This is a repository copy of *Review of valuation methods of preference-based measures of health for economic evaluation in child and adolescent populations: Where are we now and where are we going?*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/154523/

Version: Accepted Version

Article:

This is a post-peer-review, pre-copyedit version of an article published in PharmacoEconomics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s40273-019-00873-7.

Review of valuation methods of preference-based measures of health for economic evaluation in child and adolescent populations: Where are we now and where are we going?

Donna Rowen[1], Oliver Rivero-Arias[2], Nancy Devlin[3], Julie Ratcliffe[4]

[1] School of Health and Related Research (ScHARR), University of Sheffield, UK

[2] National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, UK

[3] Centre for Health Policy, Melbourne School of Population and Global Health, University of Melbourne, Australia

[4] Health and Social Care Economics Group, College of Nursing and Health Sciences, Flinders University, Australia

Correspondence to: Donna Rowen, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, S1 4DA, UK. d.rowen@sheffield.ac.uk
ORCID: 0000-0003-3018-5109

SHORT TITLE: Valuing health for economic evaluation in child and adolescent populations

ABSTRACT

Methods for measuring and valuing health benefits for economic evaluation and health technology assessment in adult populations are well developed. In contrast, methods for assessing interventions for child and adolescent populations lack detailed guidelines, particularly regarding the valuation of health and quality of life in these age groups. This paper critically examines the methodological considerations involved in the valuation of child and adolescent-specific health related quality of life by existing preference-based measures. It also describes the methodological choices made in the valuation of existing generic preference-based measures developed with and/or applied in child and adolescent populations: AHUM, AQoL-6D, CHU9D, EQ-5D-Y, HUI2, HUI3, QWB, 16D and 17D. The approaches used to value existing child and adolescent-specific generic preference-based measures vary considerably. Whilst the choice of whose preferences and which perspective to use is a matter of normative debate and ultimately for decision by reimbursement agencies and policy makers, greater research around these issues would be informative and would enrich these discussions. Research can also inform the other methodological choices required in the valuation of child and adolescent health states. Gaps in research evidence are identified around: the impact of the child described in health state valuation exercises undertaken by adults including possibility of informed preferences; the appropriateness and acceptability of valuation tasks for adolescents in particular tasks involving the state 'dead'; anchoring of adolescent preferences; and generation and use of combined adult and adolescent preferences.

KEY POINTS

- The paper critically examines the methodological considerations involved in the valuation of child and adolescent-specific preference-based measures
- The paper concludes that whilst the choice of whose preferences and which perspective to use is a matter of normative debate, and ultimately for decision by reimbursement agencies and policy makers, greater research around these issues would be informative and enrich these discussions
- Gaps in research evidence are identified, including the anchoring of adolescent preferences for the calculation of quality-adjusted life years (QALYs), and generation and use of combined adult and adolescent preferences

1.0 Introduction

Economic evaluation is increasingly used to inform resource allocation decisions in healthcare, often assessing benefits using quality adjusted life years (QALYs) or disability-adjusted life-years (DALYs). The methodology for assessing interventions and measuring and valuing health benefits in adult populations for economic evaluation is well developed, including detailed guidance from many international agencies (for example [1-2]), and good practice guidelines, for example ISPOR guidance [3-4]. However, the methods for assessing interventions for child and adolescent populations in particular often lack detailed guidelines, or implicitly assume that what is recommended for adults is also what it most appropriate for children and adolescents, despite there being special considerations for children (see for example [5]). One important aspect concerns the valuation of health and/or quality of life for child and adolescents for use in health technology assessment and economic evaluation, in particular to generate QALYs.

The Quality Adjustment weight of the QALY is often generated through application of a preference-based measure accompanied by off-the-shelf utilities. Preference-based measures can be generic or condition-specific, and population-specific, including child and adolescent measures as well as adult measures. A child and adolescent preference-based measure is designed to measure and value the health of children typically aged from around 7 to 17 years (specific target ages vary between measures and there are cases where they are used from age 4, for example CHU9D [6]). Adult measures are generally designed to measure and value the health of adults aged 18 onwards. Some measures are intended for use in children, adolescents and adults (for example HUI2 and HUI3).

Child and adolescent measures differ to adult measures in important ways (see [7] for an overview). For example, child and adolescent measures may regularly need to be proxy-reported by carers as well as self-reported, since younger (aged below 7) and intellectually impaired respondents may be unable to self-report their own health. This raises important considerations around the classification systems used to measure health (for further discussion see [8-11]), for example content that must be appropriate and understandable as the person ages, and suitable for both self-report and proxy report (see for example [12] for issues around self and proxy-reporting).

However, one potential key difference between adult and child and adolescent-specific preference-based measures relates to their *value sets*, the scoring to generate utilities for economic evaluation, and this is the main focus of this paper. The valuation of any preference-based measures requires methodological decisions: whose preferences; which perspective; elicitation technique and mode of administration. If the elicitation technique does not produce scores onto the QALY scale, then methods to anchor onto the 1-0 full health-dead scale required for economic evaluation are needed. Some of these methodological questions differ for child and adolescent measures in comparison to adult measures, and whilst many can be informed by research some are normative and ultimately require a value judgement. Some international agencies also have requirements around the methods used to generate value sets for measures used in health technology assessment (for example see [1]). The issue of comparability with adult utilities and consistency of technology assessments across conditions and populations should be considered. It is also important to consider the comparability to adult utilities within a condition and within a cost-effectiveness model, as QALYs may include impacts on length of life and quality of life from childhood into adulthood.

The aims of this paper are to: 1) to identify current available child and adolescent-specific generic preference-based measures; 2) summarise and provide a critical assessment of the methodological considerations in the valuation of child and adolescent-specific preference-based measures; 3) review the existing literature on generating value sets for child and adolescent preference-based measures; and 4) identify current gaps in research evidence and methods regarding valuation of child and adolescent-specific preference-based measures.

2.0 Child and adolescent-specific generic preference-based measures

A recent review [13] of generic multi-attribute preference-based instruments in paediatric populations has identified and provided an overview of the measures AHUM [14], AQoL-6D [15-16], CHU9D [6,17-25], EQ-5D-Y [26-29], HUI2 [30-31], HUI3 [32] QWB [33], 16D [34] and 17D [35], that are summarised in Table 1.

This review focusses on child and adolescent-specific generic preference-based measures, though there are some examples of child and adolescent-specific condition-specific preference-based measures (for example for dermatitis [36] and asthma [37]) with others also in development.

3.0 Methodological considerations in the valuation of child and adolescent-specific measures

This section outlines the main issues and critically assesses options available to researchers, clinicians and other key stakeholders. Decisions relating to valuation may be influenced by the measure under consideration or by recommendations from reimbursement agencies. However, several important methodological considerations in the valuation of child and adolescent-specific measures can be informed by economic theory and research, for example by identifying good practice through understanding the strengths and limitations of different approaches when applied in different modes of administration, to different populations, using different perspectives. Table 2 presents an overview of key methodological considerations in this context.

3.1 Whose preferences

Utilities that are used to generate the value set for preference-based measures can be elicited from adults (members of the general public, parents, or healthcare professionals), young adults, adolescents and children. The choice of whose preferences is important, and research has shown that different populations provide different preferences [22, 38-39] and arguments can be made to involve the differing perspectives of both child and adult preferences in medical decision making [40].

Adult preferences

Adult preferences can be advocated on the grounds that adults ultimately fund healthcare through taxation, and hence their preferences should be used to determine how healthcare resources are allocated. Value sets for preference-based measures for adults are typically generated using adult general population preferences elicited for hypothetical health states,

and hence it can be argued that the elicitation of adult preferences for child and adolescent-specific preference-based measures can provide comparability in the methodology used to elicit preferences for adults, children and adolescents. However, whilst this provides comparability in the population used to elicit preferences, this does not guarantee comparability in the utilities that are elicited, for example see the discussion below around perspective. The comparability in methods but not the resulting utilities can generate issues for HTA where utilities are modelled over time as the patient ages from childhood through adulthood.

In general it may be argued that adults have a greater understanding than children and adolescents of preference elicitation tasks, that can be cognitively demanding both in terms of understanding the task and being able to make a choice (though this will differ at the individual level) [19]. In addition, it is widely regarded as being ethically acceptable to ask adults to compare health states to being dead, without causing unnecessary distress. However, whilst adults may have greater cognitive understanding of the tasks, they may not understand the child and adolescent health states and their impact, and this is something discussed further below regarding perspective. In addition, previous research has demonstrated that adult preferences can differ to child and adolescent preferences [22], and therefore utilities derived from adult preferences should not be viewed as interchangeable with those derived from children and adolescents.

Child and adolescent preferences

Child and adolescent preferences can be argued for on the grounds that it is children and adolescents who experience the health states, and some institutions regard adolescent views as an important consideration for any assessment of health interventions [41-44]. However, younger children aged around 7-10 years of age are unlikely to fully understand the tasks and unlikely to be able to make a choice. The ability to understand and choose is impacted not only by age but also may be impacted by educational ability, experience of ill health and socio-demographic characteristics, meaning some younger children may be able to undertake these tasks and some older children may be unable to undertake them [45]. The type of elicitation approach adopted, the number of tasks that are presented, framing of questions, the complexity of wording, the number of dimensions in health states, and health state selection for valuation (and comparisons) may also affect the difficulty of the tasks (for an example of how methodological choices may impact see [45]). Presentation and design can be tailored to the population asked to value health states to ease comprehension and reduce difficulty, for example colour coding to highlight differences/similarities, boldening/graying of severity levels, allowing dimensions to vary for only a subset of dimensions within or between tasks (for an example of these types of approaches in an adult population see [46]). Research has found internally valid responses for adolescents valuing hypothetical health states using best-worst scaling and DCE, suggesting that an appropriate selection of task, design, framing and presentation can be used to elicit adolescent preferences where respondents have good understanding and make reasoned choices [22,47, 48]. It should be noted that when applying the best-worst scaling approach in the valuation of CHU9D states worst choices were far less consistent than best choices [22]. This tendency was also evident in the valuation of CHU9D health states using an adult sample but was found to be more prevalent in adolescents. However, such phenomenon was not observed in the valuation of EQ-5D-Y health states in different samples of adolescents and adults in two countries [48]. Other research examining the elicitation of preferences for hypothetical health states has found that children aged 10-17 can complete

best-worst scaling tasks, and children aged 14-17 years can undertake pairwise comparison tasks [45].

Questions have been raised around the acceptability and appropriateness of asking preference elicitation tasks that involve consideration of the state of being dead with adolescents. This raises two issues: first whether adolescents are able to understand and make reasoned choices in questions involving consideration of being dead; and second whether the use of elicitation techniques involving consideration of being dead would cause distress or upset for adolescents and therefore cause concerns for research ethics committees. Some studies have been undertaken involving consideration of being dead with adolescents [49], suggesting that if appropriate design and framing is used these tasks may be appropriate, and further guidance for ethics committees is required for this to be an option pursued in the future as currently there is little guidance on these issues.

The inability of younger children to value health states raises the issue of whether it is more acceptable for adolescents than adults to value health states experienced by young children. Either argument can be made around whose preferences should be used to value health states for young children, but for these children their own preferences cannot be taken into account, meaning that it is a normative decision around whose preferences to use.

Hypothetical preferences, experience-based preferences or patient preferences

Preferences can be elicited for hypothetical health states, where people imagine health states, termed hypothetical general population preferences, and these could be provided by general population adults or adolescents. However it is possible to ask adolescent patients in ill health, to value hypothetical health states, and this is referred to as *patient preferences*. Another alternative is to ask adolescent patients in ill health to value their own health state, which generates *experience-based preferences.* An experience-based value set has been estimated for the EQ-5D-Y in Canada, which estimates a regression with own VAS as the dependent variable and the EQ-5D-Y classification system as the independent variables from respondents aged primarily 10 to 11 years [50], though note that this uses a 1-0 scale where 1 equals best state and 0 equals worst state. There are theoretical and practical arguments around the advantages and limitations of both experience-based preferences [51] and patient preferences [52] that have been discussed for adult utilities, and many of these arguments are likely to apply for child and adolescent-specific preference-based measures.

Combined preferences

One option is to extend the definition of the general population to include adolescents when valuing health states, to generate a value set that combines both adult and adolescent preferences together. Since adolescent and adult preferences may differ, sampling strategies around age and gender would need to be carefully considered to achieve an appropriate sample. Alternatively, both adolescent and adult value sets could be generated and both used to inform analyses (for an analogous argument for general population and patient preferences see [52]; this is also relevant for the 2nd Washington Panel on Cost Effectiveness [53]). However, careful consideration of the appropriate elicitation technique and perspective would be required.

Informed preferences

Informed preferences have been used in the elicitation of adult utilities as a way of obtaining preferences from the general population that are more informed about what it is like to live in ill health using information from patients experiencing health states [54]. This technique could be used to provide adolescents more information about what it is like to experience ill health, since their experiences of ill health may be limited; or could be used to provide adults more information about how ill health impacts on children and adolescents when they are valuing health states in the context of imagining what it is like for a child (see section on perspective below). For example, information that is provided could involve child and adolescent experience-based preferences (see section above), or child and adolescent patient preferences (see section above). This is not something that we are aware has been undertaken in the literature, and further research may be worthwhile.

3.2 Perspective

In hypothetical health state valuation tasks participants are asked to imagine someone in a health state, and indicate how good or bad the health state is for that person. The term perspective' is used to indicate who the person is that they are imagining is experiencing the health state, for example, the person could be themselves, a child, or another adult. The elicitation of preferences from adolescents would usually involve valuation from their own perspective, where they are imagining that they are experiencing the health state. However, adolescents could be asked to value health states experienced by someone else, an 'other' perspective, but this is likely to be more cognitively challenging.

The elicitation of preferences from adults can involve multiple different perspectives: own health as an adult; health state for themselves as a child; health state in the context of a child at a specified age; health state for another adult.

Own perspective for adults can be argued for on the basis that the adult is under a 'veil of ignorance' where they do not know who is experiencing the health state, and hence the value they provide is not influenced by any views around children or child health. It can be argued that this provides comparability with the methodology used to elicit hypothetical adult preferences for adult health states. In addition, if child health is valued more highly by society than adult health, this can be taken into account in the resource allocation process using, for example, QALY weighting or deliberation, where there is no risk of double counting as the utilities are not in any way influenced by participants preferences around child health. However, the classification system of child and adolescent-specific preference-based measures may involve terms that are inappropriate for adults, for example CHU9D mentions homework and schoolwork in one dimension (though there is an adult version that instead refers to work [22]). If these were to remain in their original wording this would likely cause confusion and a lack of engagement, and would lead participants to the view that they are being asked to imagine themselves as a child. Alternatively some dimensions can be reworded, meaning that the definition of this dimension is not analogous to the aspect of health-related quality of life that the child or adolescent are reporting using the classification system, creating a discrepancy in what is valued in the value set and what is reported using the measure [24,55]. Another example is daily routine, where although the dimension would not be reworded in a valuation task, a child's daily routine will differ to the daily routine adults imagine for themselves.

Adults could be asked to imagine the health state in the context of a child of a specified age, where often a 10 year old child is specified though this could be any age. However, the child that the participant imagines may matter: whether for example it is their own child,

grandchild, child they have strong feelings about, or a child they do not know. These preferences may be influenced by participants' views about children and child health, meaning that the elicited preferences may include both how good or bad the health state is, but also how good or bad it is that the child they are imagining is in this state of ill health. It can be argued that the use of these preferences to inform policy, for example to generate QALYs for HTA, should take this into consideration, since any QALY weighting or deliberation that gives a higher weight to child health relative to adult health may be double counting. There is also the issue around the age of the child that adults should be asked to imagine. There is a possibility that the age of the child participants are asked to imagine impacts on preferences, and this is an area currently under research.

Adults could be asked to imagine the health states for themselves as a child, but this is prone to recall bias, as they will not be able to accurately recall what it was like to be a child. Their preferences may also be influenced by views around child health, their childhood and their experiences as a parent/guardian if they have children.

3.3 Elicitation technique and mode of administration

Table 2 outlines the different preference elicitation techniques that can be used in studies eliciting valuations from adolescents and adult populations: best-worst scaling; discrete choice experiment (DCE); ranking; rating scale/visual analogue scale (VAS); DCE with duration; time trade-off; standard gamble. Each of these elicitation techniques is theoretically plausible for use with adolescents and adults, though there may be ethical and practical concerns around the acceptability and appropriateness of use of some of these techniques in adolescents.

Best-worst scaling, ranking and discrete choice experiment are all ordinal techniques that provide relative weightings of dimensions and severity levels, and are all generally considered as being easy to understand. These methods do not *require* any consideration of being dead, and so are considered ethically acceptable and appropriate for use in adolescents. However, all these methods only generate anchored preferences onto the 1-0 full health-dead scale if there is mention of being dead and duration of health states. For example in DCE with duration this is achieved by including duration as an additional attribute [56-57] (see section below for methods of anchoring). VAS tasks do not require inclusion of dead as a state in the task, but if dead is included the generated preferences can be directly anchored onto the 1-0 full-health-dead scale.

Best-worst scaling has been criticised in the literature when used to value health states in adults, and a small number of studies have found that the preferences it generates differ to other elicitation techniques [58-59], though further research studies examining this are recommended. DCE may be cognitively challenging particularly where there are several dimensions of health and these vary across the profiles within a choice set. Ranking over a large number of health states can become laborious and time consuming with a large amount of reading and recall of the other states each state is being ranked alongside. VAS has been criticised in the literature as it does not involve sacrifice or opportunity cost meaning that it may not accurately reflect the value of a health state, though there is no consensus on this issue [60]. Participants have been found to spread the set of states (or dimensions) they are valuing across the scale, meaning that the value of states can be impacted by the states they are valued alongside, avoid the ends of the scale, and display a tendency to prefer numbers ending in 5 or 0 (50, 55, 60) [61] though digit preferences can

also be observed using other cardinal elicitation techniques. However in VAS valuation studies the impact of these may be reduced through careful design.

Time trade-off, standard gamble and DCE with duration are cardinal techniques that generate utilities on the 1-0 full health-dead scale. These techniques involve imagining being dead, and as discussed above, questions have been raised around the acceptability and appropriateness of asking adolescents to complete these tasks. An option to remove consideration of dead is chained time trade-off or chained standard gamble, where an impaired health state is valued relative to a worse health state, with no mention of dead. The utility for the impaired health state is then anchored on the 1-0 full health-dead scale using the utility for the worse health state which is elicited using standard time trade-off or standard gamble, and these utilities could be elicited from adults (see the section below on anchoring for discussion of some of the issues this raises). DCE with duration has not to our knowledge been undertaken with adolescents, and may be too cognitively challenging since it involves both trading between length of life and health and simultaneously considering multiple profiles of health. DCE with duration will not generate appropriate responses if respondents do not trade between length of life and health, and hence this should be established prior to use of this technique. Standard gamble involves consideration of risk, and adolescents may have different attitudes to risk than adults which could impact on elicited standard gamble preferences. Time trade-off is often used to generate value sets for adult preference-based measures, and the use of this technique may provide greater comparability of methods used to generate adult value sets for these measures, provided this can be used appropriately given the methodological choices of whose values and which perspective to use in the valuation survey.

The choice of perspective combined with technique should be carefully considered since this can impact on preferences. Research using visual analogue scale has shown that adults valuing health states from the perspective of a child of a specified age can generate lower utilities than adults valuing health states for themselves [62]. However the reverse has been found using time trade-off where participants trade between health and length of life to indicate their preferences for health states, where utilities elicited using an adult own health perspective can be lower that utilities elicited considering the perspective of a child [55] i.e. adults were less willing to trade off length of life for children. Potentially this may also occur for DCE with duration and standard gamble due to the risk of death. Potentially this may occur because participants are more unwilling to state that a child should die sooner than to state that they themselves should die sooner.

Valuation studies for adult preference-based measures have been conducted using online surveys, computer assisted personal interviews (CAPI), face-to-face interviews and hall tests across a range of different elicitation techniques, and table 2 highlights the use of classroom tests for adolescents. Appropriate design, framing and presentation can make a difference not only around the appropriateness of the task but also around the appropriateness of the mode of administration used to elicit preferences, and careful piloting is recommended.

3.4 Anchoring

Best-worst scaling, ranking and discrete choice experiment do not automatically provide utilities that are anchored onto the 1-0 full health-dead scale (see discussion above about the protocols that enable these methods to directly generate utilities on the 1-0 scale). This presents the key challenge of how to anchor these utilities onto the 1-0 full health-dead scale. Anchoring requires the use of utilities for the classification system that are anchored

onto the 1-0 full health-dead scale, and these could be elicited using time trade-off, standard gamble or DCE with duration.

Possible methods for anchoring include: mapping the ordinal preferences via regression analysis to cardinal utilities; rescaling using cardinal utilities for worst state/small numbers of states; and a hybrid model simultaneously modelling both ordinal and cardinal data [61] (to our knowledge the hybrid model has not been currently applied to the valuation of child health states). Both the mapping method and hybrid model have been found to be more accurate at predicting time trade-off utilities when mapped from DCE preferences than the rescaling method [63]. The mapping method approach will simply anchor the ordinal preferences, whereas the hybrid model will simultaneously consider both the ordinal and cardinal data and hence will produce utilities that combine the data. The selection of which method to apply may therefore depend upon whether the researcher or policy maker aims to generate combined preferences. For example, in the case of the elicitation of adolescent preferences, the mapping approach may be selected if adult preferences are obtained solely for the purpose of anchoring, rather than to generate combined value sets. The anchoring of utilities for child and adolescent preference-based measures in particular is an important area that has been under-researched and has not been fully debated to date.

4.0 Review of methods used to generate value sets for child and adolescent-specific generic preference-based measures

Table 3 provides a summary of the value set methodologies of child and adolescent-specific generic preference-based measures. Note that AHUM, CHU9D, EQ-5D-Y, 16D and 17D are the only measures intended only for use in children and/or adolescents, all of the other measures are also appropriate (and derived) for use in adults. For a more detailed overview of each valuation study of each measure see [13].

There is no consensus in the methodology used in the valuation across the measures, for the case of the CHU9D, HUI2 and EQ-5D-Y for valuations in different countries, and for the 16D and 17D across a suite of measures.

4.1 Whose preferences

Adolescent preferences are solely used to generate value sets for AQoL-6D and 16D; adolescent preferences anchored using young adult preferences are used to generate CHU9D value sets in Australia and China; adult general population preferences are used to generate value sets for AHUM, CHU9D in the Netherlands and UK, EQ-5D-Y in the US, HUI2 in the UK, HUI3 and QWB; parent preferences are used to generate value sets for HUI2 in Canada and 17D.

Samples

Sample size ranges from 115 for AQoL-7D to 4155 for EQ-5D-Y. Some differences in sample size would be expected due to differences in the elicitation technique and mode of administration as well as the choice of modelling and selection of health states for valuation. However, three samples are below 200 (HUI2 valued in Canada and UK and 17D). Sample representativeness in terms of the approach used to ensure that the sample is representative of the population varies across the studies. The 16D and 17D studies recruiting children and adolescents recruited both school children and patients, whilst CHU9D in China recruited only school children to form the adolescent sample and CHU9D in

Australia recruited a community-based sample via parents. Most of the studies involving adult general population aimed to obtain national representativeness, with the notable exceptions that AHUM recruited participants both by word of mouth and an existing panel of potential participants and the sampling method is not specified for the AQoL-6D valuation. Three of the studies were published in 1996 (HUI2 Canada, 16D, 17D), one study in 2002 (HUI3), one study in 2005 (HUI2 UK), one study in 2008 (QWB) and the remainder 2010 onwards. However, many of the valuation studies may have been conducted many years prior to publication, for example the HUI3 valuation was undertaken in 1994.

4.2 Perspective

Adolescent preferences are elicited using their own perspective. Valuation studies where parent preferences are elicited use the perspective of a 10 year old child for HUI2 in Canada and a child aged 8-11 for 17D. Valuation studies where adult general population preferences are elicited use their own perspective for AHUM, CHU9D in the UK and the Netherlands, and HUI3, and use the perspective of a 10 year old child for HUI2 in the UK and EQ-5D-Y in the US.

4.3. Elicitation technique and mode of administration

There is considerable variation in the preference elicitation tasks used, with AHUM and AQoL-6D using time trade-off, CHU9D using different techniques in different countries with best-worst scaling and time trade-off, discrete choice experiment with duration, and standard gamble, HUI2 and HUI3 using standard gamble and visual analogue scale, EQ-5D-Y using discrete choice experiment with duration and QWB, 16D and 17D using a VAS. Adolescent preferences are elicited in a classroom setting and online survey, and adult preferences are elicited using face-to-face interviews and online surveys.

4.4 Anchoring

Most studies employ techniques that are directly elicited using conventional valuation approaches on the 1-0 full health-dead scale, with the exception of CHU9D in Australia and China. Both HUI2 value sets and the HUI3 value set apply multi-attribute utility theory to combine standard gamble and visual analogue scale data.

5.0 Discussion

This paper has critically examined the methodological considerations involved in the valuation of child and adolescent-specific measures, with reference to the methodological choices made to date in the valuation of child and adolescent-specific generic preference-based measures. The approaches used to value existing child and adolescent-specific generic preference-based measures are varied, with no commonality across the measures, or for some measures, within the choices made to value the measure in different countries. Sample size for some studies is small (HUI2 Canada [30] and UK [31] and 16D [34]) given the size of the classification systems and the intended use of the valuation study to generate value sets for use to inform policy. Some of the value sets were published over 20 years ago [30,34,35] (and the valuation studies underpinning these are likely to have been undertaken years earlier), and preferences may have changed over this time and there have been

methodological advances in the health valuation literature. The methodological choices made to generate existing value sets indicate both what has been done and what is possible, yet there are many possibilities for future research around both what else could be done, and scope for recommendations around good practice. Whilst many of the considerations are normative, meaning it is perfectly acceptable and expected that a range of approaches are used to generate existing value sets, both economic theory and empirical research can be used to generate good practice guidelines and maximise the quality of research in this area.

Currently there is limited guidance from international agencies around how to generate QALYs and hence utilities for use in health technology assessment of interventions affecting young populations. For example, whilst the NICE Methods Guide is prescriptive for the methods that should be used to generate utilities for adults, limited guidance is given around how to generate, source and model utilities for child and adolescent-specific states. Recent reviews have found that child and adolescent-specific preference-based measures have been used only a handful of times in health technology assessments covering children and adolescents submitted to NICE [64], and published cost-utility analyses for child and adolescent populations [65], and that a large range of diverse methods are used to generate published utilities for children and adolescents [66-68].

The limited use of child and adolescent-specific preference-based measures to reflect the health and quality of life of children in health technology assessment is concerning, since we are not aware of an evidence base demonstrating that adult preference-based measures (such as EQ-5D-3L) appropriately and accurately capture the health and quality of life of children and adolescents. Evidence is required to examine the representativeness of adult measures self-completed by adults for their own health as a proxy for capturing the health of a child with the same condition, since this type of evidence has been used to inform health technology assessments [64]. In addition evidence demonstrating head-to-head comparisons of adult preference-based measures and child and adolescent specific preference-based measures would enable greater understanding of the impact of using an adult or child and adolescent-specific measure to measure the health of a child and adolescent.

The issue of comparability and consistency of utilities generated by child and adolescent-specific preference-based measures and utilities generated by adult measures is important, since for health technology assessment utilities are modelled over time as the patient ages from childhood through adulthood. Whilst it can be argued that the use of comparable valuation methodology for different preference-based measures can be used to ensure consistency when considering evidence generated using different measures (see for example [69] for this argument around condition-specific and generic preference-based measures), this does not ensure comparability in the actual utilities that are used. This is important if utility changes as the patient ages due to a change in preference-based measure or from proxy to self-reporting despite no change in health.

Use of measures such as HUI2 and HUI3 that are appropriate for use across children, adolescents and adults have the advantage of consistency and comparability of utilities across all ages of patients. The combination of utilities generated using EQ-5D-3L and EQ-5D-Y can also arguably provide some consistency in terms of the domains of health assessed, if it is appropriate to assume that domain content is the correct criteria of consistency. The CHU9D measure does have an adult version, but use of this measure in adults can be questioned since the content of the classification system was developed with children aged 7 to 11 [6,17-18].

It is unclear why child and adolescent-specific preference-based measures have not been used to a larger extent to generate utilities for child and adolescent-specific states. Potentially this could be for many reasons that are not mutually exclusive, including: a concern around the psychometric performance of these measures; limited uptake of child and adolescent preference-based measures in trials or other studies used to generate data for use in health technology assessment; concern around the appropriateness of existing value set utilities, methodology or in the case of EQ-5D-Y lack of a value set; concern around the scope and focus of these measures and whether they capture all important outcomes for health and social care; or a concern around the use of these measures alongside adult utilities generated using an adult generic preference-based measure and how to combine these utilities. Another potential reason may be that there is less emphasis placed on cost-effectiveness when making resource allocation decisions for children and adolescents. In addition the absence of recommendations for the use of child and adolescent-specific measures in guidelines by international agencies is likely to be an important factor contributing to their limited usage and developing these recommendations would encourage greater usage of these measures and would be an important step forward.

In the future there are likely to be more child and adolescent-specific generic preference-based measures, since there are existing child and adolescent-specific generic measures currently undergoing valuation in order to make them preference based, including PedsQL [70] (note there is also an adult version), and other measures that are amenable to valuation and that may be valued in the future, for example PROMIS [71]. At the time of preparation of this manuscript, the EuroQol Group is developing an international valuation protocol for the development of country-specific EQ-5D-Y value sets. This protocol has been informed by completed or in-progress studies funded by the EuroQol Group that has investigated: 1) whether current EQ-5D-3L value sets can be appropriately used with EQ-5D-Y health states [55,62]; 2) the development of a latent scale value set in the UK using adults and adolescent samples [39,47]; 3) the evaluation of different anchoring alternatives to latent scale value sets from discrete choice experiments [72]; and 4) the impact of using different perspectives when completing DCE with duration tasks to estimate an EQ-5D-Y value set.

The issue of measuring and valuing benefits for children and adolescents cannot be considered in isolation, since the impact of ill-health reaches wider than the child or adolescent to other family members. There is an important literature around the use of a family perspective in economic evaluation for children and adolescents to include spillover effects and also around joint utility estimation [5,8-9,73-76] and this is an area that deserves consideration by international agencies when they consider whether to make special recommendations around measuring and valuing health benefits in child and adolescent populations for economic evaluation.

The topic of this paper can be discussed in relation to welfarism and extra-welfarism. Welfarism has a clear theoretical position on whose preferences count in social choices, though the literature as far as we are aware does not have special considerations for children or adolescents. However, QALYs and cost-effectiveness analysis are grounded in extra-welfarism, and extra welfarism offers no such guidance. This means that the normative issues that we discuss in the paper require quite strong value judgements.

This review has examined the methodology around valuation of measures aimed at measuring and valuing the health and quality of life of children and adolescents aged 5 years and above. There are added complications of generating utilities for children below age 4, where none of the generic preference-based measures are recommended for use, meaning that there is little scope for the measurement and valuation of health and quality of

life for children of this age as reported by carers/parents. There is a quality of life measure for infants and toddlers [77-79], the infant and toddler quality of life questionnaire (ITQOL), but it is not preference-based. Valuation for health and quality of life for this age group would also present new challenges, since what is within a normal developmental range widely varies within the 0-4 age range, and any generated utilities may need to capture impairment in comparison to the normal developmental range, rather than the normal developmental stage. For example, a newborn baby will not be able to walk or talk, but arguably should not have a utility decrement reflecting their inability to walk or talk; whereas a 4 year old within the normal developmental range would walk and talk, and any impairment would likely be associated with a utility decrement. Therefore, whilst QALYs can be used to capture health benefits for children aged below 4, the estimation of utilities to generate QALYs is far from straightforward.

Conclusions

This paper has summarised and critically assessed the methodological considerations involved in the valuation of child and adolescent-specific measures; and reviewed the methodological choices made to generate value sets for child and adolescent generic preference-based measures. This paper has also identified gaps in research evidence and methods regarding valuation of child and adolescent health states, in particular around:

> *Whose preferences*: the collection of experience-based utilities; the elicitation of patient preferences; possibilities for the combination of utilities elicited from adults and adolescents; whether there is a role and how to elicit informed preferences where child and adolescent experience can be used to inform elicitation tasks undertaken by adolescents or adults;

> *Perspective*: whether the age and description of the child impacts on preferences elicited by adults valuing from the perspective of the child;

> *Elicitation technique*: greater guidance around when consideration of being dead is both appropriate and acceptable for inclusion in tasks completed by adolescents; and how to ensure tasks are designed and framed appropriately for adolescents;

> *Anchoring*: greater exploration of the anchoring of adolescent preferences using techniques applied in the valuation of adult preference-based measures.

Valuation of child and adolescent-specific preference-based measures is a challenging area of research that warrants further empirical evidence to inform best practice guidelines. Many international agencies will have a view on this, and other stakeholders including the general public, carers/parents and patients, and their views as well as economic theory will ultimately determine both the research agenda and what methodology is selected.

Compliance with Ethical Standards

UK and contributing to the development of an international protocol for valuation of EQ-5D-Y.

References

[1] National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: NICE, 2013

[2] PBAC (Pharmaceutical Benefits Advisory Committee). Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee. Australia: Australian Government Department of Health, 2013.

[3] Brazier J, Ara R, Azzabi I, et al. Identification, review, and use of health state utilities in cost-effectiveness models: an ISPOR Good Practices for Outcomes Research Task Force Report. Value Health 2019;22(3):267 – 275.

[4] Wolowacz SE, Briggs A, Belozeroff V, et al. Estimating Health-State Utility for Economic Models in Clinical Studies: an ISPOR Good Research Practices Task Force Report. Value Health. 2016; 19(6):704-719.

[5] Ungar W. Economic Evaluation in Child Health. 2009. Oxford University Press: Oxford.

[6] 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

[7] Matza LS, Patrick D, Riley AW et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: Report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. Value Health 2013:16;461-79.

[8] Ungar WJ. Challenges in Health State Valuation in Paediatric Economic Evaluation: Are QALYs Contraindicated? PharmacoEconomics, 2011; 29(8):641-652

[9] Prosser LA, Hammitt JK, Keren R. Measuring Health Preferences for Use in Cost-Utility and Cost-Benefit Analyses of Interventions in Children: Theoretical and Methodological Considerations. Pharmacoeconomics 2007; 25 (9): 713-726

[10] Petrou S. Methodological issues raised by preference-based approaches to measuring the health status of children. Health Economics, 2003; 12:697-702.

[11] De Civita M, Regier D, Alamgir AH, Anis AH, FitzGerald MJ, Marra CA. Evaluating Health-Related Quality-of-Life Studies in Paediatric Populations Some Conceptual, Methodological and Developmental Considerations and Recent Applications. Pharmacoeconomics 2005; 23 (7): 659-685

[12] Pickard S, Knight SJ. Proxy Evaluation of Health-Related Quality of Life: A Conceptual Framework for Understanding Multiple Proxy Perspectives. Medical Care, 2005; 43(5): 493–499.

[13] Chen G, Ratcliffe J (2015) A Review of the Development and Application of Generic Multi-Attribute Utility Instruments for Paediatric Populations. PharmacoEconomics 33:1013–1028.

[14] Beusterien KM, Yeung JE, Pang F, Brazier J. Development of the multi-attribute Adolescent Health Utility Measure (AHUM). Health and Quality of Life Outcomes, 2012; 10:102.

[15] Richardson J, Day N, Peacock S, et al. Measurement of the quality of life for economic evaluation and the Assessment of Quality of Life (AQoL) Mark 2 instrument. Aust Econ Hist Review 2004;37:62–88.

[16] Moodie M, Richardson J, Rankin B, et al. Predicting time trade-off health state valuations of adolescents in four pacific countries using the AQoL-6D instrument. Value in Health 2010; 13: 1014-27

[17] Stevens, K J. 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

[18] Stevens, K J. Working With Children to Develop Dimensions for a Preference-Based, Generic, Pediatric Health-Related Quality-of-Life Measure. Qualitative Health Research. 2010; vol. 20: 340 - 351

[19] Ratcliffe J, Couzner L, Flynn T, Sawyer M, Stevens K, Brazier J, Burgess L. Valuing Child Health Utility 9D health states with a young adolescent sample: a feasibility study to compare Best-Worst Discrete Choice Experiment, Standard Gamble and Time Trade Off methods. Applied Health Economics and Health Policy. 2011;9 (1)15-27

[20] Ratcliffe J, Flynn T, Terlich F, Brazier J, Stevens K, Sawyer M. Developing adolescent specific health state values for economic evaluation: an application of profile case best worst scaling to the Child Health Utility-9D. Pharmacoeconomics 2012; 30:713-27.

[21] Ratcliffe J, Chen G, Stevens K, Bradley S, Couzner L, Brazier J, Sawyer M, Roberts R, Huynh E, Flynn T. Valuing Child Health Utility 9D Health States with Young Adults: Insights from A Time Trade Off Study. Applied Health Economics and Health Policy, 2015; 13:485-492

[22] Ratcliffe J, Huynh E, Stevens K, Brazier J, Sawyer M, Flynn, T. Nothing about us without us? A comparison of adolescent and adult health-state values for the child health utility-9D using profile case best-worst scaling. Health Economics, 2016; 25: 486-496

[23] Chen G, Xu F, Huynh E, Wang Zhiyong, Stevens K, Ratcliffe J. Scoring the Child Health Utility 9D instrument: estimation of a Chinese child and adolescent-specific tariff. Quality of Life Research, 2019; 28:163-176.

[24] Rowen DL, Mulhern B, Stevens K, Vermaire E. Estimating a Dutch value set for the paediatric preference-based CHU9D using a discrete choice experiment with duration. Value in Health, 2018; 21:1234-1242.

[25] Stevens K. Valuation of the Child Health Utility 9D Index. Pharmacoeconomics 2012; 30:8: 729-747.

[26] Devlin N and Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. Applied Health Economics and Health Policy, 2017; 15:127-137.

[27] Ravens-Sieberer U, Wille N, Badia X, et al. Feasibilty, reliability and validity of the EQ-5D-Y: results from a multinational study. Qual Life Res 2010;19:87-897.

[28] Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual Life Res 2010;19:875-886.

[29] Craig B, Greiner W, Brown DS, Reeve BB. Valuation of child-related quality of life in the United States. Health Economics, 2016; 25:768-777.

[30] Torrance G, Feeny D, Furling W, et al. Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index mark 2. Med Care 1996; 34:702-22

[31] McCabe C, Stevens K, Roberts J, Brazier J. Health state values for the HUI 2 descriptive system: results from a UK survey. Health economics, 2005; 14:231-244

[32] Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, Denton M, Boyle M. Multiattribute and Single-Attribute Utility Functions for the Health Utilities Index Mark3 System. Medical Care, 2002; 40(2):113-128

[33] Seiber WJ, Groessl EJ, David KM, Ganiats TG, Kaplan RM. Quality of Well Being Self-Administered (QWB-SA) Scale: Users Manual. 2008. Health Services Research Center, University of California, San Diego.

[34] Apajasalo, M., et al., Quality of life in early adolescence: a sixteen-dimensional health-related measure (16D). Quality of Life Research, 1996a. 5: p. 205-211.

[35] Apajasalo, M., et al., Quality of life in pre-adolescence: a 17-dimensional health-related measure (17D). Quality of Life Research, 1996b. 5: p. 532-538.

[36] Stevens, K.J., et al., The development of a preference-based measure of health in children with atopic dermatitis. British Journal of Dermatology, 2005. 153(2): p. 372-377.

[37] Choiu, C.F., et al., Development of the multi-attribute Paediatric Asthma Health Outcome Measure (PAHOM). International Journal for Quality in Health Care, 2005. 17(1): p. 23-30.

[38] Saigal, S., Stoskopf, B. L., Feeny, D. and et al. (1999). Differences in preferences for neonatal outcomes among health care professionals, parents, and adolescents. JAMA; 281(21): 1991-1997.

[39] Mott DJ, Rivero-Arias O , Shah K, Ramos-Goñi JM, Devlin NJ (2019) Valuing the EQ-5D-Y using a discrete choice experiment: do adult and adolescent preferences differ? OHE Research Paper. London: Office of Health Economics.

[40] Lipstein EA, Brinkman WB, Fiks AG, Hendrix KS, Kryworuchko J, Miller VA, Prosser LA, Ungar WJ, Fox D. An Emerging Field of Research: Challenges in Pediatric Decision Making. Medical Decision Making, 2015; 35(3): 403-408.

[41] United Nations. Convention on the rights of the child. 1989, United Nation: London.

[42] Department of Health Chief Medical Officer's annual report Our Children Deserve Better: Prevention Pays. 2012, Department of Health.

[43] National Health and Hospital Report Commission. A healthier future for all Australians: final report of the National Health and Hospital Reform Commission. 2009.

[44] Ministerio de Sanidad y Consumo. Ganar Salud con la Juventud. 2008, Ministerio de Sanidad y Consumo: Madrid.

[45] Stevens, K J. Because that's what matters to me". A pilot study to test the feasibility and reliability of ordinal valuation methods for health state valuation with children. HEDS Discussion Paper, 2015 https://www.shef.ac.uk/scharr/sections/heds/discussion-papers/15-05-1.526948

[46] Norman R, Viney R, Aaronson NK, Brazier JE, Cella D, Costa DSJ, Fayers PM, Kemmler G, Peacock S, Pickard AS , Rowen D et al (2016) Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Quality of Life Research, 25(3), 637-649.

[47] Mott, D. J., Shah, K., Ramos-Goñi, J., Devlin, N. and Rivero-Arias, O. (2019). Valuing EQ-5D-Y health states using a discrete choice experiment: do adult and adolescent preferences differ? OHE Research Paper. London: Office of Health Economics.

[48] Dalziel K, Catchpool M, Garcia-Lorenzo B, Gorostiza I, Norman R, Rivero-Arias O. Feasibility and validity of adolescent and adult health state preferences for EQ-5D-Y states in Australia and Spain: an application of best-worst scaling. Accepted for publication in Pharmacoeconomics

[49] Yi, M. S., Britto, M. T., Wilmott, R. W., Kotagal, U. R., Eckman, M. H., Nielson, D. W., Kociela, V. L. and Tsevat, J. (2003). Health values of adolescents with cystic fibrosis. The Journal of Pediatrics; 142(2): 133-140.

[50] Wu XY, Ohinmaa A, Johnson JA, Veugelers PJ. Assessment of children's own health status using visual analogue scale and descriptive system of the EQ-5D-Y: linkage between two systems. Quality of Life Research, 2014; 23:393-402.

[51] Brazier J, Rowen DL, Karimi M, Peasgood T, Tsuchiya A, Ratcliffe J. Experience-based utility and own health state valuation for a health state classification system: why do it and how to do it. The European Journal of Health Economics, 2018; 19:881-891.

[52] Brouwer W, Versteegh M. Patient and general public preferences for health states: A call to reconsider current guidelines. Social Science & Medicine; 2016: 165:66-74.

[53] Sanders GD, Neumann PJ, Basu A, Brock DQ, Feeny D, Krahn M, Kuntz KM, Meltzer DO, Owens DK, Prosser LA, Salomon JA, Sculpher MJ, Trikalinos TA, Russell LB< Siegel JE, Ganiats TG. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA, 2016;316:1093-1103.

[54] McTaggart-Cowan H. Elicitation of informed general population health state utility values: a review of the literature. Value in Health. 2011;14(1153-1157).

[55] Kreimeier S, Oppe M, Ramos-Goni JM, Cole A, Devlin N, Herdman M, Mulhern B, Shah KK, Stolk E, Rivero-Arias O, Greiner W. Valuation of EuroQol Five-Dimensional Questionnaire, Youth Version (EQ-5D-Y) and EuroQol Five-Dimensional Questionnaire, Three-Level Version (EQ-5D-3L) Health States: The Impact of Wording and Perspective, 2018; Value in Health; 21:1291-1298.

[56] Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate societal health state utility values. J Health Econ. 2012;31:306–18.

[57] Norman, R., Mulhern, B. and Viney, R. (2016). The Impact of Different DCE-Based Approaches When Anchoring Utility Scores. PharmacoEconomics; 34(8): 805-814.

[58] Krucien N. Watson V. and Ryan, M. Is best-worst scaling suitable for health state valuation? A comparison with discrete choice experiments, Health Economics. 2017; 26:12.

[59] Krucien N Sicsic J and Ryan M. For better or worse? Investigating the validity of best-worst discrete choice experiments in health. Health Economics. 2019; 28:572-586.

[60] Parkin D and Devlin N. Is there a case for using visual analogue scale valuations in cost utility analysis? Health Economics, 2006; 15:653-664.

[61] Brazier J, Ratcliffe J, Saloman J & Tsuchiya A (2016) Measuring and Valuing Health Benefits for Economic Evaluation. Oxford: Oxford University Press

[62] Kind P, Klose K, Gusi N, Olivares PR, Greiner W. Can adult weights be used to value child health states? Testing the influence of perspective in valuing EQ-5D-Y. Quality of Life Research, 2015; 24:2519-2539

[63] Rowen D, Brazier J, Van Hout B. A Comparison of Methods for Converting DCE Values onto the Full Health-Dead QALY Scale. Medical Decision Making, 2015; 35:328-340.

[64] Hill H, Rowen D, Pennington D, Wong R, Wailoo A. NICE DSU Report. A review of the methods used to estimate and model utility values in NICE technology appraisals for paediatric populations. 2019 forthcoming.

[65] Kromm, S.K., Bethell, J., Kraglund, F. et al. Characteristics and quality of pediatric cost-utility analyses. Quality of Life Research, 2012; 21:1315.

[66] Herdman M, Cole A, Hoyle CK, Coles V, Carroll S, Devlin N. Sources and Characteristics of Utility Weights for Economic Evaluation of Pediatric Vaccines: A Systematic Review. Value in Health, 2016; 19:255-266.

[67] Kwon J, Wook Kim S, Ungar WJ, Tsiplova K, Madan J, Petrou S. A Systematic Review and Meta-Analysis of Childhood Health Utilities. Medical Decision Making, forthcoming.

[68] Thorrington D, Eames K (2015) Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. PLoS ONE 10(8): e0135672. https://doi.org/10.1371/journal.pone.0135672

[69] Brazier JE, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, Barkham M & Ibbotson R (2012) 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, 16(32), 1-114

[70] Varni JW, Seid M, Rode CA. The PedsQL™: measurement model for the pediatric quality of life inventory. Medical care 1999:126-139.

[71] DeWalt DA, Gross HE, Gipson DS, Selewski DT, DeWitt EM, Dampier CD, Hinds PS, Huang I, Thissen D, Varni JW. PROMIS Pediatric Self Report Scales Distinguish Subgroups of Children Within and Across Six Common Pediatric Chronic Health Conditions. Qual Life Res. 2015 Sep; 24(9): 2195–2208.

[72] Ramos-Goñi JM, Shah K, Kreimeier S, Devlin NJ (2019) Anchoring latent scale values for the EQ-5D-Y at 0 = dead. OHE Research Paper. London: Office of Health Economics.

[73] Lavelle TA, D'Cruz BN, Mohit B, Ungar WJ, Prosser LA, Tsiplova K, Vera-Llonch M, Lin PJ. Family Spillover Effects in Pediatric Cost-Utility Analyses. Applied Health Economics and Health Policy (2019) 17:163–174

[74] Tilford JM, Payakachat N. Progress in measuring family spillover effects for economic evaluations. Expert Rev Pharmacoecon Outcomes Res. 2015;15(2):195–8.

[75] Prosser LA, Wittenberg E. Advances in Methods and Novel Applications for Measuring Family Spillover Effects of Illness. PharmacoEconomics 2019; 37:447-450.

[76] Brouwer WR. The Inclusion of Spillover Effects in Economic Evaluations: Not an Optional Extra. PharmacoEconomics 2019; 37:451-456.

[77] Raat H, Landgraf JM, Oostenbrink R, Moll HA, Essink-Bot ML. Reliability and validity of the Infant and Toddler Quality of Life Questionnaire (ITQOL) in a general population and respiratory disease sample. Quality of Life Research, 2007; 16:445-460.

[78] Landgraf JM, Vogel I, Oostenbrink R, van Baar ME, Raat H. Parent-reported health outcomes in infants/toddlers: measurement properties and clinical validity of the ITQOL-SF47. Quality of Life Research, 2013; 22(3): 635-646

[79] Volger S, Landgraf JM, Mao M, Ge J, Northington R, Hays NP. Feasibility and Psychometric Properties of the Infant Toddler Quality of Life (ITQOL) Questionnaire in a Community-Based Sample of Healthy Infants in China. Maternal and Child Health Journal. 2018 22(5):702-712.

Table 1: *Summary of the classification systems of child and adolescent-specific generic preference-based measures*

Measure	Age appropriate to measure health for (years)	Classification system content	Number of dimensions	Response levels	Number of health states	References
AHUM	12-18	Self-care; pain; limitations walking around (mobility); perceptions of strenuous activities; self-image; health perceptions	6	4-7	16,800	[14]
AQoL-6D	Unclear	Independent living; relationships; mental health; coping; pain; senses	6	4-6	7.8×10^{13}	[15]
CHU9D	4-17	Worry; sadness; pain; tiredness; annoyance; school; sleep; daily routine; activities	9	5	1,953,125	[6,17-18]
EQ-5D-Y	4-15	Mobility; looking after myself; doing usual activities; having pain or discomfort; feeling worried, sad or unhappy	5	3	243	[26-28]
HUI2	5 upwards	Sensation; mobility; emotion; cognition; self-care; pain; fertility	7	3-5	24,000	[30]
HUI3	5 upwards	Vision; hearing; speech; ambulation; dexterity; emotion; cognition; pain	8	5-6	972,000	[32]
QWB	Unclear	Chronic symptoms or problems; acute physical problems; mental health; mobility; physical activity; social activity	3 plus 58 symptoms	2-4	945	[33]
16D	12-15	Mobility; vision; hearing; breathing; sleeping; eating; elimination; speech; mental function; discomfort and symptoms; school and hobbies; friends; physical appearance; depression; distress; vitality	16	5	1.5×10^{11}	[34]
17D	8-11	Mobility; breathing; school and hobbies; friends; hearing; vision; eating; elimination; vitality; sleeping; anxiety; discomfort and symptoms; learning and	17	5	7.6×10^{11}	[35]

Measure	Age appropriate to measure health for (years)	Classification system content	Number of dimensions	Response levels	Number of health states	References
		memory; ability to concentrate; depression; speech; physical appearance				

Table 2: Considerations and study characteristics in the valuation of child and adolescent-specific-specific measures

Whose values	Children and adolescents						Adults						
Perspective	Own Other[1]						Child and adolescent Own Other						
Elicitation technique	BWS	DCE	Ranking	VAS	Time trade-off	Standard gamble	BWS	DCE	Ranking	VAS	DCE with duration	Time trade-off	Standard gamble
Method of anchoring	Mapping via regression analysis to anchored values e.g. time trade-off values Rescaling using anchored adult/young adult values for worst state/small numbers of states using DCE with duration, time trade-off or standard gamble Hybrid model to combine, for example, DCE and time trade-off values			Using MAUT to combine, for example VAS and standard gamble values Values may be anchored onto 1-0 scale directly where dead is also rated alongside the health state on the same scale	Values anchored onto 1-0 scale directly		Mapping via regression analysis to anchored values e.g. time trade-off values Rescaling using anchored adult/young adult values for worst state/small numbers of states using DCE with duration, time trade-off or standard gamble Hybrid model to combine, for example, DCE and time trade-off values			Values may be anchored onto 1-0 scale directly depending on VAS protocol used Using MAUT to combine, for example VAS and standard gamble values	Modelled latent scale values are anchored using duration coefficient	Values anchored onto 1-0 scale directly	
Considerations	May require adult/young adult values to anchor onto 1-0 scale			VAS can be used for dimensions or states Often argued that generates preference not values as no consideration of opportunity cost To enable values on 1-0 scale and negative values must involve	May not be deemed ethically acceptable and appropriate May not be appropriate for online administration		Will require other values to anchor onto 1-0 scale			VAS can be used for dimensions or states, often argued that generates preference not values as no consideration of opportunity cost	May be inappropriate for use when using child and adolescent perspective	May be inappropriate for use when using child and adolescent perspective May not be appropriate for online administration	

Whose values	Children and adolescents			Adults			
		rating of dead alongside the health state on the same scale					
Mode of administration	Hall tests			Classroom tests			
	Face-to-face interview CAPI Online						

Notes: BWS: best-worst scaling; DCE: discrete choice experiment; VAS: visual analogue scale; MAUT: multiattribute utility theory; CAPI: computer assisted personal interview. [1]This would be another person, for example another child or adolescent.

Table 3: *Summary of the value set methodologies of child and adolescent-specific generic preference-based measures*

Measure	Country	Whose values	Perspective	Elicitation technique	Mode of administration	Method of anchoring onto 1-0 full health-dead scale	References	Year of publication	Sample size	Sample recruitment	Sample age range (years)
AHUM	UK	Adult	Own health	Time trade-off	Face-to-face interview with props	Time trade-off utility values are directly generated onto the 1-0 scale	[14]	2012	312	Participants recruited via word-of-mouth and an existing general population panel	18 upwards
AQoL-6D	Australia Fiji New Zealand Tonga	Adolescents	Own health	Time trade-off	Class test of 10-15 participants with 2 facilitators	Time trade-off utility values are directly generated onto the 1-0 scale	[16]	2010	279	Not specified	Not specified
CHU9D	Australia	Adolescents	Own health	Best-worst scaling	Online	Time trade-off utility values elicited from a sample of young adults	[19-21]	2011-2016	2076	Community based adolescent sample recruited through parents	11-17 years
	China	Adolescents	Own health	Best-worst scaling	Classroom	Time trade-off utility values elicited from a sample of young adults	[23]	2019	BWS 902, TTO 38	BWS: Multi-stage sampling method of school children; TTO: convenience sample of undergraduate students	BWS: 9-17; TTO: 18-19
	Netherlands	Adult general population	Own health	Discrete choice experiment	Online survey	Modelled latent scale values anchored	[24]	2018	1276	Participants sampled to be nationally representative using existing	18 upwards

Measure	Country	Whose values	Perspective	Elicitation technique	Mode of administration	Method of anchoring onto 1-0 full health-dead scale	References	Year of publication	Sample size	Sample recruitment	Sample age range (years)
				with duration		using duration coefficient				online panel of general population	
	UK	Adult general population	Own health	Standard gamble	Face-to-face interview with props	Standard gamble utility values are directly generated onto the 1-0 scale	[25]	2012	282	Local sample, random street sample	18 upwards
EQ-5D-Y	US	Adult general population	7 or 10 year old child	Discrete choice experiment involving problems with one attribute for *x* years followed by full health for *y* years	Online survey	Modelled latent scale values, argued are directly on 1-0 scale	[29]	2016	4155	Participants sampled to be nationally representative using existing online panel of general population	18 upwards
HUI2	Canada	Parents of school-aged children (subsample of parents of childhood cancer patients)	Child aged 10 years	Standard gamble and visual analogue scale	Face-to-face interview with props	Standard gamble utility values are directly generated onto the 1-0 scale	[30]	1996	194	Parents of childhood cancer patients, and parents of similarly aged school children from the general population (n = 293), sampling methods not specified	Not specified

Measure	Country	Whose values	Perspective	Elicitation technique	Mode of administration	Method of anchoring onto 1-0 full health-dead scale	References	Year of publication	Sample size	Sample recruitment	Sample age range (years)
	UK	Adult general population	Child aged 10 years	Standard gamble and visual analogue scale	Face-to-face interview with props	Standard gamble utility values are directly generated onto the 1-0 scale	[31]	2005	176	Not specified, aimed for national representativeness	18 upwards
HUI3	Canada	Adult general population	Own health	Standard gamble and visual analogue scale	Face-to-face interview with props	Standard gamble utility values are directly generated onto the 1-0 scale	[32]	2002	256 (additional verification sample of 248)	Local sample, random household sample	16 upwards
QWB	US	Adult general population	Own health	VAS	Unclear	VAS values elicited assuming 0=dead/worst state	[33]	2008	430	Local sample recruited in primary care clinics and college campuses	18 upwards
16D	Finland	Adolescents aged 12-15	Own health	VAS	Classrooms after oral instruction	Value of dead elicited on VAS	[34]	1996	213	School children, patients (inpatients and outpatients)	12-15
17D	Finland	Parents	8-11 year old child	VAS	Unclear	Value of dead elicited on VAS	[35]	1996	115	School children, patients	8-11