



This is a repository copy of *How Well Do the Generic Multi-attribute Utility Instruments Incorporate Patient and Public Views Into Their Descriptive Systems?*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/90861/>

Version: Accepted Version

---

**Article:**

Stevens, K.J. (2015) How Well Do the Generic Multi-attribute Utility Instruments Incorporate Patient and Public Views Into Their Descriptive Systems? Patient - Patient-Centered Outcomes Research. ISSN 1178-1653

<https://doi.org/10.1007/s40271-015-0119-y>

---

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

**How well do the generic multi attribute utility instruments incorporate patient and public views in their descriptive systems?**

Katherine J Stevens, PhD

Health Economics and Decision Science

ScHARR, The University of Sheffield, UK

Address for correspondence:

Dr Katherine Stevens

Health Economics and Decision Science

School of Health and Related Research

The University of Sheffield

Regent Court, 30 Regent Street

Sheffield S1 4DA, UK

Telephone: 0114 222 0841

Fax: 0114 272 4095

Email: [K.Stevens@Sheffield.ac.uk](mailto:K.Stevens@Sheffield.ac.uk)

Short running title: Incorporating patient and public views in multi attribute utility instruments **Abstract**

Multi attribute utility instruments (MAUI) are being increasingly used to generate utility data which can be used to calculate quality adjusted life years (QALYs). This 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 in 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 an 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 was used. If we wish to incorporate patient/public views in future development of MAUIs, then qualitative methods are recommended.

### **Key Points for Decision Makers**

- The majority of multi attribute utility instruments had very little involvement of patients in the development of their descriptive systems
- The descriptive systems of multi attribute utility instruments were mostly developed using top down methods, which made use of existing literature and/or views of experts in determining what should be included
- If patient or public views are to be incorporated in 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 formalised in many countries through the establishment of decision making bodies who 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 which 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 of life years and a quality adjustment weighting, ranging on a scale from 0 (equivalent to being dead) to 1 (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 ( $8 \times 0.6$ ) 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).

A MAUI is a measure of health related quality of life that consists of two components, a descriptive system and a set of preference weights for all the possible health states defined by this descriptive system. The descriptive system is typically constructed from a number of domains of health related quality of life, 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 EQ5D 3L [2] consists of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression), each with 3 levels and therefore has 243 possible health states, whereas the Health Utilities Index 3 (HUI3) has 8 dimensions each with 5 or 6 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 MAUIs 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) and 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 is specific to a clinical area or condition, for example asthma or diabetes. Conditions specific (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 developing their descriptive systems they are able to use methods that have been evolving, such as the use of 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 [7]). As tax payers representing 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 for what should be in a descriptive system have not been incorporated.

## **2. The importance of incorporating patient views**

When a MAUI is used, the ideal situation is that patients self complete the descriptive system which defines them into 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 health related quality of life 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 is appropriate [9]. It will also improve the responsiveness to change of a measure as it will ensure only outcomes of relevance to the patient are included [9].

The U.S Food and Drug Administration (FDA) require that instruments show evidence that their items have been generated through 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 recognised [11].

## **3. Methods of descriptive system development**

There are three key stages to developing an MAUI: firstly, creation of the descriptive system; secondly, valuation of a subset of the health states and thirdly, 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 to additional constraints compared to 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 a limited (no more than  $7\pm 2$ ) domains and also ideally have a series of response options which 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 selected/refined on the basis of these. For example, it may be found that some items are measuring the same thing and thereby 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 for selecting items for including and/or reducing the number of item levels [15]. The majority of these methods offer scope for the inclusion of patient/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 could be tested on a patient or general population group and the practicality, validity and reliability 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 for 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 EuroQoL 5D (EQ5D); the Short Form 6D (SF6D); The Health Utilities Index Mark 2 and Mark 3 (HUI2/3); the 15D; the Australian Quality of Life (AQoL) and the Quality of Well Being (QWB). Richardson et al (REF 2011 paper 64) recently reviewed the use of them in the literature. They found that the EQ5D was by far the most commonly used (63.2% of studies). Use of the HUI3 was 9.8%, the SF6D 8.8%, the 15D 6.9%, the HUI2 4.6%, AQoL 4.3% and the QWB 2.4%. All these six measures are included in the review. The EQ-5D and AQoL both have versions for children/adolescents (EQ-5D Youth (EQ-5D-Y) [2] and AQoL 6D respectively [17]). 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 below shows a summary of the included generic MAUIs, including the number of dimensions, levels and whether there was patient and/or public involvement at each of the three key stages. The country of origin and year preference weights for each instrument became available is also shown, although instruments typically take a number of years to develop. Each of these MAUIs is then considered in turn.

Table 1: Summary of whether there was patient/public involvement at each stage.

MAUI	Country of Origin	Year*	Number of dimensions	Number of levels	Patient/public involvement		
					Item/domain generation	Refinement	Testing
<b>EuroQoL-5D-3L</b>	<b>Europe/UK</b>	<b>1995</b>	<b>5</b>	<b>3</b>	<b>no</b>	<b>Yes</b>	<b>no</b>
<b>EuroQoL-5D-5L</b>	<b>Europe/UK</b>	<b>2012**</b>	<b>5</b>	<b>5</b>	<b>no</b>	<b>yes</b>	<b>yes</b>
<b>EuroQoL-5D-Y</b>	<b>Europe/UK</b>	<b>Preference weights not yet available</b>	<b>5</b>	<b>3</b>	<b>no</b>	<b>yes</b>	<b>yes</b>
<b>Short Form-6D</b>	<b>UK/USA</b>	<b>2002</b>	<b>6</b>	<b>4-6</b>	<b>no</b>	<b>no</b>	<b>no</b>

<b>Health Utilities Index 2</b>	<b>Canada</b>	<b>1996</b>	<b>6</b>	<b>4-5</b>	<b>no</b>	<b>yes</b>	<b>no</b>
<b>Health Utilities Index 3</b>	<b>Canada</b>	<b>2002</b>	<b>8</b>	<b>5-6</b>	<b>no</b>	<b>no</b>	<b>No</b>
<b>15D</b>	<b>Finland</b>	<b>1989</b>	<b>15</b>	<b>4-5</b>	<b>no</b>	<b>yes</b>	<b>yes</b>
<b>Australian Quality of Life-8D</b>	<b>Australia</b>	<b>2009</b>	<b>8</b>	<b>4</b>	<b>no</b>	<b>no</b>	<b>no</b>
<b>Australian Quality of Life-6D</b>	<b>Australia</b>	<b>2004</b>	<b>6</b>	<b>4-6</b>	<b>no</b>	<b>no</b>	<b>yes</b>
<b>Quality of Well Being</b>	<b>USA</b>	<b>1976</b>	<b>3 plus 27 symptoms/problems</b>	<b>2-3</b>	<b>no</b>	<b>no</b>	<b>no</b>
<b>Child Health Utility 9D</b>	<b>UK</b>	<b>2012</b>	<b>9</b>	<b>5</b>	<b>yes</b>	<b>yes</b>	<b>yes</b>

**\*Year preference weights available**

**\*\*Interim scoring available**

### **EQ-5D-3L**

The EQ-5D-3L was developed by a group of researchers across 5 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 6 dimensions each with 2 or 3 levels. There were some experiments with this and the result was that a number of changes were made and it became a 5 dimensional classification system with 3 levels each [19]. A large national survey of lay concepts of health was carried out by Van Dalen [20] and following this, there was work by Gudex [21] to determine whether an additional dimension of energy should be added but it was found that it was not necessary [22].

As the initial pool of items for consideration was generated from existing literature, there was no involvement of patients or the public in the first stage, however there was some involvement of patients/the public in the subsequent refinement of the instrument from the input into the survey of lay concepts of health.

### **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 3 levels. In response to this, a 5 level version 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 5 (based on evidence from psychometric literature and other sources). Potential labels for the new 5 levels were generated from a review of existing health-related quality-of-life 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 dimensions 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 2 versions which went forward for further testing in 8 focus group (4 consisting of healthy people and 4 consisting of people with a chronic illness). This testing aimed to assess the ease of use, comprehension, interpretation, and acceptability of the 2 versions and decide which was to be the final one [23]. The result was the new 5 dimension, 5 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 in the second stage of the descriptive system development. The dimensions were kept the same as the original 3 level version and a pool of potential labels for the new levels was generated from existing literature. Patients and the public were used extensively in the 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.

### **EQ-5D-Y**

In 2006, a child friendly version of the EQ-5D was developed and named the EQ-5D Youth (EQ-5D-Y). The descriptive system has the same 5 dimensions as the EQ-5D-3L, but uses a child friendly wording (mobility, looking after myself, doing usual activities, having pain or discomfort, feeling worried, sad or unhappy). There are 3 levels for each dimension (no problems, some problems, a lot of problems) [24]. It is recommend for children age 8-15 years old, although the developers note that for children age 12-15 years , it is also possible to use the adult EQ5D and for children age 4-7 years old there is a proxy version. The EQ-5D-Y was developed collaboratively by teams from 7 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 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 this to alter the instructions, dimension descriptions and wording of response levels where they felt it was necessary [25]. The resulting descriptive system was then translated into several languages and qualitative assessment 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 South Africa [26]. These results were not used to refine the descriptive system.

Patients/the public were not involved in the initial development stage as the dimensions and levels for consideration were to be kept the same as the existing adult instrument. Children were involved in 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 in 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.

### **EQ-5D Bolt ons**

There has also been recent work looking at the use of bolt ons for the EQ-5D. Bolt 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.

### **SF-6D**

The SF-6D was developed by a team at The University of Sheffield. It takes its content from the Short Form-36, which is a widely used health status measure 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 8 dimensions on health related quality of life [29]. The team revised the SF-36 into a 6 dimensional health state classification system in order to make it amenable to valuation [30]. The team made use of extensive factor analysis that had been carried out by Ware et al [31] already in informing their selection of dimensions.

Patients or the public were not involved in the first stage of generating potential items to include as this research took an existing instrument which was already developed. The public were not involved in the second stage as use was made by the team 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.

### **HUI2/HUI3**

The HUI2 consists of 6 dimensions (sensation, mobility, emotion, cognition, self-care and pain), each with between 4 and 5 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, to select attributes for inclusion. The populations were sampled from schools in Hamilton, Ontario, Canada and the child was in grade 7 or 8 at school (age 12/13 years) [33].

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

The Health Utilities Index 3 (HUI3) was developed from the HUI2 by increasing the number of dimensions to 8 (through the separating out of some dimensions and the removal of others) and increasing the number of levels for all dimensions to between 5 and 6 [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].

### **15D**

The 15D was based on a review of the 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 in the first stage however as part of the process of refinement, patients were involved by giving feedback as users and also later, in determining whether dimensions should be added or omitted.

### **AQoL-8D**

The Assessment of Quality of Life descriptive system was developed from a literature review of 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 was then used to select items for inclusion. A survey to determine values for a selection of health states was also undertaken with 629 respondents (half patients, half 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 in 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.

### **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 AQoL. A subsequent study refined it for application in adolescents by interviewing adolescents and testing the semantics and language [17]. It has six

dimensions, (independent living, social relationships, physical senses, psychological wellbeing, pain and coping) [17]. Patients/the public were not involved in the first or second stages, as this was derived from an existing measure. They were involved in the third stage to some extent as the semantics and language was tested with them.

## **QWB**

The Quality of Well Being (QWB) consists of 3 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 3 stages as items were taken from existing survey instruments and selected by researchers.

## **CHU9D**

The CHU9D was developed by Stevens [39-41]. It was developed from the start to be a generic paediatric health related quality of life 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 was 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 (through schools, n=150) and a clinical population including medical, surgical and day case patients (recruited through a hospital, 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 in all 3 stages of development here, in stage 1, item generation (qualitative interviews), in stage 2, selection of items for inclusion and in stage 3, testing and refinement of the instrument.

## **5. Discussion**

The majority of the generic MAUIs have used a top down approach in the development of their descriptive systems, that is, the content has been derived from existing literature, instruments and health surveys. Patient/public involvement, if any, was generally within 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 lean themselves better towards patient/public involvement as they typically use

focus groups and/or individual interviews when generating items for consideration, testing items and refining 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 used to develop the dimensions. This approach has been followed by the majority of the MAUIs reviewed here. This approach is becoming less common in general HRQoL instrument development due to a wider adoption of qualitative techniques and impact following the FDA requirement for the development of PRO measures [42] which requires that instruments 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 when developing new MAUIs and in particular those for condition specific MAUIs [6]. The most recently developed generic MAUI, the CHU9D, used qualitative methods and a bottom up approach and involved patients/the general paediatric population at all stages of its development. The recent developments of the EQ-5D-5L and EQ-5DY 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 like this could easily incorporate the views of patients/the public through qualitative methods.

The advantages of the bottom up approach over a 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 an improved responsiveness to change, as it includes outcomes directly from patients that they feel are relevant [12]. There are also recent initiatives from health care providers such as the UK NHS, to focus care and health service research around meeting patient priorities and inclusion within decision making [44], again, encouraging the need for patient/public involvement.

Involving children (as both patients and 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 the child as the language and content was all developed directly from children [39]. The AQoL6D undertook semantic testing of the descriptive system with adolescents to ensure that the measure was understood and the EQ-5DY development also involved some element of input from children as to the appropriate wording that should be used. The AQoL-6D and EQ-5DY were both derived from existing adult measure 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 ICECAP A and ICECAP O are measures of capability for use in adults and older people respectively [45, 46]. The descriptive systems of both these measures 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 in the development of descriptive systems then the use of qualitative methods in the initial stage is ideal. This allows for the greatest input. Later stages could follow a more mixed methods approach, such as the 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 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 traditional focus group and interview techniques, future development of descriptive systems could also 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 EQ5D.. It seems more likely that the 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 the use of qualitative methods and if so, would only serve to increase the validity of MAUIs.

## **6. Conclusions**

Of all 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 has seen a much greater level of patient/public involvement and it is likely that the use of qualitative methods and patient/public involvement will increase in the future.

### **Study Funding and Conflicts of Interest**

**There was no funding source for this paper. The author is the developer of The Child Health Utility 9D (CHU9D) which is included in this paper**



## References

1. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A Measuring and Valuing Health Benefits for Economic Evaluation. Oxford University Press. 2007; chapter 2.
2. EuroQol. <http://www.euroqol.org/> accessed 23/09/2014
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. Medical Care. 2002; 40(2): 113-128.
4. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A Measuring and Valuing Health Benefits for Economic Evaluation. Oxford University Press. 2007; chapter 6.
5. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A Measuring and Valuing Health Benefits for Economic Evaluation. Oxford University Press. 2007; chapter 4.
6. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Brazier JE, Rowen D, Mavranouzouli I, Tsuchiya A, Young T, Yang Y, Barkham M, Ibbotson R. Health Technology Assessment. 2012;16(32):1-114.
7. The National Institute for Health and Care Excellence. <https://www.nice.org.uk/> accessed 23/09/2014
8. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A Measuring and Valuing Health Benefits for Economic Evaluation. Oxford University Press. 2007; chapter 5.
9. McColl E. Developing questionnaires. In: Fayers, P, Hays R, editors. Assessing Quality of Life in Clinical Trials. 2nd edition. Oxford University Press. 2005. pp 9-23.
10. U.S. Department of Health and Human Services. Food and drug administration. 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 24/09/2014.
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 in 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 in Health* 2012; 15(8): 991-998.
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 Technology Assessment*. 2005; 9 (10):1-93.
14. Hunt SM, McEwen J, McKenna SP. *The Nottingham Health Profile User's Manual*, 1981.
15. Young T et al. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. *Quality of Life Research*. 2009 Mar;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, Monash University Australia. 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, Monash University Australia. 2009; 43.
18. Stevens K. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics* 2012; 30:8: 729-747
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*. Springer. 2005.
20. van Dalen H, Williams A, Gudex C. Lay people's evaluations of health: are there variations between different subgroups? *Journal of Epidemiology and Community Health*. 1994; 48:248-253.
21. Gudex C. Are we lacking a dimension of energy in the EuroQol Instrument? In: Bjork S editors. *Euroqol Conference Proceedings, Lund, October 1991*. IHE Working Paper 92:2. Lund, Swedish Institute for Health Economics 1992:61-72.
22. Brooks R. Descriptive System. In: Brooks R, editor. *The Euroqol Group after 25 years*. Springer. 2013. Pp. 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). *Quality of Life Research*. 2011;20(10):1727-36.
24. <http://www.euroqol.org/about-eq-5d/publications/user-guide.html>. Accessed 12/11/2014
25. Wille N, Badia X, Bonsel G et al. Development of the EQ-5DY: a child friendly version of the EQ-5D. *Quality of Life Research*. 2010; 19:875-886.

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. *Quality of Life Research*. 2010; 19(6): 887–897.
27. Longworth L, Yang Y, Young T, Mulhern B, Hernandez M, Mukuria C, Rowen D, Tosh J, Tsuchiya A, Evans P, Keetharuth A and 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 Technology Assessment* 2014; 18:9.
28. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford University Press. 2007; chapter 8.
29. Short Form 36. <http://www.sf-36.org/tools/SF36.shtml#VERS2>. Accessed 13/03/2008.
30. Brazier JE, Roberts J, Deverill M. The estimation of a preference-based single index measure for health from the SF-36. *Journal of Health Economics*.2002; 21;2.
31. Ware, JE, Kosinski M, Keller SD. *How to score the SF-12 physical and mental health summaries: a user's manual*. The Health Institute, New England Medical Centre, Boston, MA. 1995.
32. Torrance G W, Feeny D H, Furlong W J, Barr R D, Zhang Y, Wang Q A. 1996. Multi-attribute utility function for a comprehensive health status classification system. *Health Utilities Mark* 2. *Medical Care*. 34(7), 702-722.
33. Cadman D, Goldsmith C, Torrance GW, et al. *Development of a health status index for Ontario children*. McMaster University, Hamilton, Ontario. 1986.
34. McCabe C, May 2003. *Estimating preference weights for a paediatric health state classification (HUI2) and a comparison of methods*. Ph.D. Thesis. The University of Sheffield.
35. Sintonen H, Pekurinen M. A generic 15 dimensional measure of health-related quality of life (15D). *Journal of Social Medicine*. 1989;26:85-96.
36. Sintonen H, Pekurinen M. *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, Monash University Australia. 1994; (41).
37. Sintonen H. *The 15D instrument of health-related quality of life: Properties and applications*. *Annals of Medicine*. 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.pp 51-77.
39. Stevens KJ. Working With Children to Develop Dimensions for a Preference-Based, Generic, Pediatric Health-Related Quality-of-Life Measure. *Qualitative Health Research*. 2010; 20: 340 – 351.
40. Stevens KJ. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Quality of Life Research*. 2009; 18 (8): 1105-1113.
41. Stevens K J. 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. *Applied Health Economics and Health Policy*. 2011; 9(3): 157-169.
42. Farnik M, Pierzchala W. Instrument development and evaluation for patient related outcome assessments. *Patient Related Outcome Measures*. 2012; 3:1-7.
43. Brod M, Tesler L, Christensen T. Qualitative research and content validity: developing best practices based on science and experience. *Quality of Life Research*. 2009;18 (9): 1263-1278.
44. UK Department of Health  
[http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_17353](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_17353) Accessed 24/09/2014.
45. Al-Janabi H, Flynn T, Coast J. (2012) Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research* 21: 167-176
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? *Social Science and Medicine* 2006; 62:1891-1901.
47. Keeley T, Al-Janabi H, Lorgelly P & Coast J. (2013) 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* 8(12)
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. *Applied Research in Quality of Life*.
49. Al-Janabi H, Coast J, Flynn TN. What do people value when they provide unpaid care? A meta-ethnography with interview follow-up. *Social Science & Medicine*. 2008;67:111-121.