CrossMark

REVIEW ARTICLE

How Well Do the Generic Multi-attribute Utility Instruments Incorporate Patient and Public Views Into Their Descriptive Systems?

Katherine J. Stevens

Abstract   Multi-attribute utility instruments (MAUIs) are increasingly being used to generate utility data, which can be used to calculate quality-adjusted life-years (QALYs). These QALY data can then be incorporated into a cost–utility analysis as part of an economic evaluation, to inform health care resource allocation decisions. Many health care decision-making bodies around the world, such as the National Institute for Health and Care Excellence, require the use of generic MAUIs. Recently, there has been a call for greater input of patients into the development of patient-reported outcome measures, and this is now actively encouraged. By incorporating the views of patients, greater validity of an instrument is expected and it is more likely that patients will be able to self-complete the instrument, which is the ideal when obtaining information about a patient's health-related quality of life. This paper examines the stages of MAUI development and the scope for patient and/or public involvement at each stage. The paper then reviews how much the main generic MAUIs have incorporated the views of patients/the public into the development of their descriptive systems at each of these stages, and the implications of this. The review finds that the majority of MAUIs had very little input from patients/ the public. Instead, existing literature and/or the views of experts were used. If we wish to incorporate patient/public views into future development of MAUIs, qualitative methods are recommended.

K. J. Stevens (✉)
Health Economics and Decision Science, School of Health and
Related Research (ScHARR), The University of Sheffield,
Regent Court, 30 Regent Street, Sheffield S1 4DA, UK
e-mail: k.stevens@sheffield.ac.uk

Key Points for Decision Makers

The majority of multi-attribute utility instruments (MAUIs) had very little involvement of patients in the development of their descriptive systems.

The descriptive systems of MAUIs were mostly developed using top-down methods, which made use of existing literature and/or the views of experts in determining what should be included.

If patient or public views are to be incorporated into the development of descriptive systems in the future, qualitative methods are recommended.

1 Introduction

Health care decision-making is increasingly using economic evaluation to help inform the allocation of health care resources. This has been formalized in many countries through the establishment of decision-making bodies, which require submission of evidence on the cost effectiveness of interventions as part of their requirements when deciding whether to recommend interventions. These decision-making bodies often have sets of guidelines that give guidance on the methods to use to provide this evidence. Typically, the preferred form of economic analysis is cost–utility analysis (CUA), with the outcomes measured in quality-adjusted life-years (QALYs). This gives the advantage to the decision maker of being able to compare interventions both within and across clinical areas, as the QALY is a common metric for measuring health outcomes [1]. QALYs are composed of two components: the number

△ Adis

of life-years and a quality-adjustment weighting, ranging on a scale from 0 (equivalent to being dead) to 1 (equivalent to having full health). These two components are combined to calculate the number of QALYs. For example, 8 years of life with a quality weighting of 0.6 would equal 4.8 (8 × 0.6) QALYs. To obtain the quality-adjustment weight, a common approach is to make use of off-the-shelf preference-based measures (PBMs), sometimes called 'multi-attribute utility instruments' (MAUIs).

An MAUI is a measure of health-related quality of life (HRQoL), which consists of two components: a descriptive system and a set of preference weights for all possible health states defined by this descriptive system. The descriptive system is typically constructed from a number of domains of HRQoL, each with a number of response options (levels). A patient will be asked to complete the instrument by answering a series of questions about what level they are at for each of the domains. The answers to these questions categorize the patient into what is termed a 'health state'. Each MAUI has a different number of possible health states that can be defined by its descriptive system. For example, the EQ-5D 3 Level (EQ-5D-3L) [2] consists of five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression), each with three levels, and therefore it has 243 possible health states, whereas the Health Utilities Index Mark 3 (HUI3) has eight dimensions, each with five or six levels, giving a possible 972,000 health states [3]. For each MAUI, there is a pre-existing set of preference weights, which give a utility value to each of the health states defined by the MAUI's descriptive system. These preference weights have been developed by valuing a subset of the total health states in a descriptive system and then modelling to predict a value for every health state. Typically, the health states have been valued using a choice-based technique, such as the standard gamble (SG) or time trade-off (TTO), with members of the adult general population [4].

A generic MAUI is one that is intended to be applicable to all clinical areas and the domains are not specific to any particular health condition. In contrast to this, a condition-specific MAUI (CS MAUI) is specific to a clinical area or condition—for example, asthma or diabetes. CS MAUIs are often called for when generic measures are considered to be insensitive to the health condition being considered [5].

Until recently, research into MAUIs had typically focussed on the valuation side (generating the preference weights) rather than the measurement side (the descriptive system used) [5]—in particular, with the generic measures. As CS MAUIs have generally been developed more recently, when their descriptive systems are being developed it is possible to use methods that have been evolving, such as qualitative techniques to inform item development and Rasch analysis to inform item selection and refinement of the descriptive system [6].

MAUIs take into account people's preferences for the different domains within a measure. Typically, in the generic measures, these preferences are taken from the general population [as recommended by agencies such as the National Institute for Health and Care Excellence (NICE)] [7]. As taxpayers represent society and potential users of the healthcare system, this is often felt to be the most appropriate population [8]. However, in terms of what should be valued, it may not be the case that patients and/or the general public have been as involved in determining what should be included and what is important to measure, and therefore their views on what should be in a descriptive system have not been incorporated.

2 The Importance of Incorporating Patient Views

When an MAUI is used, the ideal situation is that patients self-complete the descriptive system, which defines them as being in a particular health state. The pre-existing preference weights are then applied. By patients self-completing the descriptive system, we ensure that the information about their HRQoL at any point in time comes directly from them. It is known that when instruments are completed by proxy, bias can be introduced, and so, where possible, it is best practice to obtain the information directly from the patient [9].

Given this, it is important that the instruments are amenable to completion directly by the patient and are reliable and valid measures.

Involving patients/users in the development of a descriptive system helps to increase the validity of an instrument—in particular, the content validity and relevance of an instrument, whereby the items and response options included are relevant to the population and the language and terminology used to describe them are appropriate [9]. It will also improve the responsiveness to change of a measure, as it will ensure that only outcomes of relevance to the patient are included [9].

The US Food and Drug Administration (FDA) requires that instruments show evidence that their items have been generated by taking account of the experience and perspective of the patient group [10]. In more recent years, the importance of involving patients and/or lay people in the development of quality-of-life measures has been more widely recognized [11].

3 Methods of Descriptive System Development

There are three key stages in developing an MAUI: first, creation of the descriptive system; second, valuation of a

subset of the health states; and third, modelling to produce a value for every health state [16]. In this paper, the focus is on the development of the descriptive system.

The development of an MAUI has to work within additional constraints, compared with a non-utility-based instrument, as it has to be amenable to valuation of the health states. To be amenable, descriptive systems should ideally contain limited (no more than 7 ± 2) domains and also should ideally have a series of response options that are ordinal and range in levels of increasing (or decreasing) severity/frequency [5].

Within these constraints, the development of a descriptive system can be broken down into a number of stages:

1. Generation of items/domains for potential inclusion.
2. Selection and/or refinement of items.
3. Testing of the descriptive system.

Each of these stages has the potential for patient or public involvement, and these are now considered in turn.

3.1 Generation of Items/Domains for Potential Inclusion

Two contrasting methods of item/domain development have been reported by Stevens and Palfreyman [12]. They are a bottom-up methodology and a top-down methodology. A bottom-up methodology involves working with patients and/or members of the public, and seeks their views on how their life is affected by their particular health problem or condition. This approach typically requires the use of qualitative methods to generate items—for example, through focus groups or individual interviews [9]. Stevens and Palfreyman give two examples of non-MAUIs that have taken this approach: the DEMQol, where both patients and carers were interviewed to identify items [13]; and the Nottingham Health Profile, which used patients and the general public [14]. In contrast, a top-down methodology generally takes information from existing sources, such as the literature, other instruments and health surveys, and uses these to generate a pool of items for potential inclusion in the instrument. Clearly, there is greater scope for patients/the public to be involved through the use of a bottom-up methodology.

3.2 Selection and/or Refinement of Items

This stage involves selecting out the items that are going to be included in the measure. There are a number of ways of doing this. It could be done through the use of qualitative methods, whereby either patients/the public or perhaps clinical/other experts are asked what they think and are asked to select items. Other methods include using psychometric testing, whereby the validity and reliability are tested and items are selected/refined on the basis of these. For example, it may be found that some items are measuring the same thing and therefore one can be removed. Rasch analysis and factor analysis are also methods that can be used [15]. Factor analysis is useful for establishing instrument dimensions, and Rasch analysis is useful for selecting items for inclusion and/or reducing the number of item levels [15]. The majority of these methods offer scope for the inclusion of patients' or the public's views.

3.3 Testing of the Descriptive System

Final testing of a descriptive system is useful, as it can highlight any issues/problems with completion and also offers further scope for refinement before the final descriptive system is decided on. An instrument can be tested on a patient group or general population group, and the practicality, validity and reliability can be measured. This stage again offers scope for patients or the general population to be involved.

4 Review of the Main Generic MAUIs

This paper reviews to what extent existing generic adult MAUIs intended for use in calculating QALYs took account of patient and/or public input in the development of their descriptive systems, and considers the implications of this.

The generic MAUIs have all been developed using different methods. This paper reviews the amount of patient/public input for each MAUI at the three key stages of descriptive system development outlined previously. For each MAUI, the key literature describing its development was identified by searching the literature, reviewing MAUI websites and contacting developers where necessary. There are currently six generic MAUIs that enable the calculation of QALYs [16]: the EQ-5D; the Short Form 6D (SF-6D); the Health Utilities Index Mark 2 (HUI2) and HUI3; the 15D; the Assessment of Quality of Life (AQoL) and the Quality of Well-Being (QWB) scale. Richardson et al. [16] recently reviewed the use of them in the literature. They found that the EQ-5D was by far the most commonly used (63.2 % of studies). The HUI3 was used in 9.8 % of studies, the SF-6D in 8.8 %, the 15D in 6.9 %, the HUI2 in 4.6 %, the AQoL in 4.3 % and the QWB in 2.4 %. All six measures are included in this review. The EQ-5D and AQoL both have versions for children/adolescents (the EQ-5D Youth version [EQ-5D-Y] [2] and the AQoL-6D [17], respectively). In addition, the Child Health Utility 9D (CHU9D) has recently been developed as a new paediatric MAUI [18]. These child instruments are also included in the review.

Table 1 shows a summary of the included generic MAUIs, including the numbers of dimensions and levels, and whether there was patient and/or public involvement at each of the three key stages. The countries of origin and the year when preference weights for each instrument became available are also shown, although instruments typically take a number of years to develop. Each of these MAUIs is then considered in turn.

4.1 EQ-5D-3L

The EQ-5D-3L was developed by a group of researchers across five countries [2]. The researchers used their own expertise, together with a review of other generic HRQoL measures available at the time, to generate a core set of domains, which they felt reflected the most important concerns of patients themselves [19]. The resulting descriptive system consisted of six dimensions, each with two or three levels. There were some experiments with this, and the result was that a number of changes were made and it became a five-dimensional classification system with three levels in each dimension [19]. A large national survey of lay concepts of health was carried out by van Dalen et al. [20] and, following this, there was work done by Gudex [21] to determine whether an additional dimension of energy should be added, but it was found that this was not necessary [22].

As the initial pool of items for consideration was generated from the existing literature, there was no involvement of patients or the public at the first stage; however,

there was some involvement of patients/the public in the subsequent refinement of the instrument from input into the survey of lay concepts of health.

4.2 EQ-5D-5L

The EQ-5D-3L has existed for a number of years and has been widely used and validated. However, there is evidence that it is not always sensitive enough, with just three levels. In response to this, the EQ-5D 5 Level (EQ-5D-5L) was created by a EuroQoL Group Task Force, with the aim of increasing sensitivity and reducing ceiling effects. This was undertaken simultaneously in England and Spain. After discussion by the Task Force, the dimensions were kept the same but it was decided to increase the number of levels to five (on the basis of evidence from the psychometric literature and other sources). Potential labels for the new five levels were generated from a review of existing HRQoL instruments, a review of the literature on response scaling, hand searching of dictionaries and thesauruses, and informal interviews with lay respondents to find out how they described different severities of health problems [23]. The existing structure of the EQ-5D-3L was kept, and the new labels had to fit within this. Pilot work was undertaken to reduce the pool down to a manageable level of 10–12 labels per dimension for consideration. A response scale exercise was done with lay respondents in order to select labels from the pool. Respondents were also asked for their input on whether the new labels suited the dimensions (or not). This exercise produced two versions, which went

Table 1 Summary of generic multi-attribute utility instruments (MAUIs) and whether there was patient/public involvement at each stage of their development

MAUI	Countries of origin	Year preference weights available	Number of dimensions	Number of levels	Patient/public involvement		
					Item/domain generation	Refinement	Testing
EQ-5D	Europe/UK	1995	5	3	No	Yes	No
EQ-5D-5L	Europe/UK	2012[a]	5	5	No	Yes	Yes
EQ-5D-Y	Europe/UK	Not yet available	5	3	No	Yes	Yes
SF-6D	UK/USA	2002	6	4–6	No	No	No
HUI2	Canada	1996	6	4–5	No	Yes	No
HUI3	Canada	2002	8	5–6	No	No	No
15D	Finland	1989	15	4–5	No	Yes	Yes
AQoL-6D	Australia	2004	6	4–6	No	No	Yes
AQoL-8D	Australia	2009	8	4	No	No	No
QWB	USA	1976	3 plus 27 symptoms or problems	2–3	No	No	No
CHU9D	UK	2012	9	5	Yes	Yes	Yes

AQoL Assessment of Quality of Life, *CHU9D* Child Health Utility 9D, *EQ-5D-5L* EQ-5D 5 Level, *EQ-5D-Y* EQ-5D Youth version, *HUI2* - Health Utilities Index Mark 2, *HUI3* Health Utilities Index Mark 3, *QWB* Quality of Well-Being, *SF-6D* Short Form 6D

[a] Interim scoring available

forward for further testing in eight focus groups (four consisting of healthy people and four consisting of people with a chronic illness). This testing aimed to assess the ease of use, comprehension, interpretation and acceptability of the two versions and to decide which was to be the final one [23]. The result was the new five-dimensional, five-level version. Testing has since been carried out to compare it with the EQ-5D-3L in various clinical populations, but not with the purpose of refining the descriptive system.

As with the EQ-5D-3L, there was mainly patient/public input at the second stage of the descriptive system development. The dimensions were kept the same as those in the original three-level version, and a pool of potential labels for the new levels was generated from the existing literature. Patients and the public were used extensively in selection of the levels, both in the response scaling task and then in the subsequent focus groups, which included both well people and chronically ill people.

4.3 EQ-5D-Y

In 2006, a child-friendly version of the EQ-5D—the EQ-5D-Y—was developed.

The descriptive system has the same five dimensions as those in the EQ-5D-3L but uses child-friendly wording ('mobility', 'looking after myself', 'doing usual activities', 'having pain or discomfort', 'feeling worried, sad or unhappy'). There are three levels for each dimension ('no problems', 'some problems', 'a lot of problems') [24]. It is recommend for children aged 8–15 years, although the developers note that for children aged 12–15 years, it is also possible to use the adult EQ-5D, and for children aged 4–7 years, there is a proxy version. The EQ-5D-Y was developed collaboratively by teams from seven countries, who formed a Task Force on behalf of the EuroQoL Group. A decision was made to keep the existing concepts for the dimensions the same as those in the adult version. The Task Force considered evidence from studies where the EQ-5D-3L had been used in younger populations and results from previous qualitative assessments, and used these to alter the instructions, dimension descriptions and wording of response levels where they felt such changes were necessary [25]. The resulting descriptive system was then translated into several languages, and qualitative assessment was undertaken with children. Some versions were subsequently altered to take into account cultural differences, but no changes were felt to be necessary for the English-language version. Subsequently, psychometric testing was undertaken with children in a range of European countries and in South Africa [26]. These results were not used to refine the descriptive system.

Patients/the public were not involved at the initial development stage, as the dimensions and levels for consideration were to be kept the same as those in the existing adult instrument. Children were involved at the refinement stage, as the Task Force took account of previous qualitative assessments with children as to the language used in the descriptive system. The results of this were used (along with consideration of studies where the EQ-5D-3L had been used in younger people) to help determine what wording should be used in the final version in order that children would be able to understand the original concepts [25]. Children were involved at the final stage, which was testing the instrument with the population. Children were also involved in the refinement of some of the translated versions, in order to make sure the instrument was culturally valid.

4.4 EQ-5D Bolt-Ons

There has also been recent work looking at the use of bolt-ons for the EQ-5D. Bolts-ons are dimensions that are added to an instrument to overcome inadequacies in a particular population [27]. Three bolt-ons were developed: hearing, vision and tiredness. The wording and development of these bolt-ons all came from the literature and decisions made by the research team. There was no patient or public involvement.

4.5 SF-6D

The SF-6D was developed by a team at The University of Sheffield. It takes its content from the Short Form 36, a health status measure used widely around the world [28]. The SF-36 takes its content from existing surveys used in health research and subsequent refinement in a series of medical outcome studies. It assesses patients on eight dimensions of HRQoL [29]. The team revised the SF-36 into a six-dimensional health state classification system in order to make it amenable to valuation [30]. The team made use of extensive factor analysis, which had previously been carried out by Ware et al. [31], to inform their selection of dimensions.

Patients or the public were not involved at the first stage of generating potential items to include, as this research took an existing instrument that had already been developed. The public were not involved at the second stage, as the team made use of the results of studies involving factor analysis where the SF-36 had been administered to patients. There was no third stage of testing prior to valuation in this study.

4.6 HUI2 and HUI3

The HUI2 consists of six dimensions (sensation, mobility, emotion, cognition, self-care, pain), each with between

four and five levels, and was designed for use with children. The instrument was originally developed for use in childhood cancer but has subsequently been used as a generic measure [32]. The HUI2 was developed from a review of epidemiological surveys and a review of the literature, which generated a large pool of potential attributes. A sample of 84 child and parent pairs of the same gender living in the same household then rated these items, for selection of attributes for inclusion. The populations were sampled from schools in Hamilton, ON, Canada, and the children were in grade 7 or 8 at school (aged 12/13 years) [33].

The HUI2 did not involve patients or the public at the first stage, as the generation of potential items came from a review of the existing literature. The public were involved at the second stage of selecting items, through the rating work done as child and parent pairs. Whilst children were involved at the rating stage, along with their parents, the investigators made an expert judgment as to what attributes were relevant to the purpose for which the instrument was being developed when they formed the initial list of attributes [34].

The HUI3 was developed from the HUI2 by increasing the number of dimensions to eight (through the separating out of some dimensions and the removal of others) and increasing the number of levels in all dimensions to between five and six [3]. It was designed for use by adults. The development was carried out by the research team who developed the HUI2, and the decisions concerning what dimensions to include were based on experience and evidence from using the HUI2. The aim of the HUI3 was to have full structural independence [3].

4.7 15D

The 15D was based on a review of Finnish health policy documents [16]. It originally had 12 dimensions and then was revised to 15 following feedback from users and health professionals [35]. Two large patient surveys were then carried out, in which respondents were asked to identify dimensions that should be omitted or added. These findings, combined with factor analysis, resulted in the final version [36, 37].

There was no involvement of patients/the public at the first stage; however, patients were involved by giving feedback as users during the process of refinement, and also later, during the process of determining whether dimensions should be added or omitted.

4.8 AQoL-8D

The AQoL descriptive system was developed from a literature review of the existing instruments, focus groups with clinicians and construction surveys [16]. These construction surveys administered large numbers of items to selected patients and the public. Factor analysis and structural equation modelling were then used to select items for inclusion. A survey to determine values for selection of health states was also undertaken with 629 respondents (half patients, half members of the general population). The results of this were used to refine the descriptive system, and the AQoL-8D was produced.

Patients/the public were not directly involved at any of the stages of development. The results of the surveys conducted with patients and the public were used to inform the selection of items for inclusion, but this was not direct involvement.

4.9 AQoL-6D

The AQoL-6D was derived from the existing AQoL-8D adult version. It was designed to increase sensitivity to health state variations close to normal health and to extend the coverage of the AQoL. A subsequent study refined it for application in adolescents by interviewing adolescents and testing the semantics and language [17]. The AQoL-6D has six dimensions (independent living, social relationships, physical senses, psychological wellbeing, pain, coping) [17]. Patients/the public were not involved at the first or second stages, as this was derived from an existing measure. They were involved at the third stage, to some extent, as the semantics and language were tested on them.

4.10 QWB

The QWB consists of three multi-response items and 27 symptom/problem groups, giving a total of 945 states.

It draws its items mainly from an existing US Health Interview Survey, a Social Security Administration Survey and several rehabilitation scales and ongoing community surveys [38].

Patients/the public were not involved at any of the three stages, as the items were taken from existing survey instruments and selected by researchers.

4.11 CHU9D

The CHU9D was developed by Stevens [39–41]. It was developed from the start to be a generic paediatric HRQoL measure for use in economic evaluation. Dimensions were developed through 74 one-to-one interviews with children recruited through schools, who were asked to describe any health problems they had and how these problems impacted on their lives. Children with a wide range of acute and chronic health problems were included in the interviews until saturation was reached. The qualitative interview data

were also used to develop potential response-level wordings for inclusion. Ranking work with 31 children was then undertaken to determine the ordinality of the response-level wordings and also to remove any redundant wordings. A draft descriptive system was then produced, which was then tested on both a general population recruited through schools ($n = 150$) and a clinical population recruited through a hospital, including medical, surgical and day case patients ($n = 98$). The results of this testing then informed the subsequent refinement of the draft instruments to produce the final version for valuation.

The general paediatric population and patients were involved at all three stages of development: at stage 1 (item generation from qualitative interviews); at stage 2 (selection of items for inclusion); and at stage 3 (testing and refinement of the instrument).

5 Discussion

The majority of the generic MAUIs have involved use of a top-down approach in the development of their descriptive systems—that is, the content has been derived from the existing literature, instruments and health surveys. Patient/public involvement, if any, occurred generally at the second and third stages of development when an instrument was being tested. The exception to this is the CHU9D, which was derived using bottom-up methods. Bottom-up methods generally lend themselves better to patient/public involvement, as they typically use focus groups and/or individual interviews for generation of items for consideration, testing of items and refinement of an instrument [12].

This top-down approach mirrors the common approach historically taken within the general HRQoL instrument development literature [12]. This involves generating lists of items drawn from interviews, literature and expert opinion, and then a technique such as factor analysis is used to develop the dimensions. This approach has been followed in the development of the majority of the MAUIs reviewed here. This approach is becoming less common in general HRQoL instrument development because of wider adoption of qualitative techniques and because of the impact of the FDA requirement for the development of patient-related outcome (PRO) measures [42], which requires instruments to show evidence that items have been generated through taking account of the experience and perspective of the patient group [10].

In the MAUI literature, there has also recently been a move towards more use of qualitative methods, particularly in development of new MAUIs and CS MAUIs [6]. The most recently developed generic MAUI, the CHU9D, used qualitative methods and a bottom-up approach, and

involved patients/a general paediatric population at all stages of its development. The recent developments of the EQ-5D-5L and EQ-5D-Y have also made use of qualitative methods, in contrast to the development of the original EQ-5D-3L, which used purely top-down methods. The recent development of bolt-ons for the EQ-5D, however, lacked any patient/public involvement. This is one potential area where future research into bolt-ons such as this could easily incorporate the views of patients/the public through qualitative methods.

The advantages of the bottom-up approach versus the top-down approach are that the final instrument is likely to have more appropriate language and terminology, which should increase the content validity [43]. It is also likely there will be improved responsiveness to change, as it includes outcomes that come directly from patients and that patients feel are relevant [12]. There have also been recent initiatives from health care providers, such as the UK National Health Service (NHS), to focus care and health service research around meeting patient priorities and inclusion within decision-making [44]—again, encouraging patient/public involvement.

Involving children (as both patients and members of the general population) in the development of paediatric measures will also increase the likelihood that the measure is valid and reliable for the intended population. The use of qualitative methods in the development of the CHU9D allows for easy self-completion of the instrument by children, as the language and content were all developed directly from children's input [39]. The AQoL-6D undertook semantic testing of the descriptive system with adolescents to ensure that the measure was understood, and the EQ-5D-Y development also involved some element of input from children as to the appropriate wording that should be used. The AQoL-6D and the EQ-5D-Y were both derived from existing adult measures, and the first stage of development—generating items for potential inclusion—did not involve the public/patients. The disadvantage of this is that these top-down adaptations from adult measures risk missing dimensions pertinent to children and also may include dimensions that are irrelevant to children.

More recent work has seen the development of measures of capability for use in economic evaluation [45, 46]. Whilst these measures cannot be used to calculate QALYs, they still provide valuable information in assessing the benefits of interventions. The descriptive systems of the Investigating Choice Experiments for the Preferences of Older People (ICEPOP) Capability Measure for Adults (ICECAP-A) [45] and the ICEPOP Capability Measure for Older People (ICECAP-O) [46] were developed using qualitative methods, involving in-depth interviews with relevant populations (adults and older people) to identify and refine the attributes that should be included.

Subsequent validation of the measures also made extensive use of qualitative methods to provide evidence on the validity, again involving interviews with the relevant populations [47, 48]. The use of qualitative methods here is in contrast to the vast majority of the development of generic MAUIs used for calculating QALYs to date. The extensive use of qualitative methods in the development of these instruments helps to increase the validity and ensures that patient/user views are incorporated in determining what should be included in the descriptive systems and how it should be defined.

It is clear that if we wish to incorporate patient/public views into the development of descriptive systems, use of qualitative methods at the initial stage is ideal. This allows for the greatest input. Later stages could follow a more mixed-methods approach, such as use of focus groups to reflect on items for inclusion and also quantitative data collected directly from patients, which could also be used to select items for inclusion or to refine a measure. One of the problems in terms of advancing methodology in this area is that previously developed instruments are often poorly reported and it is difficult to find literature documenting the development of their descriptive systems [12]. As well as using traditional focus group and interview techniques, future development of descriptive systems could make use of other qualitative techniques to develop attributes, such as meta-ethnography, which was used in the development of the Carer Experience Scale [49].

It seems unlikely that a new generic MAUI for adults will be developed and used extensively, given the widespread use and validity of the EQ-5D. It seems more likely that use of CS MAUIs will increase, and possibly more bolt-ons to existing MAUIs will be developed [27]. Development of descriptive systems in these areas would be amenable to taking patient/public views into account through use of qualitative methods and, if so, it would only serve to increase the validity of MAUIs.

6 Conclusions

Of all of the generic MAUIs reviewed in this paper, the CHU9D has the most patient/public involvement. Children were involved at each stage of the development of the instrument, and their views about what should be included were taken into account. This measure is unique within this set of instruments in that it used a bottom-up methodology, which allows for greater patient/public input. The other MAUIs were developed using top-down methods, with a mixture of adaptation from existing instruments and/or reviews of the literature/existing measures. The most recent development in the adult MAUIs, the development of the EQ-5D-5L, saw a much greater level of patient/public involvement, and it is likely that use of qualitative methods and patient/public involvement will increase in the future.

References

1. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Introduction to the measurement and valuation of health. In: Measuring and valuing health benefits for economic evaluation; chapter 2. Oxford: Oxford University Press; 2007. p. 7–33.
2. EuroQol. http://www.euroqol.org/. Accessed 23 Sept 2014.
3. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, et al. Multi-attribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. Med Care. 2002;40(2):113–28.
4. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Modelling health state valuation data. In: Measuring and valuing health benefits for economic evaluation; chapter 6. Oxford: Oxford University Press; 2007. p. 139–56.
5. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Describing health. In: Measuring and valuing health benefits for economic evaluation; chapter 4. Oxford: Oxford University Press; 2007. p. 55–76.
6. Brazier JE, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, Barkham M, Ibbotson R. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Health Technol Assess. 2012;16(32):1–114.
7. National Institute for Health and Care Excellence. https://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf. Accessed 23 Sept 2014.
8. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Valuing health. In: Measuring and valuing health benefits for economic evaluation; chapter 5. Oxford: University Press; 2007. p. 83–117.
9. McColl E. Developing questionnaires. In: Fayers P, Hays R, editors. Assessing quality of life in clinical trials. 2nd ed. Oxford: Oxford University Press; 2005. p. 9–23.
10. Food and Drug Administration, US Department of Health and Human Services. Guidance for industry: patient-reported outcome measures Use in medical product development to support labeling claims. 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed 24 Sept 2014.
11. Patrick DL, Burke LB, Gwaltney CJ, Leidy N, Martin ML, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 1—eliciting concepts for a new PRO instrument. Value Health. 2011;14:967–77.
12. Stevens K, Palfreyman S. The use of qualitative methods in developing the descriptive systems of preference-based measures of health-related quality of life for use in economic evaluation. Value Health. 2012;15(8):991–8.
13. Smith SC, Lamping DL, Banerjee S, Harwood R, Foley B, Smith P, et al. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. Health Technol Assess. 2005;9(10):1–93.
14. Hunt SM, McEwen J, McKenna SP. The Nottingham Health Profile user's manual. Manchester: Galen Research and Consultancy; 1981.

15. Young T, et al. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. Qual Life Res. 2009;18(2):253–65.

16. Richardson J, McKie J, Bariola E. Review and critique of health related multi attribute utility instruments. CHE Research Paper Series. Melbourne: Monash University; 2011; 64.

17. Moodie M, Richardson J, Rankin B, Sinha K, Lezzi A. Predicting time trade-off health state valuations of adolescents in four Pacific countries using the AQoL-6D instrument. CHE Research Paper Series. Melbourne: Monash University; 2009; 43.

18. Stevens K. Valuation of the child health utility 9D index. Pharmacoeconomics. 2012;30(8):729–47.

19. Gudex C. The descriptive system of the EuroQol instrument. In: Kind P, Brooks R, Rabin R, editors. EQ-5D concepts and methods: a developmental history. Dordrecht: Springer; 2005.

20. van Dalen H, Williams A, Gudex C. Lay people's evaluations of health: are there variations between different subgroups? J Epidemiol Commun Health. 1994;48:248–53.

21. Gudex C. Are we lacking a dimension of energy in the EuroQol instrument? In: Bjork S, editor. EuroQol conference proceedings, Lund, October 1991. IHE Working Paper 92:2. Lund: Swedish Institute for Health Economics; 1992. pp. 61–72.

22. Brooks R. Descriptive system. In: Brooks R, editor. The EuroQol group after 25 years. Dordrecht: Springer; 2013. p. 37–64.

23. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36.

24. van Reenen M, Janssen B, Oppe M, Kreimeier S, Greiner W. EQ-5D-Y user guide. http://www.euroqol.org/about-eq-5d/publications/user-guide.html. Accessed 12 Nov 2014.

25. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5DY: a child friendly version of the EQ-5D. Qual Life Res. 2010;19:875–86.

26. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burstrom K, Cavrini G, Devlin N, Egmar A, Gusi N, Herd M. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. Qual Life Res. 2010;19(6):887–97.

27. Longworth L, Yang Y, Young T, Mulhern B, Hernandez M, Mukuria C, Rowen D, Tosh J, Tsuchiya A, Evans P, Keetharuth A, Brazier J. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. Health Technol Assess. 2014;18:9.

28. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Methods for obtaining health state values: generic preference-based measures of health and the alternatives. In: Measuring and valuing health benefits for economic evaluation; chapter 8. Oxford: Oxford University Press; 2007. p. 175–239.

29. QualityMetric Incorporated. SF tools. http://www.sf-36.org/tools/SF36.shtml#VERS2. Accessed 13 Mar 2008.

30. Brazier JE, Roberts J, Deverill M. The estimation of a preference-based single index measure for health from the SF-36. J Health Econ. 2002;21:2.

31. Ware JE, Kolinski M, Keller SD. How to score the SF-12 physical and mental health summaries: a user's manual. Boston: The Health Institute, New England Medical Centre; 1995.

32. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang QA. Multi-attribute utility function for a comprehensive health status classification system. Health Utilities Mark 2. Med Care. 1996;34(7):702–22.

33. Cadman D, Goldsmith C, Torrance GW, et al. Development of a health status index for Ontario children. Hamilton: McMaster University; 1986.

34. McCabe C. Estimating preference weights for a paediatric health state classification (HUI2) and a comparison of methods. PhD thesis. Sheffield: The University of Sheffield; May 2003.

35. Sintonen H, Pekurinen M. A generic 15 dimensional measure of health-related quality of life (15D). J Soc Med. 1989;26:85–96.

36. Sintonen H. The 15-D measure of health related quality of life: reliability, validity and sensitivity of its health state descriptive system. Centre for Health Programme Evaluation: Working Paper 41. Melbourne: Monash University; 1994.

37. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med. 2001;33(5):328–36.

38. Kaplan RM, Anderson JP. The Quality of Well-Being scale: rationale for a single quality of life index. In: Walker SR, Rosse RM, editors. Quality of life: assessment and application. London: MTP Press; 1988. p. 51–77.

39. Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric health-related quality-of-life measure. Qual Health Res. 2010;20:340–51.

40. Stevens KJ. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. Qual Life Res. 2009;18(8):1105–13.

41. Stevens KJ. Assessing the performance of a new generic measure of health related quality of life for children and refining it for use in health state valuation. Appl Health Econ Health Policy. 2011;9(3):157–69.

42. Farnik M, Pierzchala W. Instrument development and evaluation for patient related outcome assessments. Patient Relat Outcome Meas. 2012;3:1–7.

43. Brod M, Tesler L, Christensen T. Qualitative research and content validity: developing best practices based on science and experience. Qual Life Res. 2009;18(9):1263–78.

44. UK Department of Health. Equity and excellence: liberating the NHS. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_117353. Accessed 24 Sept 2014.

45. 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.

46. Grewal I, Lewis J, Flynn T, 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.

47. Keeley T, Al-Janabi H, Lorgelly P, Coast J. A qualitative assessment of the content validity of the ICECAP-A and EQ-5D-5L and their appropriateness for use in health research. PLoS One. 2013;8(12):e85287.

48. Horwood J, Sutton E, Coast J. Evaluating the face validity of the ICECAP-O capabilities measure: a "think-aloud" study with hip and knee arthroplasty patients. Appl Res Qual Life. 2014;9(3):667–82.

49. Al-Janabi H, Coast J, Flynn TN. What do people value when they provide unpaid care? A meta-ethnography with interview follow-up. Soc Sci Med. 2008;67:111–21.