single summary index by applying scores from a standard set of values (preferences) derived from general population samples [18]. In this study, the value set for the EQ-5D-5L is derived from the stated preference data of 996 members of the English general public, for which a hybrid model combining a composite TTO approach and DCE tasks was used for its direct elicitation [16]. The minimum value for the worst health state ("the pits") was−0.281, giving a scale length of 1.281 (i.e., from −0.281 to 1).

The 15D

The 15D tariff was generated using a set of preference weights elicited from several representative samples of the Finnish adult population [15]. Respondents were asked to assign the relative importance for 15D dimensions on a 0 to 100 scale, in which 100 was assigned to the most important dimension. Then, the importance of all other dimensions was assessed in relation to this most important dimension. Similarly, importance weights for levels within each dimension were produced on a 0 to 100 scale, in which the most desirable level (level 1) was assigned 100 and the desirability of all other levels was assessed in relation to level 1. In addition to the five levels, the states of "unconscious" and "dead" were also valued for each dimension. The preference weights were scaled on a 0 to 1 range, in which 0 represented "dead" and 1 represented "no problems on any dimension," and with no health state worse than being dead. The weights were obtained by using a rating scale (i.e. VAS) and then combined using a simple additive model. Hence, the 15D value set is not based on preferences that reflect the trade-offs between quality and quantity of life gains.

The Unweighted Scoring Approach

On the basis of the instruments' summary scores, unweighted health state values are developed, with each dimension assigned equal importance and each level change assigned the same weight. First, item scores are set equal to the rank order of the