Article:
Brazier, J.E., Rowen, D.L., Lloyd, A. et al. (2019) Future directions in valuing benefits for estimating QALYs: Is time up for the EQ-5D? Value in Health, 22 (1). pp. 62-68. ISSN: 1098-3015

https://doi.org/10.1016/j.jval.2018.12.001

VALUE IN HEALTH 22 (2019) 62–68

Available online at www.sciencedirect.com

ScienceDirect

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

Themed Section: Evolution of EuroQoL

Future Directions in Valuing Benefits for Estimating QALYs: Is Time Up for the EQ-5D?

John Edward Brazier, PhD [1,*], *Donna Rowen, PhD* [1], *Andrew Lloyd, DPhil* [2], *Milad Karimi, PhD* [1,3]

[1]*Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, Sheffield, UK;* [2]*Acaster Lloyd Consulting Ltd, London, UK;* [3]*Institute of Health Policy and Management, Erasmus University, Rotterdam, The Netherlands*

ABSTRACT

The widespread adoption of the EuroQol 5-dimensional questionnaire (EQ-5D) has been important for the comparability, transparency, and consistency of economic evaluations for informing resource allocation in healthcare. The objectives of this article were to (1) critically assess whether the widespread adoption of the EQ-5D and its time trade-off–based value sets to inform economic evaluation is likely to continue and (2) speculate about how benefits may be measured and valued to inform economic evaluation in the future. Evidence supports the use of the EQ-5D in many areas of health, but there are notable gaps. Furthermore, there has been interest among some policy makers in measuring changes in well-being, and in using common outcomes across sectors. Possibilities for measuring well-being alongside health can be achieved through bolt-on dimensions or an entirely new measure capturing both health and well-being. Nevertheless, there are significant concerns about the logic of estimating a common utility function. The development of online valuation methods has had a major impact on the field, which is likely to continue. We, however, recommend more allowance for respondents to consider their answers. There is an ongoing debate on the role of patient values or experience-based values. To date, this has seen limited take-up by decision makers and there are significant technical problems to obtaining representative and meaningful values. Policy makers and the general population must decide on the focus and scope of benefits that are incorporated into economic evaluation, and current evidence on this is mixed. In part, this will determine whether the widespread adoption will continue.

Keywords: bolt-on dimensions, capabilities, carer quality of life, economic evaluation, experience-based utilities, health-related quality of life, QALYs, utilities, well-being

Introduction

Economic evaluation has been used to inform the allocation of scarce healthcare resources for many years. The most commonly used technique has been to estimate the incremental cost per quality-adjusted life-year (QALY) of new health technologies. QALYs provide a way to capture the benefits of healthcare in terms of impact on survival and health-related quality of life (HRQOL). A QALY is generated by multiplying life-years by a quality adjustment weight or health utility, Q, that is used to reflect the HRQOL of the person.[1] To derive the health utility requires both a *description* of health and a *value* of the state of health. The description is usually generated using self-complete responses to a questionnaire about health that allow the respondent to be assigned to a single multidimensional health state. The value of this health state is estimated to lie on a scale where 1 is equivalent to full health and 0 is equivalent to being dead, and values less than 0 indicate that the state of health is worse than being dead. These values reflect preferences around how good or bad different health states are; they are used to inform health technology assessment (HTA) of different health interventions and are typically elicited from members of the general population.

To inform economic evaluation in healthcare, generic preference-based measures (GPBMs) are most commonly

Conflicts of interest: The authors have indicated that they have no conflicts of interest with regard to the content of this article.
Source of financial support: There was no funding received for this project.

* *Address correspondence to:* John Brazier, Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent St, Sheffield S1 4DA, UK.
E-mail: j.e.brazier@sheffield.ac.uk

1098-3015 - see front matter
https://doi.org/10.1016/j.jval.2018.12.001

recommended by reimbursement agencies to both describe and value health.[2] These measures are generic because they are intended for use across conditions and treatments and, it is hoped, to provide consistency and comparability between economic evaluations. A GPBM of health has 2 components: a classification for describing health states and a set of utility values, called a value set or tariff, for each health state defined by the classification or a scoring algorithm. Different GPBMs have different classification systems, meaning that the dimensions of health, their descriptions, and their severity levels differ across GPBMs. There are also differences in valuation methods and protocols, and in the population used to provide the values. For these reasons, GPBMs generate different utility values (and differences in change in utility over time and across treatments) irrespective of whether there are differences in the health status of the population.[1] This has led to some agencies preferring 1 measure to be used in HTA submissions (eg, National Institute for Health and Care Excellence [NICE][3] and the *Zorginstituut Nederland* [ZIN][4]).

The most widely used GPBM is the EuroQol 5-dimensional questionnaire (EQ-5D).[5] This was developed to provide a simple and easy-to-use measure of HRQOL for administration alongside disease-specific measures and to provide a measure that could be readily valued by the general population.[6] The EQ-5D describes health status across 5 dimensions—mobility, self-care, usual activities, pain/discomfort, and anxiety/depression—and sets of preference weights that can be applied to convert it into a GPBM. In response to concerns about the insensitivity of the 3 severity levels used in the EQ-5D, a 5-level version has been developed.[7] The original 3-level version defines 243 states and the new 5-level version defines 3125. These states have been valued using various valuation techniques. The most widely used in the context of economic evaluation have been variants of the time trade-off (TTO) technique across numerous countries and more recently versions of discrete choice experiments (DCEs). The EQ-5D can be administered as a patient-reported outcome measure without scoring responses using the value set, meaning it can be used outside of applications in economic evaluation (eg, in routine outcome assessment).

The widespread adoption of the EQ-5D and associated value sets across different countries is a testament to the vision of the pioneers who developed it (eg, Brooks and the EuroQol Group[6]). But how much longer will the EQ-5D (3- or 5-level version) be the dominant measure in HTA? It is not in other related sectors such as social (or personal) care that have used different measures (eg, The Adult Social Care Outcomes Toolkit (ASCOT)[8]) and impact on carers (eg, Carer-Qol[9–11] and Carer Experience Scale[12]). These measures cover nonhealth domains such as autonomy, social participation, and meaning. Indeed, outside of healthcare, economists typically use monetary measures of all relevant benefit. Even within HTA, the impact of interventions extends beyond healthcare to other sectors and therefore to nonhealth benefits. Furthermore, the EQ-5D was never intended to cover all dimensions of health, some of which may be important for conditions such as vision, hearing, and some mental health disorders.[6] For this reason, the EQ-5D may not accurately capture the benefits of some interventions. This means that different types of measures sometimes need to be used, which can affect consistency in resource allocation informed by economic evaluation. There has also been growing interest in different conceptualizations of benefit such as well-being and capability[13–17] with claims that they may achieve consistency and capture benefits more relevant to decision making.[16,17]

At the same time there have been important developments in the valuation of the EQ-5D and other measures in the last 2 decades. Much of this has been reviewed in an article by Stolk et al[18] in this special issue. In our article we do not address the more technical issues around the different variants of TTO or DCEs or how they should be analyzed, but the more broader issues in the field such as apparently competing approaches to enhance the role of deliberation compared with the growing use of online data collection and the role of experience-based or patient values.

The objectives of this article were to (1) critically assess whether the widespread adoption of the combination of the EQ-5D and its corresponding TTO-based value sets to inform economic evaluation is likely to continue and (2) draw on recent developments in the field to speculate how benefits may be measured and valued to inform economic evaluation in the future. Nevertheless, inevitably the article will reflect the interests of the authors and involve speculation beyond what is known about future health policy. The article documents the rise to dominance of the EQ-5D, outlines recent developments in the scope of what is measured by the classification system and valuation, and speculates about what the future may bring.

The Rise to Dominance of the EQ-5D

A review of articles published on Web of Science (2004-2010) reported that 63.2% of studies using a GPBM use the EQ-5D, followed by the Health Utilities Index Mark 3 (9.8%) and the 6-dimensional health state short form (8.8%).[5] How did the EQ-5D come to dominate economic evaluation? When it came onto the scene in the early 1990s, most cost-per-QALY studies used the "vignette" approach, where researchers would construct bespoke descriptions of how health conditions had an impact on the lives of patients (eg, Sackett and Torrance[19]). There are major limitations with using vignettes arising from the poor linkage to evidence and the lack of comparability between studies. For this reason, researchers welcomed the development of standardized measures of health that could be completed very simply by the patient and used to estimate utilities from existing sets of preference weights.

The first to be developed was the Quality of Well-Being Index and its predecessor the Index of Well-Being.[20–22] There are a number of other GPBMs that have been available for the last 2 decades that can be used in the same way.[5] As mentioned earlier, they have been shown to produce different values on the same patients and so the choice of measure can have important implications for the final incremental cost effectiveness of an intervention. One solution to this problem has been the call to adopt 1 measure as a reference case, but why was the EQ-5D adopted by NICE and more recently by ZIN in the Netherlands and appears in a large number of HTA guidelines? One reason could be evidence of its psychometric performance. Indeed there is a growing literature on the measurement properties of GPBMs of health in different conditions[1,23,24] As reviewed elsewhere, the EQ-5D has been found to perform well in many conditions, although some problems have been identified in more complex areas of mental health,[25] dementia,[26] and sensory conditions.[23] Nevertheless, most of the available evidence relies on relatively crude tests of validity, such as whether they reflect differences between disease severity groups or changes over time after treatment. The evidence base is often quite patchy with very little head-to-head comparison across the measures. A key problem is the absence of a criterion standard, making it impossible to demonstrate validity absolutely, but instead weighing up the strength of evidence supporting a measure. Nonetheless, the EQ-5D has been shown to perform well in terms of many of these tests of measurement sensitivity in conditions as wide-ranging as type 2 diabetes, cardiovascular disease, and sexual functioning.[27–29]

The adoption of the EQ-5D may be partly because it was developed by researchers from many countries with the widespread availability of approved translations, whereas most other instruments were developed in 1 country (eg, the Quality of

Well-Being in the United States[22], the Health Utilities Index Mark 3 in Canada,[30,31] and the 15D in Finland[32]). Anecdotally, many clinicians suggested that the instrument had face validity and liked its short format. It has always been free to use for academic research. These features are not unique to the EQ-5D, and its widespread adoption is likely to be as much to do with the influence of the developers as the science per se. A key boost to its uptake was its adoption in the United Kingdom by NICE for use in economic evaluation in HTA.[3,33,34] A number of other key agencies (eg, ZIN in the Netherlands[4]) also prefer utility values generated by the EQ-5D. This and the use of the EQ-5D in many clinical studies and large surveys become self-fulfilling because consistency with past evidence is best ensured by using the EQ-5D in any future study. Nevertheless, will this dominance of the EQ-5D be undermined by recent developments in the assessment of benefits?

Recent Developments in Assessing Benefits of Healthcare

To consider the future use of GPBMs of health, and the EQ-5D in particular, we will examine recent developments in benefit assessment in healthcare and their likely implications for the future of the field including the development of new broader measures of well-being and capabilities and developments in valuation.

Going Beyond Health: Well-Being and Capabilities

The EQ-5D was not intended to cover all dimensions of health, and there is some evidence that this results in insensitivity or inappropriateness in some patient groups, such as those listed earlier. Furthermore, it is a measure of health and was not intended to measure the nonhealth impacts of some healthcare interventions and nor those of nonhealthcare interventions. The most closely related sector is social care, which is concerned with meeting the needs of service users with such things as feeding, clothing, cleaning, and participation. For social care, palliative care, and the care of many long-term conditions, the outcomes of care are not simply improved health per se, but include nonhealth outcomes such as dignity, autonomy and control, satisfaction with relationships, and having meaningful activities.[8] The importance of social care is expected to grow in developed countries, with the aging population effectively increasing the number of years people spend in ill-health.[35] Furthermore, there are important interactions between health and social care with predictions of greater integration in some countries. These include developments such as new models of care facilitating integration across acute, primary, mental, specialist, and social care services in the United Kingdom.[36] There are also important implications for informal carers, who actually provide much of the social care. This means that agencies seeking to compare outcomes across sectors cannot use a sector-specific measure such as the EQ-5D. In recent years there have been important developments within and outside health to look at alternatives to health outcomes, and these have included well-being and capabilities.

Well-being

This raises the issue of what is meant by "well-being." A broad conception of well-being is how well an individual's life is going on.[37] Subjective well-being (SWB) has been described or categorized into 3 types: hedonism (well-being increases when an individual experiences more pleasure and/or less pain), flourishing theories (well-being increases when an individual fulfills their nature as a human being, or "flourishes"), and life evaluation or life satisfaction (well-being increases when an individual positively assesses his or her life).[38] Traditionally, there are also objective list accounts of well-being including items such as literacy, accommodation, and ability to see.[39] There are a number of tools available to measure SWB, including simple self-reported items on happiness and life satisfaction, and multi-item measures of psychological well-being such as the Warwick-Edinburgh Mental Well-Being Scale.[40] Here we focus on the different approaches to using well-being, rather than a detailed review of the different measures (for this, see study by Peasgood et al[37]).

There are 2 lines of argument for using well-being measures. One is that they might be better able to take into account nonhealth dimensions alongside health dimensions. Well-being could provide an overall indicator of quality of life, which could be used between health and other sectors and so overcome some of the limitations of different measures in health and social care. Another argument for well-being is that the EQ-5D and other GPBMs use (typically) general population preferences, elicited using techniques such as TTO where respondents are being asked to imagine health or social care states. These approaches assume that individuals are able to predict the likely impact of the health state being described on their future lives, but this has been shown not to be the case in health and other contexts.[16] General population respondents usually do not take into account the extent of any adaptation they may make over time,[41] meaning that their preferences will provide a poor indicator of the actual impact on their well-being. A more direct description of well-being might provide a more accurate basis for members of the general population imagining a state.

The use of well-being measures, particularly SWB measures, is not without its critics. There are many different measures of SWB, but none currently provide a basis for estimating QALYs, because most are not preference-based and none generate values on the QALY scale. Nevertheless, it might be possible to construct a classification system on the basis of well-being dimensions and value it on a QALY scale. Well-being as a concept also has a lack of agreed definition and little evidence of measurement accuracy. It has been suggested that "wellbeing policy is running ahead of the evidence."[42] Policy makers in health and social care wishing to assess the role of well-being in informing priorities face a number of uncertainties about the validity and merits of well-being measures. Psychometric evidence suggests that the items do not seem to form the constructs originally intended, with little difference between items intended to tap quite different constructs.[43,44] SWB measures also have been shown to be less sensitive to differences than the EQ-5D across key health groups.[44] Although this is perhaps not surprising because SWB is a different construct, it has implications for sample size and for their use in decision making. The main limitation to using just well-being in healthcare is evidence that many policy makers continue to be interested in traditional physical health outcomes.[37]

Capabilities

The notion of SWB is often confused with capabilities, but this concept has a different genus. It can be attributed to Amartya Sen, who argued that society is interested with what you *can* do or be (ie, capabilities), and not just what you actually choose (or happen) to do or be (which he calls *functioning*).[45] This addresses the fact that we may not choose to walk to the shops, but we value the ability to do so, although he does not seem to suggest that only capabilities should matter. Sen argues it is not possible to come up with a list of attributes suitable for all contexts, although there are important examples, and 1 major attempt is the ICECAP measure. The 5-item ICECAP-A is similar to GPBMs of health with a multidimensional classification defining capability states valued using a preference elicitation technique. It aims to measure generic capabilities that are required to have a high quality of life. ICECAP is a measure of capabilities for use in both health and social

care.[15,46–48] It has been examined for construct validity[49] and test-retest reliability.[50] Initial evidence suggests there may be some small differences between item wording capturing whether individuals "can" have key functioning rather than actually "do" have, although further research is recommended.[51]

Although it has been used in health and social care and so could provide a common measure across sectors and the content is closer to SWB, there are concerns with using capabilities, many of which are similar to SWB. The valuation method of ICECAP-A did not include trade-offs between improvements in capabilities and years of life/survival, and instead assumes that having no capabilities is equivalent to being dead. This means that it is not valued on the 1-0 full health–dead scale used to generate QALYs, and so it is not directly comparable with other instruments designed to measure QALYs. It is unclear whether ICECAP is actually consistent with the capability approach because (1) it seems to suggest that only capability counts, whereas functioning may still matter to decision makers, and (2) it is not clear whether it measures a capability "set" because it does not allow for the natural interdependence between dimensions.[13] It has been shown to be less sensitive than the EQ-5D for many physical conditions,[44] and as for SWB, many decision makers remain interested in physical outcomes for their own sake.[37] This suggests that ICECAP in particular does not fully address the decision context of HTA.

Ways Forward: Moving Beyond Health?

Decision makers have a broad range of opinions regarding what outcomes matter and the role of well-being in particular. A study conducted through focus groups and interviews with members of the NICE technology appraisal, social care, and public health committees and members of their citizens' council found that "outcomes such as relationships, a sense of control, being able to do the things you want to, and positive emotion were considered important aspects of quality of life; current measures, such as EQ-5D have insufficient content capturing social and emotional health; that health (including physical functioning) continues to be important; and decision makers lack the tools necessary to consistently incorporate wellbeing into decision making (i.e. valid, well understood measures)."[37] There was little support for relying solely on SWB. One implication is that there is a need for a measure that captures both health and key aspects of non–health-related quality of life. This can be achieved either by incorporating additional dimensions into the EQ-5D or by developing a new measure.

Bolting on missing dimensions

When evidence has suggested that the EQ-5D is unable to capture change in a specific aspect of health, such as hearing or vision problems, then there has been research to explore how these domains could be added to the instrument. Although still at an exploratory stage, this so-called "bolt-on" research has expanded significantly in recent years. There has been some exploratory work on a number of EQ-5D bolt-ons or extra dimensions, but as yet similar work does not seem to have been undertaken with other GPBMs of health. Existing areas of bolt-on research include sleep,[52] vision and hearing,[23,53] and cognition,[54] although not all bolt-ons seem to have a significant impact on health state values (eg, sleep). This work could be extended to include some of the aforementioned well-being dimensions such as autonomy and control, relationships, and positive emotion.

Nevertheless, bolt-on items have been shown to have an impact on the coefficients of the 5 original dimensions rather than being simply additive.[23,53] Given there is overlap between health and well-being dimensions, particularly with mental health,[44] this then means that any new additional dimension requires a new preference-based value set. The overlap may raise more fundamental concerns about the appropriateness of combining different concepts such as health and well-being (discussed later). There are also concerns that bolt-ons may be developed to capture the specific benefit of a drug or intervention in an attempt to maximize a utility gain. More generally, it has the potential to undermine the generic nature of the EQ-5D and to affect the comparability of utilities from different disease areas.

Developing a new measure of quality of life

Another approach to moving beyond health would be to develop a new measure that includes important health dimensions alongside those linked to well-being. This approach has already been used in the development of the Assessment of Quality of Life instruments by Richardson et al in Australia.[55–57] The Assessment of Quality of Life 8D measures HRQOL in a broader way than do usual instruments because of the inclusion of 5 psychosocial aspects (mental health, happiness, self-worth, relationships, and coping) alongside health. It is a long instrument (35 items) that was developed in only 1 country. More evidence regarding its feasibility in real-life settings is required because it takes rather longer to complete than the EQ-5D, for example. There is a new measure that is being developed at the time of writing that aims to cover health and those aspects of well-being considered important by service users.[37,58] This international research is looking at the feasibility and implications of a broader measure that is jointly funded by the UK Medical Research Council and the EuroQol Group.

This approach raises important concerns. It remains theoretically unclear whether a single measure can meaningfully capture health and other quality-of-life considerations into a single classification system that can be valued alongside each other because health is an aspect of quality of life and also a facilitator or enabler of other aspects of quality of life. How will respondents weigh up dimensions where one is probably causal to another in the same state being valued? Although this is an issue for measures such as the EQ-5D, because symptom items such as pain and depression have implications for usual activities, adding in well-being items is likely to make this problem more challenging.

Recent Developments in Valuation

There have been major developments in valuation methods, and this literature is far too large to review here. We will focus on 3 broad developments: the growing use of online computer-based methods, the use of more deliberative and informed methods, and "experience-based" utility. Readers interested in the details of specific methods of elicitation and the many debates between the methods (eg, TTO vs standard gamble [SG]) are invited to consult key texts on the subject (eg, studies by Brazier et al[1] and Drummond et al[59]).

Use of Online Samples

Computer-based methods for administering preference elicitation tasks such as TTO and SG have been available for nearly 2 decades, but it is only recently that the real potential of computer-based methods has started to be realized. A key development has been the use of online panels that enable quick and comparatively cheap ways of collecting large samples. There are, however, concerns about the impact of online methods for the quality of data. Conventional valuation tasks such as TTO or SG are unfamiliar and often complex, and evidence suggests that respondents benefit from having the human contact of an interviewer and even the use of a physical prop.[60] Nevertheless, there has been great interest in recent years in valuation research on

the use of preference methods such as DCEs and best worse scaling, which are arguably more feasible to complete online without an interviewer being present. Furthermore, using DCE with duration as an additional dimension enables the preference weights to be placed on the QALY scale. Experimentation and piloting of valuation methods have become much simpler and cheaper, which in turn will allow researchers to make more rapid progress in the development of valuation methods. Nevertheless, the level of engagement and impact on data quality needs to be examined rigorously with qualitative and quantitative methods before they be used to inform decision making.

Experience-Based Utility

Experience-based utility values are when respondents value their own current health state rather than asking members of the general public to value hypothetical health states. This approach does not require people to imagine hypothetical ill-health states because it asks people about the state they are experiencing (although they are being asked to imagine full health), and hence may provide more accurate values of what it is like to live in a state. Evidence shows that experience-based utility values differ from hypothetical health state values.[61–63] Many reasons may be attributed to this; for example, patients have a different perspective and better understanding of the health state because of the knowledge of actual impact rather than perceived impact from abstract descriptions (eg, EQ-5D states), and patients may have adapted to their health state.[64] Although it continues to be explored in published literature, there has been little interest among policy makers with the notable exception of Sweden.[65] Most have been influenced by the argument that because the general public pays for healthcare through taxation (in many but not all countries), then arguably it should be their hypothetical values that set the priorities for the health service rather than values from patients experiencing the health problems.[66]

There are important methodological questions surrounding how to elicit experience-based utilities in a way that could be used in economic evaluation (see, eg, the study by Brazier et al[67]). There are substantial problems with eliciting values from those experiencing the state—namely, patients—many of whom will be suffering from severe illness. The general population respondent is poorly informed about the ill-health state, but then many patients may have little memory of full health. There are also ethical issues in asking people in poor states to value their own health by considering being dead or being fully healthy again. This can have an impact on the elicitation technique because studies often used the visual analogue scale to avoid these problems (see, eg, Sun et al[68]). It is also unclear whether what they are really valuing in TTO or the visual analogue scale is consistent with the QALY, because they may not be imagining a fixed health state for the rest of their life, but an unknown lifetime profile of health that will be dependent on their current condition. Furthermore, those who respond to a survey in a severe disease, such as advanced cancer or a severe chronic disease, are unlikely to be representative of those in their health condition.

Deliberative Approaches

In contrast, with the growing use of online methods, an alternative approach is the use of methods that allow respondents more time to reflect on their valuations, consider more evidence (such as the views of patients about the state, or evidence regarding adaptation), and/or deliberate with other respondents, friends, and relatives. Evidence is mixed on whether longer time for reflection and deliberation alters the values, with some evidence for change[69–71] compared with another study with no change,[72] but research into this has been comparatively small-scale. This could be seen as a proxy for experience-based utility values that does not suffer from the methodological challenges of collecting and analyzing experience-based utility data, or a radically different means to getting the views of society.

Options for obtaining better informed preferences involve more deliberative techniques including the use of citizens' jury or a process similar to multicriteria decision analysis, with a small sample of participants who become informed during the jury process through reviewing and examining evidence on the health state, who then deliberate on health state values and reach democratic recommendations. Numerical utility values can be elicited through the use of traditional health state elicitation techniques such as TTO, or DCE, or those more commonly used in multicriteria decision analysis such as swing weighting, during or after the deliberation process.[67]

Deliberative approaches are an important methodological development, although their uptake in valuing health states is likely to be limited because of the time it takes and the problems of standardizing methods given the large range of options.

Conclusions

The widespread adoption of the EQ-5D has been an important enabler of comparability, transparency, and consistency of economic evaluations for informing resource allocation in

Table 1 – Moving beyond health to include other aspects of quality of life: pros, cons, and challenges.

Pros	Cons	Challenges
• The EQ-5D is a measure of health that was not intended to measure the nonhealth impacts of healthcare interventions nor the impact of nonhealthcare interventions • In social care, palliative care, and the care of many long-term conditions, outcomes of healthcare include nonhealth outcomes, such as dignity • Enables measurement of all outcomes occurring from both health and social care • Enables consistency in resource allocation decisions across sectors and easier comparisons of outcomes across sectors	• Resource allocation decisions and policies may be informed by the ability of healthcare to have an impact on nonhealth outcomes, where this may not be their remit • May not be sensitive to healthcare interventions or to differences across key health groups	• To inform economic evaluation, a single measure of utility to generate QALYs is recommended (rather than 1 measure for health and 1 for other outcomes) • It is unclear whether a single measure can meaningfully capture health and other quality-of-life considerations into a single classification system that can be valued alongside each other • Health is a facilitator or enabler of other aspects of quality of life, and the logic of estimating a common utility function requires careful consideration

QALY indicates quality-adjusted life-year.

healthcare. Nevertheless, is it a tool that remains fit for purpose, where that purpose is primarily to measure the benefits for the patients receiving healthcare technologies? Evidence in healthcare supports the use of the EQ-5D in many areas of health, although there are some notable gaps because it does not cover all dimensions of health.

In some countries, economic evaluation is used to cover a broader range of sectors including public health, social care, and other nonhealthcare sectors and there is a degree of integration between them and healthcare (but this may not be relevant to all countries). In countries where economic evaluation covers broader sectors, we expect to see that there will be a move to the use of measures that capture not only health, but also well-being. Nevertheless, this is not without its challenges, and this is not to state that measures of only health will not also be used or used alongside.

Policy makers and the general population must ultimately make a judgment around the focus and scope of benefits that are incorporated into economic evaluation. Table 1 provides a summary of the pros, cons, and challenges associated with moving beyond health to include other aspects of quality of life. In the United Kingdom, the picture around the views of policy makers and the general population is mixed, with some decision makers favoring the inclusion of nonhealth outcomes related to well-being such as relationships, control and autonomy, and positive emotion, but there seems little support for relying solely on these well-being outcomes. Decision makers and the general public want to retain more conventional health outcomes such as pain, mobility, and psychological health in any assessment alongside any well-being components. Measures should also reflect what matters to the users of services. Whether consideration of well-being can be achieved through the addition of bolt-ons or an entirely new measure capturing both health and well-being remains to be seen. There are significant theoretical concerns about the way health may be valued in its own right as well as through well-being, and the logic of estimating a common utility function needs to be examined.

With the growing proliferation of measures, it might be argued that a new measure is not what the field needs at the moment. Without doubt, any new measure requires considerable input into its development and careful psychometric testing and validation before any potential adoption to inform policy. Nevertheless, although the development of a new measure provides considerable challenges and resources, this should not act as discouragement to the enterprise because no existing measure should be used simply because it already exists if the scope and focus may be considered inappropriate for policy or a particular research question. The marginal benefit of adopting a new measure should be weighed against the loss of consistency.

The development of online valuation methods has had a major impact on the field, and this is likely to continue. Nevertheless, we would voice a note of caution and recommend that more time be spent in allowing respondents time to deliberate on their answers because the numbers will have such an important application in informing the allocation of scarce resources.

Finally, there has been an ongoing argument on the role of patient values or experience-based values, and we expect this debate to continue. To date, this perspective has seen little take-up by decision makers and there are significant technical problems to obtaining representative and meaningful values that reflect just the health state. We anticipate that the use of experience-based values may play into an agenda to increase the patient voice, but the issue is a normative one of whose values should be used to inform resource allocation in a publicly funded system.

Many decision makers are likely to continue to want to focus on health and maintain consistency with past decisions. As the decision of whether and when to adopt the 5-level EQ-5D by NICE shows, consistency with past decisions is key, and any new measure, change in whether the measure captures beyond health, or any method of valuation will be subject to far greater scrutiny than the original 3-level EQ-5D and its value sets.

Acknowledgment

J.B. and A.L. are members of the EQ Group Executive. We thank the Yorkshire and Humber CLAHRC for support.

REFERENCES

1. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford, UK: Oxford University Press; 2017.
2. Rowen DL, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. International regulations and recommendations for utility data for health technology assessment. Pharmacoeconomics 2017;35(suppl 1):11–9.
3. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal. London: National Institute for Health and Care Excellence; 2013.
4. Busschbach JV, van Hout BA, de Wit A. BIJLAGE 2: QALY en kwaliteit: van leven metingen. Diemen, The Netherlands: Zorginstituut Nederland; 2016.
5. Richardson J, McKie J, Bariola E. Multiattribute utility instruments and their use. In: Culyer AJ, editor. Encyclopaedia of Health Economics, Vol. 2. San Diego, CA: Elsevier; 2014. p. 341–57.
6. Brooks REuroQol Group. EuroQol: the current state of play. Health Policy 1996;37:53–72.
7. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of a new five-level version of the EQ-5D—the EQ-5D-5L. Qual Life Res 2011;20(10):1727–36.
8. Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference-weighted measure. Health Technol Assess 2012;16:1.
9. Brouwer WB, Van Exel NJ, Van Gorp B, Redekop WK. The CarerQol instrument: a new instrument to measure care-related quality of life of informal caregivers for use in economic evaluations. Qual Life Res 2006;15(6):1005–21.
10. Hoefman RJ, Van Exel NJA, Rose JM, et al. A discrete choice experiment to obtain a tariff for valuing informal care situations measured with the CarerQol instrument. Med Decis Making 2014;34(1):84–96.
11. Hoefman RJ, van Exel J, Brouwer WBF. Measuring care-related quality of life of caregivers for use in economic evaluations: CarerQol tariffs for Australia, Germany, Sweden, UK, and US. Pharmacoeconomics 2017;35(4):469–78.
12. Al-Janabi H, Flynn T, Coast J. Estimation of a preference-based Carer Experience Scale. Med Decis Making 2011;31(3):458–68.
13. Karimi M, Brazier J, Basarir H. The capability approach: a critical review of its application in health economics. Value Health 2016;19(6):795–9.
14. Brazier J, Tsuchiya A. Improving cross-sector comparisons: going beyond the health related QALY. Appl Health Econ Health Policy 2015;13(6):557–65.
15. Grewal I, Lewis J, Flynn TN, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: preferences or capabilities? Soc Sci Med 2006;62:1891–901.
16. Dolan P, Kahneman D. Interpretations of utility and their implications for the valuation of health. Econ J 2008;118(525):215–34.
17. Well-being in public policy. https://www.whatworkswellbeing.org/wp-content/uploads/2018/03/Overview-incorporating-wellbeing-in-policy-analysis-vMarch2018.pdf. [Accessed January 8, 2019].
18. Stolk E, Rand K, Ludwig K, Van Hout B, Ramos-Goni J. Overview, update and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. Value Health 2019;22(1):23–30.
19. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. J Chronic Dis 1978;31:697–704.
20. Fanshel S, Bush JW. A health-status index and its application to health-services outcomes. Oper Res 1970;18(6):1021–66.
21. Patrick DL, Bush JW, Chen MM. Measuring levels of well-being for a health status index. Health Serv Res 1973;8:228–45.
22. Kaplan RM, Anderson JP. A general health policy model: update and application. Health Serv Res 1988;11:478–507.
23. Longworth L, Yang Y, Young T, et al. Use of generic and condition specific measures of health related quality of life in NICE decision making. Health Technol Assess 2014;18(9):1–224.

24. Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. European J Health Economics 2017:1–14.
25. Brazier J, Connell J, O'Cathain A. Do EQ-5D and SF-6D ask the right questions in mental health? A content validation using interviews with patients. Value Health 2014;17(3):A194.
26. Hounsome N, Orrell M, Edwards RT. EQ-5D as a quality of life measure in people with dementia and their carers: evidence and key issues. Value Health 2011;14(2):390–9.
27. Clarke P, Gray A, Holman R. Estimating utility values for health states of type 2 diabetic patients using the EQ-5D. Med Decis Making 2002;22(4):340–9.
28. Briggs AH, Parfrey PS, Khan N, Tseng S, Dehmel B, Kubo Y, et al. Analyzing health-related quality of life in the EVOLVE trial: the joint impact of treatment and clinical events. Med Decis Making 2016;36(8):965–72.
29. Kerr C, Lloyd A, Rowen D, Maslen T, Brazier J. Androgen deprivation therapy for prostate cancer prevention: what impact do related adverse events have on quality of life? Health Outcomes Res Med 2012;3(3):e169–80.
30. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. A multi-attribute utility function for a comprehensive health status classification system: Health Utilities Mark 2. Med Care 1996;34:702–22.
31. Feeny DH, Furlong WJ, Torrance GW, et al. Multiattribute and single-attribute utility function: the Health Utility Index Mark 3 system. Med Care 2002;40:113–28.
32. Sintonen H, Pekurinen M. A fifteen-dimensional measure of health-related quality of life (15D) and its applications. Quality of Life Assessment: Key Issues in the 1990s. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1993. p. 185–95.
33. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal. London: National Institute for Health and Care Excellence; 2004.
34. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal. London: National Institute for Health and Care Excellence; 2008.
35. Corbett J, d'Angelo C, Gangitano L, Freeman J. Future of health: findings from a survey of stakeholders on the future of health and healthcare in England. The Department of Health. https://www.rand.org/pubs/research_reports/RR2147.html. [Accessed January 8, 2019].
36. National Health Service England. Five year forward view. https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf. [Accessed January 8, 2019].
37. Peasgood T, Karimi M, Brazier J. A Review of Health and Well-Being Measures: What Is Their Conceptual Basis, What Is in Them and How Do They Perform. London: National Institute for Health and Care Excellence; 2016.
38. Haybron D. The Pursuit of Unhappiness. Oxford, UK: Oxford University Press; 2008.
39. Varelius J. Objective explanations of individual well-being. J Happiness Stud 2004;5:73–91.
40. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007;5(1):63.
41. Karimi M, Brazier J, Paisley S. How do individuals value health states? A qualitative investigation. Soc Sci Med 2017;172:80–8.
42. Davies SC. Annual Report of the Chief Medical Officer 2013. Public Health Priorities: Investing in the Evidence. London: Department of Health; 2014.
43. Peasgood T, Brazier J, Mukuria C, Rowen D. A Conceptual Comparison of Well-Being Measures Used in the UK: EEPRU Report 026. London: Economics Evaluation Policy Research Unit; 2014.
44. Mukuria C, Peasgood T, Rowen D, Brazier J. An Empirical Comparison of Wellbeing Measures Used in the UK. In: Policy Research Unit in Economic Evaluation of Health and Care Interventions. EEPRU Research Report 048. Universities of Sheffield and York; 2016.
45. Sen A. Commodities and Capabilities. Amsterdam, The Netherlands: North-Holland; 1985.
46. Coast J, Flynn T, Natarajan L, et al. Valuing the ICECAP capability index for older people. Soc Sci Med 2008;67(5):874–82.
47. Al-Janabi H, Flynn T, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Qual Life Res 2012;21:167–76.
48. Flynn TN, Huynh E, Peters TJ, et al. Scoring the ICECAP-A capability instrument: estimation of a UK general population tariff. Health Econ 2015;24:258–69.
49. Al-Janabi H, Peters T, Brazier J, et al. An investigation of the construct validity of the ICECAP-A capability measure. Qual Life Res 2013;22(7):1831–40.
50. Al-Janabi H, Flynn TN, Peters TJ, Bryan S, Coast J. Test-retest reliability of capability measurement in the UK general population. Health Econ 2014;24(5):625–30.
51. Al-Janabi H. Do capability and functioning differ? A study of U.K. survey responses. Health Econ 2018;27:465–79.
52. Yang Y, Brazier J, Tsuchiya A. Effect of adding a sleep dimension to the EQ-5D descriptive system: a "bolt-on" experiment. Med Decis Making 2014;34(1):42–53.
53. Yang Y, Rowen D, Brazier J, Tsuchiya A, Young T, Longworth L. An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. Value Health 2015;18(1):52–60.
54. Krabbe PF, Stouthard ME, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. J Clin Epidemiol 1999;52(4):293–301.
55. Richardson J, Peacock S, Hawthorne G, Iezzi A, Elsworth G, Day N. Construction of the descriptive system for the assessment of quality of life AQoL-6D utility instrument. Health Qual Life Outcomes 2012;10:38.
56. Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the Assessment of Quality of Life (AQoL-8D) multi attribute utility instrument. Patient 2014;7(1):85–96.
57. Richardson J, Sinha K, Iezzi A, Khan MA. Modelling utility weights for the Assessment of Quality of Life (AQoL) 8D. Qual Life Res 2014;23:2395–404.
58. Extending the QALY. https://scharr.dept.shef.ac.uk/e-qaly/. [Accessed Januarty 8, 2019]
59. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. In: Methods for the Economic Evaluation of Health Care Programmes. 4th ed. Oxford, UK: Oxford University Press; 2015.
60. Shah KK, Lloyd A, Oppe M, Devlin NJ. One-to-one versus group setting for conducting computer-assisted TTO studies: finding from pilot studies in England and the Netherlands. Eur J Health Econ 2013;14(suppl 1):66–73.
61. Mann R, Brazier J, Tsuchiya A. A comparison of patient and general population weightings of EQ-5D dimensions. Health Econ 2009;18:363–72.
62. Little MH, Reitmeir P, Peters A, Leidl R. The impact of differences between patient and general population EQ-5D-3L values on the mean tariff scores of different patient groups. Value Health 2014;17:364–71.
63. Rand-Hendriksen K, Augestad LA, Kristiansen IS, Stavem K. Comparison of hypothetical and experienced EQ-5D valuations: relative weights of five dimensions. Qual Life Res 2012;21(6):1005–12.
64. Karimi M, Brazier J, Paisley S. Are preferences over health states informed? Health Qual Life Outcomes 2017;15:105.
65. General guidelines for economic evaluations from the Pharmaceutical Benefits Board (LFNAR 2003:2). https://www.tlv.se/download/18.2e53241415e842ce95514e9/1510316396792/Guidelines-for-economic-evaluations-LFNAR-2003-2.pdf. [Accessed January 8, 2019].
66. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Making decisions in health care. In: Methods for the Economic Evaluation of Health Care Programmes. Oxford, UK: Oxford University Press; 2015.
67. Brazier J, Rowen D, Karimi M, Peasgood T, Tsuchiya A, Ratcliffe J. Experience-based utility and own health state valuation for a health state classification system: why and how to do it. European J Health Economics 2017. https://doi.org/10.1007/s10198-017-0931-5.
68. Sun S, Chen J, Kind P, Xu L, Zhang Y, Burström K. Experience-based VAS values for EQ-5D-5L health states in a national general population health survey in China. Qual Life Res 2015;24:693–703.
69. McTaggart-Cowan HM, O'Cathain A, Tsuchiya A, Brazier JE. Using mixed methods research to explore the effect of an adaptation exercise on general population valuations of health states. Qual Life Res 2012;21(3):465–73.
70. Robinson S, Bryan S. Does the process of deliberation change individuals' health state valuations? An exploratory study using the person trade-off technique. Value Health 2013;16(5):806–13.
71. Mcintosh CN, Gorber SC, Bernier J, Berthelot JM. Eliciting Canadian population preferences for health states using the Classification and Measurement System of Functional Health (CLAMES). Chronic Dis Can 2007;28:29–41.
72. Stein K, Ratcliffe J, Round A, Milne R, Brazier J. Impact of discussion on preferences elicited in a group setting. Health Qual Life Outcomes 2006;4:22.