VALUE IN HEALTH 15 (2012) 1059–1068

Available online at www.sciencedirect.com

SciVerse ScienceDirect

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

Preference-Based Assessments

Comparison of Generic, Condition-Specific, and Mapped Health State Utility Values for Multiple Myeloma Cancer

Donna Rowen, PhD, MSc, BA[1,*], *Tracey Young, PhD, CStat, MSc, BSc*[1,2], *John Brazier, PhD MSc BA*[1], *Sabine Gaugris, MSc*[3]

[1]*Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, UK;* [2]*NIHR Research Design Service for Yorkshire and the Humber, University of Sheffield, Sheffield, UK;* [3]*Janssen-Cilag Ltd., High Wycombe, UK*

ABSTRACT

Objective: Resource allocation informed by cost-utility analysis requires that the benefits be comparable across patient groups and interventions. One option is to recommend the use of one generic utility measure, but this raises the issue of comparability when the preferred measure is inappropriate or unavailable. Many cancer trials do not include generic measures such as the EuroQol five-dimensional (EQ-5D) questionnaire and instead include condition-specific measures and use these to generate utility estimates. We analyze the comparability of generic, condition-specific, and mapped utility values for a multiple myeloma cancer patient data set. **Methods:** Generic EQ-5D, condition-specific EORTC-8D, and EQ-5D utility values mapped from the EORTC QLQ-C30 were compared by using psychometric and statistical analysis to determine discrimination across severity groups, responsiveness, and agreement. **Results:** Generic, condition-specific, and mapped utility estimates were responsive over time and show discriminative validity. The EQ-5D had higher responsiveness and detected a greater change across severity groups and treatment periods than did the EORTC-8D but has a higher proportion of responses at full health (12.8%). Differences in the EQ-5D and the EORTC-8D were due at least in part to differences in the classification system. Mapped EQ-5D estimates had a smaller SD and do not reflect the severe range of health states reported by using the EQ-5D. **Conclusions:** Our findings suggest that condition-specific EORTC-8D or mapped EQ-5D utility estimates are broadly comparable to directly obtained EQ-5D utilities for a multiple myeloma patient data set. However, EORTC-8D estimates captured changes in quality of life for patients in mild health states that were not captured by the EQ-5D, but estimated lower utility gains than did the use of the EQ-5D directly.

Keywords: condition-specific measures, mapping, preference-based measures, QALYs, utility.

Introduction

Resource allocation informed by economic evaluation using cost-utility analysis has become increasingly popular in recent years. This analysis requires that the measures of benefit and cost for each evaluation be comparable across both different patient groups and different interventions. Payers in various European jurisdictions prefer the expression of benefit in quality-adjusted life-years (QALYs), which are a measure of both quantity and quality of life (QOL). Often, generic preference-based measures such as the EuroQol five-dimensional (EQ-5D) questionnaire [1], health utilities index 3 [2], or six-dimensional health state short form (derived from short form 36 health survey) [3,4] are used to calculate the "Q" component of the QALY. It is well documented, however, that different generic measures produce different results when applied to the same patient group at the same point in time [5]. This raises issues for comparability, and one solution is to recommend the use of a single measure for all evaluations. This is the approach taken by the National Institute of Health and Clinical Excellence (NICE) [6] where the most commonly used generic measure, the EQ-5D, is recommended for use in all technology appraisals. This raises the question of how utility values should be generated if the EQ-5D is either unavailable or inappropriate, and the comparability of evaluations undertaken in these circumstances.

Cancer is one condition in which it remains unclear whether the generic EQ-5D is appropriate, but the issue is further complicated by the fact that the EQ-5D is often unavailable because many cancer trials do not include it. NICE state that if a measure is thought inappropriate, empirical evidence should be provided demonstrating why it is inappropriate, covering properties such as content validity, construct validity, responsiveness, and reliability. A recent report argues that the EQ-5D may not be sufficiently sensitive to capture changes in the health status of cancer patients, as, for example, there is no EQ-5D dimension to specifically capture changes in vitality or energy [7]. There is little guidance, however, provided by NICE or similar agencies of when a measure can be deemed inappropriate for a patient group

* *Address correspondence to:* Donna Rowen, Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK.

E-mail: d.rowen@sheffield.ac.uk.

1098-3015 – see front matter Copyright © 2012, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2012.08.2201

or intervention, and this is an area requiring further research and guidance. If the EQ-5D is inappropriate, NICE states that other measures can be used [6].

Clinicians and researchers often choose to include condition-specific profile measures in trials rather than generic preference-based measures such as the EQ-5D. Condition-specific profile measures, such as the EORTC QLQ-C30, are often included because these capture the effects of interventions across a wide range of relevant symptoms, side effects, and aspects of functioning and QOL and their validity is well established. These profile measures have great clinical utility and are recommended by the US Food and Drug Administration [8], whereas the EQ-5D is recommended for effectiveness studies and economic evaluation and can be viewed as being an additional burden for completion for patients who are very unwell. These condition-specific profile measures, however, typically provide a description rather than a valuation of health and cannot be used to populate cost-effectiveness models. In recent years, there has been a growth in preference-based measures derived from existing condition-specific measures that enable these measures to be used directly to generate utilities. [9–11] The EORTC-8D is a recently developed condition-specific preference-based measure derived from the EORTC QLQ-C30 for use in patients with cancer [12]. This measure allows a utility estimate to be generated for every individual each time the EORTC QLQ-C30 is used and enables the direct estimation of utility without placing any burden on patients to complete an extra measure or additional questions. Mapping is an alternative method that can be used to obtain utility values when only a condition-specific non–preference-based measure was included in the trial. Mapping applies the statistical relationship between, for example, the QLQ-C30 and the EQ-5D to obtain predicted EQ-5D values from QLQ-C30 data. This relationship is typically obtained by estimating regressions on a separate data set that has patient characteristics similar to those of the trial. Published mapping algorithms are available that map the condition-specific QLQ-C30 onto the EQ-5D, and these algorithms can be applied to the trial data set to produce EQ-5D estimates. If the EQ-5D is unavailable in a trial, NICE [6] recommends that either mapping or other validated measures be used to produce utility values. NICE stipulates that the mapping must be based on empirical data and the other measures should have valuation methods that are comparable to those used for the EQ-5D (Measuring and Valuing Health [MVH] tarif) [13]. The validity of mapping has been questioned, not least because it relies on substantial overlap between both measures. For an overview of recommendations for the development and use of mapping algorithms, see Longworth and Rowen [14].

A small number of studies have examined the impact of using mapped EQ-5D estimates rather than directly generated EQ-5D utilities, finding different results across studies [15–17]. A large number of studies compare the performance of the EQ-5D with the performance of other main generic preference-based measures such as six-dimensional health state short form (SF-36 health survey) and HUI2 [5], but there are few comparisons of condition-specific and generic preference-based measures [18]. Furthermore, as far as the authors are aware, no study has examined the comparability of all preferred options for use in technology appraisals to agencies such as NICE; although the EQ-5D is the preferred option, under certain circumstances other generic, condition-specific, or mapped EQ-5D utility estimates can be used.

This article compares utility values generated by using the EQ-5D, a condition-specific preference-based measure, and mapping for a cancer patient data set. We compare utility values obtained by using the generic preference-based EQ-5D, the condition-specific preference-based EORTC-8D derived from the EORTC QLQ-C30, and three published algorithms mapping the QLQ-C30 onto the EQ-5D. We further compare the performance of EORTC-8D and EORTC QLQ-C30 summary scores to determine whether the EORTC-8D maintains the desirable properties of the original measure. This article seeks to inform researchers and policymakers in their choice of the source of utility values and the interpretation of these values regarding discrimination across severity groups, responsiveness, and agreement.

Summary of Measures

The EQ-5D

The EQ-5D has five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each with three levels of severity ranging from "no problems" to "severe problems" [1]. The health state classification system describes 243 unique health states and utility values that range from 1 to −0.594 for the UK value set collected in the Measuring and Valuing Health study [13].

The EORTC QLQ-C30 and the EORTC-8D

The QLQ-C30 is widely used in cancer clinical trials in Europe and Canada [19] and has been found valid for many cancer conditions. The QLQ-C30 has 30 questions that cover functioning (physical, role, social, emotional, and cognitive functioning) and common cancer symptoms (pain, fatigue, nausea, vomiting, dyspnea, appetite loss, sleep disturbance, constipation, and diarrhea) plus the financial impact of the disease and treatment (excluded from analyses here because this is inappropriate for inclusion in health-related QOL measurement to generate QALYs). The QLQ-C30 has 14 summary scales ranging from 0 to 100, each representing an aspect of functioning (five summary scales, higher scores represent higher functioning) or a particular symptom (nine summary scales, higher scores represent greater symptoms), with one additional global QOL scale.

The EORTC-8D has eight dimensions (physical functioning, role functioning, pain, emotional functioning, social functioning, fatigue and sleep disturbance, nausea, and constipation/diarrhea), each with four or five levels of severity. The health state classification system was derived from 10 QLQ-C30 items and describes 81,920 unique health states with a range of utility values from 1 to 0.291 [12].

Methods

Utility values were generated by using the available preference weights for the EQ-5D and the EORTC-8D for each patient at each time point in the data set. Mapped utility values were also estimated for each patient at each time point by using published algorithms described below.

Estimating EQ-5D Utilities by Mapping the QLQ-C30 Onto the EQ-5D

The easiest way to produce mapped estimates is to use published algorithms. Nine published algorithms use mapping to produce utilities by using EORTC QLQ-C30 data, three of which were used here [15,20,21]. The other six algorithms are not used here because one article requires FACT data not available in our data set [22], one article maps to patient time trade-off values rather than the EQ-5D [23], one article does not publish the mapping function [24], one article maps to the EQ-5D scored by using the US value set [25], one article maps to the EQ-5D scored by using the Korean value set [26], and one article maps to the EQ-5D scored by using the Dutch value set [27]. Patient valuations of own health by using preference elicitation techniques such as time trade-off or visual analogue scales are not preferred by agencies such as NICE or the Washington Panel of Cost Effectiveness [28] because public preferences are preferred given that public funding is often used to provide health care. Patient values