Valuation of Child Health-Related Quality of Life by Child Age and Problem Duration Using a Novel Method

ABSTRACT

Objectives: In the U.S., the valuation of health-related quality of life (HRQoL) on a quality-adjusted life years scale has been criticized due to its discriminatory aspects. Therefore, in this EQ-5D-Y-3L valuation study, we assess the value of child HRQoL on a novel "experience" scale, which ranges from 0 ("no experience"/unconscious) to 1 ("experience full health"), and test whether value depends on the child's age and the duration of the problems.

Methods: In Spring 2024, 2,577 U.S. adults were assigned to one of 16 arms that differed only in the child's age and problem duration. Each participant completed a discrete choice experiment (DCE) with five paired comparisons and 20 kaizen tasks. Using this stated preference evidence, we estimated a conditional logit model for each arm and tested for differences in EQ-5D-Y-3L values.

Results: In each arm, all 10 incremental differences in levels were positive (p-value < 0.05) with the highest value on alleviating a child's pain and discomfort. Child age affected the value of self-care, such as relieving "problems taking a bath or showering by myself or getting dressed by myself". The results didn't indicate any differences in main effects by problem duration. The value of the worst EQ-5D-Y-3L profile (33333) relative to being "in a coma" differed by child age and problem duration, ranging from 0.202 to -0.346 across 16 arms (p-value < 0.05).

Conclusion: In the U.S., the value of relieving problems with self-care depends on a child's age, possibly due to perceived differences in developmental norms.

Keywords: Quality-adjusted life years, Inflation Reduction Act, Discrete choice experiments, Kaizen task, EQ-5D-Y-3L

Highlights

- This study addresses the gap in U.S. child HRQoL valuation by introducing an "experience" scale for the EQ-5D-Y-3L, responding to regulatory concerns about the discriminatory aspects of traditional QALY-based scaling.
- Using DCE and kaizen tasks with 2,577 U.S. adults, pain and discomfort relief had the highest value, self-care valuation varied by child age. The value of the worst EQ-5D-Y-3L profile (33333) relative to being "in a coma" differed by child age and problem duration, ranging from 0.202 to -0.346 across 16 arms.
- These findings inform healthcare decision-making by providing age-sensitive HRQoL values for children that align with current U.S. regulations restricting QALY-based approaches in value assessment.

1. INTRODUCTION

Health-related quality of life (HRQoL) is defined as the consequences of a person's health (including healthcare) on their quality of life, emphasizing the causal relationship between health and quality of life[1,2,3]. In health preference research, several methods, such as time trade-off (TTO), standard gamble (SG), and choice tasks (e.g., paired comparisons), are used to summarize gains in HRQoL. Among these, discrete choice experiments (DCE) include choice tasks to elicit preferences from the perspective of the general population, particularly when assessing multi-attribute trade-offs using hypothetical scenarios. These DCEs produce value sets that may be used in health technology assessments (HTAs) and other decision analytics[4,5,6].

However, DCE estimates are on a latent utility scale and must therefore be anchored to the quality-adjusted life year (QALY) scale for the HTA agencies (e.g., NICE in the UK, ICER in the USA, PBAC in Australia) to reliably use in their economic evaluations. A QALY scale expresses the value of health outcomes on a scale anchored at 0 and 1, where the lower anchor represents "dead" and the upper anchor represents "starting today, one year with no health problems, then die."

In 2013, Craig and colleagues conducted a DCE study with paired comparisons analyzing child health valuation and produced U.S. EQ-5D-Y-3L (Y-3L) values on a QALY scale[7]. However, this study and subsequent child health valuation studies revealed that the respondents dislike making choices involving the lives of children because it evokes a feeling of abandoning a child[8,9], which prompted a reexamination of QALY scaling in general[10].

Furthermore, the summarizing of HRQoL and life expectancy on a QALY scale has faced various criticisms, particularly regarding its potential discriminatory aspects toward persons with disabilities, older adults, and the most vulnerable individuals in society[11,12]. Since some groups generally have lower HRQoL, QALY gains from extending their life (e.g., 10 additional years) are less than for other groups[13]. Although economists remain supportive of QALYs[14,15,16], in the U.S., these criticisms have led to lobbying efforts by various interest groups and disability rights advocates against the valuation of HRQoL on a QALY scale, particularly in the context of federal programs.

For example, the Inflation Reduction Act (IRA) of 2022[17], under section 1194e(2), states, "*The Secretary shall not use evidence from comparative clinical effectiveness research in a manner that treats extending the life of an elderly, disabled, or terminally ill individual as of lower value than extending the life of an individual who is younger, nondisabled, or not terminally ill*." This prohibits the use of QALYs or other similar measures in the Medicare Drug Price Negotiation Program, in part because the use of QALYs can lead to discrimination against those who have serious illnesses, are older, or have disabilities.

Therefore, in alignment with the EuroQol conceptual framework[18], Jumamyradov et al[19] put forth an alternative approach to summarize gains in HRQoL apart from life expectancy. Specifically, in their protocol paper, they proposed to use kaizen tasks to estimate Y-3L values on an "experience" scale instead of a QALY scale. To test the feasibility of such approach, in 2023, Jumamyradov et al[20] conducted a preference elicitation study using DCE paired comparisons and kaizen tasks to estimate U.S. (Y-3L) values on an "experience" scale instead of a QALY scale.

Each kaizen task displayed a single profile of a child's HRQoL using the Y-3L descriptive system and asked respondents to select first, second, and third for the child's problems[21]. The ranking of these three improvements creates a preference path from an origin profile (without the improvements) to a destination profile (with the three improvements, i.e., origin-destination pairs). Throughout this paper, we will refer to the work of Jumamyradov et al[20] as the "original study."

As a complement to the kaizen tasks, the original study also included paired comparisons, asking whether having the problems is preferable to "being in a coma" for a week (i.e., whether "no experience" is a relief in certain cases). Using the kaizen and paired comparison evidence, they produced U.S. Y-3L values on an "experience" scale, wherein a positive value signifies an experience that is better than having no experience at all (i.e., "being in a coma"), while a negative value indicates an experience worse than having no experience. By replacing "dying immediately" with "being in a coma," the original study responds to the United States regulations prohibiting the use of QALY scaling in value assessment.

Their approach addressed two key challenges in previous child health valuation studies. First, kaizen tasks elicit preferences through the sequential relief of child health problems (i.e., preference paths), which is less burdensome than paired comparisons (i.e., choosing between children). Second, they replaced the lower anchor of the QALY scale ("dying immediately") with "being in a coma" for a week, adopting the EuroQol's definition of "unconscious"[*]. As such, their respondents were able to focus on relieving the suffering of children instead of choosing between health outcomes for terminally ill children. Based on feedback from clinical experts, 'unconscious' alone is not an appropriate description due to its heterogeneity. Instead, 'being in a coma' is a 'lengthy deep state of unconsciousness' [24], which is a more specific profile and suitable to serve as the lower anchor of the experience scale [25,26,27,28].

[*] Each person's health resides between what they perceive to be the worst and the best imaginable states[18] (i.e., the most undesirable health state a person can imagine, in which they are unable to function independently in all areas of life, and the most desirable, ideal health state a person can imagine). Otherwise, the person is either unconscious (alive but not awake or aware of their surroundings) or dead (deprived of life). If alive and conscious, a person's health problems may reduce their quality of life along five domains, each of which is known to be a "highly valued" outcome[22,23] (i.e., similar in value to life extensions).

However, in the original study, respondents sequentially relieved the health problems of a 10-year-old child for 1 week. Furthermore, the internation valuation protocol for the Y-3L frame its hypothetical questions for a health state of a 10-year-old child[28]. It is possible that the novel approach from Jumamyradov et al[20] may produce different results across different child ages and problem durations. For instance, some studies found that child age[30,31] and duration of health problems[7] are crucial in valuing child HRQoL. Adult respondents may take into account different factors when asked to value child HRQoL with different ages and problem durations. Therefore, in this paper, we extend the original study[20] to assess the effects of child age and problem duration on U.S. Y-3L values. More specifically, we compare the results of 16 DCE surveys, interacting {children with 4, 7, 10, and 13 years of age} and {health problems lasting for 1, 2, 6, and 12 months}.

Although numerous U.S. studies have conducted DCEs for adult HRQoL valuation, only a limited number of studies have conducted DCEs for child health valuation[7,32,33,34]. Many child health valuation studies have been conducted using the international valuation protocol to produce Y-3L value sets[29]. As described above, U.S. regulations, such as the Inflation Reduction Act (IRA) of 2022, prohibit the summary of HRQoL on a QALY scale in health technology assessment. This U.S. study is among the first research efforts to produce value sets on an experience scale and to assess the effects of child's age and problem duration on values.

2. METHODS

In the Y-3L descriptive system, child HRQoL is described using five attributes[35]: walking around (i.e., mobility, MO), looking after myself (i.e., self-care, SC), doing usual activities (i.e., usual activities, UA), having pain/discomfort (PD), and feeling worried, sad, or unhappy (i.e., anxiety/depression, AD). Each attribute ranges from Level 1 (no problems) to Level 3 (a lot of problems) depending on its severity. Thus, there are 243 (3^5) possible Y-3L profiles, ranging from 11111 (best) to 33333 (worst). For instance, the profile 33211 represents the HRQoL of a child with a lot of problems in mobility and usual activities (Level 3), some problems in self-care (Level 2), and no problems in pain/discomfort and anxiety/depression (Level 1). Among the EQ-5D concepts, unconscious is defined as "being alive but not awake or aware of surroundings" and resides outside the five attributes[36]. However, this term is imprecise and can refer to a variety of experiences; therefore, each child in this study is either described using the five attributes of a Y-3L profile or as "in a coma" (i.e., "a lengthy deep state of unconsciousness")[19].

This study did not describe dead children ("deprived of all life; all functions have ceased and are incapable of being restored or revived")[36]; instead, respondents were explicitly informed that all children survived and that their recovery is the same regardless of the choices made (i.e., acute episodes).

2.1. Survey instrument

The survey instrument has five sections: consent and screener, background, paired comparisons (Y-3L vs. "being in a coma"; one warmup exercise and five tasks), kaizen tasks (Y-3L preference paths; one warmup and 20 tasks), and follow-up. The instrument is identical to what was used in the original study[20], except for the experimental design and the addition of the PROMIS-16 in the follow-up component (Appendix A). Figure 1 and Figure 2 show examples of a paired comparison and kaizen task, respectively. The survey instrument also includes a quiz consisting of five questions that were shown to deter fraudulent respondents (i.e., bots) and promote a common understanding of what "coma" entails[19]. The instrument automatically excludes respondents who failed the quiz after three attempts, a threshold set based on evidence from a pilot study with 100 respondents.

In the original study, we examined the HRQoL of a 10-year-old child with health problems lasting for a week (N = 631). In this one, we analyze four different age groups (4, 7, 10, and 13 years of age) and four different problem durations (1, 2, 6, and 12 months), resulting in 16 arms with a target sample size of 150 respondents each.

As in the original study, we recruited adults aged 18 or older from the U.S. general population and 18 demographic quotas based on gender (female and other), age (18–34, 35–54, and 55 and older), and race/ethnicity (Hispanic, non-Hispanic Black, and non-Hispanic other). To promote internal validity, we randomly assigned respondents from each of the 18 quotas to one of the 16 arms. We also randomized the task sequence and attributed order.

2.2.Experimental design

Each respondent was assigned the same five paired comparisons and 20 kaizen tasks, regardless of arm (i.e., single block design). In each kaizen task, only one attribute (i.e., hold-out attribute) was prevented from improvement. This hold-out attribute remained constant across all alternatives within a preference path for a given task. Its presence prevents the formation of lexicographic preferences, which is the cases where individuals rank attributes in order of priority and make decisions based on the highest priority attribute.

To select the 20 origin-destination (OD) pairs for the kaizen tasks, we started with the 405 candidate OD pairs in the Y-3L descriptive system (i.e., five possible hold-outs x three possible improvements per attribute [Level 1 to 2, 2 to 3, and 1 to 3]$^{\text{four differential attributes}}$ = 5 x 3^4). From this candidate pool, we iteratively selected a block of 20 OD pairs based on the D-optimal design. A D-optimal design minimizes the determinant of the information matrix based on a conditional logit model using priors from the original study[20] (i.e., D-error). Equivalently, it maximizes the determinant of the Fisher information matrix, hence "D" for determinant. After generating the D-optimal design, we evaluated the design based on a series of worst-case scenarios and assigned levels to the hold-out attributes such that all 40 origin profiles were unique (i.e., 20 unique OD pairs).

To select the five pairs for the paired comparison, we began with the 20 candidate pairs in the original study (i.e., four blocks with five paired comparisons each) and selected five profiles that showed level balance (i.e., a level must not repeat more than three times for each attribute). The full experimental design is available in Appendix B.

2.3.Analysis plan

For descriptive analysis, we compared the sample of respondents who completed the survey (Table 1) with those who dropped out using chi-squared tests and with the 2021 American Community Survey (ACS) 1-year estimates[37] based on 66% and 133% thresholds (i.e., 1/3 rule; Appendix C). Furthermore, we analyzed the median and interquartile range (IQR) of survey time overall for each DCE component (i.e., paired comparison and kaizen tasks) and by task, arm, task sequence, and attribute order (Appendix D). Using chi-square tests of independence, we tested whether response frequency varied by experimental characteristics, namely arm and task sequence. We also assessed two forms of straight-lining behavior in the paired comparisons (always HRQoL and always coma) and vertical behavior in the kaizen tasks (top down) as well as their association with arm and task sequence. Lastly, we examined respondent feedback on the three debriefing questions and open-ended comments.

For the primary analysis, let V be the value of a child health outcome. To incorporate the quality and quantity of an outcome in a single metric, one can write the value of a health outcome (V) as a product of two values representing HRQoL (V^H) and life years (V^Y):

$$V = V^H \times V^Y \tag{1}$$

We also assumed that the value of HRQoL is linear $V^H = 1 - X\beta$, where X is a vector of 10 incremental indicators of level changes (e.g., from 3 to 2 or from 2 to 1) in five of the attributes described above (i.e., MO_1, MO_2, SC_1, SC_2, UA_1, UA_2, PD_1, PD_2, AD_1, or AD_2) as well as an 11th indicator for "being in a coma" (COMA). $\beta = \{\beta_1, \ldots, \beta_{11}\}$ is a vector of the main effects (i.e., independent effects of a single attribute level). Thus, we expressed the value of HRQoL as follows:

$$V^H = 1 - \begin{pmatrix} \beta_1 MO_1 + \beta_2 MO_2 + \beta_3 SC_1 + \beta_4 SC_2 + \beta_5 UA_1 + \\ \beta_6 UA_2 + \beta_7 PD_1 + \beta_8 PD_2 + \beta_9 AD_1 + \beta_{10} AD_2 + \beta_{11} COMA \end{pmatrix} \tag{2}$$

Therefore, under this specification, the value of the best-case scenario (i.e., 11111) is 1 due to the absence of problems, and the value of the worst-case scenario (i.e., 33333) is 1 minus the sum of all 10 Y-3L main effects ($1 - \sum_{m=1}^{10} \beta_m$). In addition, we set the value of "being in a coma" at 0 by constraining its effect ($\beta_{11} = 1$). Therefore, since the lower anchor in our study is "being in a coma" instead of "dying immediately," the Y-3L values are expressed on an experience scale instead of a QALY scale. On an experience scale, 1 represents experiencing no health problems, and 0 represents experiencing nothing (i.e., "being in a coma").

Following McFadden[38], we expressed the probability of choosing alternative $j \in \{1, \ldots, J\}$ as follows:

$$P_j = \frac{\exp(\sigma V_j)}{\sum_{k=1}^{J} \exp(\sigma V_k)} \tag{3}$$

where σ represents the scaling factor. The set size J depends on the choice task: paired comparison ($J = \binom{2}{1} = 2$), first improvement ($J = \binom{4}{1} = 4$), first and second improvements (($J = \binom{4}{2} = 6$)), and first, second and third improvements (($J = \binom{4}{3} = 4$)). Using the conditional logit (CL) model for each arm, we estimated 12 parameters by maximum likelihood: 10 main effects and the log of the two scaling factors, $ln(\sigma)$, one for kaizen tasks and another for paired comparisons. Furthermore, we combined arm-specific samples to assess the linear relationship of child age and problem duration on the parameter estimates. To account for within-respondent correlations, all p-values and standard errors (SEs) were estimated using cluster bootstrap techniques with replacement and 1,000 iterations[39].

3. RESULTS

Between March 18th and April 6th, 2024, we conducted an online DCE survey and recruited 4,989 adults aged 18 or older from the U.S. general population. Among these adults, 248 dropped out before they were assigned to a study arm, and 1,742 (35%) were excluded as a result of failing to pass the coma quiz after three attempts. Among those who passed the screening and background components, 680 voluntarily dropped out during the preference elicitation tasks or follow-up questions, leaving an analytical sample of 2,577 (52%) respondents who completed the survey fully.

Table 1 shows that these respondents form a nationally representative sample with an underrepresentation of Hispanic or Latino respondents (12.0% vs. 16.9%) and respondents of two or more races (4.0% vs. 10.8%). We observed some differences in the demographic characteristics between respondents who completed the survey and those who dropped out (p-value < 0.05); however, there were no differences by U.S. region (p-value = 0.0788), and respondents from all

50 U.S. states and Washington, D.C. participated. Furthermore, the respondents with certain socioeconomic characteristics (never married, some college, and $150,000 or more annual income) who completed the survey did not align with the 1/3 rule when compared to the ACS estimates (Appendix C).

The median completion time was 17.87 minutes for the survey (IQR 13.40, 24.94), 1.04 minutes (0.72, 1.46) for the five paired comparisons, and 4.93 minutes (3.52, 7.08) for the 20 kaizen tasks. The time per task decreased by task sequence (Appendix Table D2). The audit of the stated preference evidence confirmed randomization by arm, attribute order, and task sequence. In the paired comparisons, the likelihood of choosing the coma alternative ranged from 22.8% to 50.3% and was not associated with arm or task sequence (Appendix Tables D3 and D4). However, 1,005 (38.9%) respondents choose the HRQoL profile in all five paired comparisons, and 394 (15.3%) chose the coma in all five paired comparisons. In the kaizen tasks, the frequency of the preference paths was associated with arm but not task sequence (Appendix Tables D5 to D44). Due to attribute randomization, each response pattern should occur about 4% (1 in 24); however, the vertical pattern occurred in 7.2% of the kaizen responses across arms and task. Generally, straight-lining and vertical behavior were not associated with arm or task sequence (Appendix Table D45). In relation to the paired comparisons, a greater number of respondents reported that the kaizen tasks are easier to complete (20% vs. 15.7%), easier to understand (20% vs. 15.7%), and preferred (28.2% vs. 12.7%), confirming prior qualitative feedback (Appendix Table D46).

In Table 2, we present the gains in HRQoL for a 6-month duration by child age and assess the linear relationships between child age and main effects on an experience scale. Across all four different child ages, all main effects are positive (p-value < 0.05), except the gain from "a bit worried, sad, or unhappy" to "not worried, sad, or unhappy" (AD1) for a 7-year-old child. As expected, the relief of pain or discomfort had the largest incremental effects, where the estimate of PD1 ranged from 0.188 to 0.232 and PD2 ranged from 0.275 to 0.305. For each child age, the value of the worst profile (33333) was less than zero: -0.152 for a 4-year-old child (p-value < 0.01), -0.180 for a 7-year-old child (p-value < 0.01), -0.163 for a 10-year-old child (p-value < 0.01), and -0.097 for 13-year-old child (p-value = 0.018), which implies that "being in a coma" has a higher value. Furthermore, the main effects are not linearly associated with child age (p-value > 0.05), except for "looking after myself" (p-value = 0.0003 SC1 and < 0.0001 SC2).

Figure 3 illustrates the value of "taking a bath or shower by myself or getting dressed by myself" for each duration by child age. Straight lines represent gains from some to no problems (SC1), and dashed lines represent gains from a lot to some problems (SC2). As child age increases, the value of SC1 and SC2 appear to increase for each problem duration.

In Table 3, we present the gains in HRQoL for a 10-year-old child by duration and assess the linear relationships between duration and main effects on an experience scale. Across all four different problem durations, all main effects are positive (p-value < 0.05) except for the gain from "a bit worried, sad, or unhappy" to "not worried, sad, or unhappy" (AD1) for a 6-month duration. As expected, the relief of pain or discomfort has the largest incremental effects, and the value of the worst profile (33333) is less than zero. The main effects are not linearly associated with problem duration (p-value > 0.05).

Overall, all of the 160 main-effect estimates are positive, and 14 of the 16 worst profiles have a value less than zero (p-value < 0.05) on an experience scale. To see the full results, Appendix E contains main-effect estimates by child age and problem duration (Tables E1 to E16), and Appendix F contains the combined results with linear relationships by child age (Table F1) and problem duration (Table F2).

4. DISCUSSION

This study estimates the value of child HRQoL on an experience scale for different child ages and problem durations. Our results suggest that U.S. adults value all 10 improvements in child HRQoL and that "being in a coma" is generally preferred to the Y-3L worst-case scenario (33333). Furthermore, we found that child age affects the value of "looking after myself"; however, the results did not indicate any other linear relationships between the main effects and child age or problem duration.

In the U.S. Y-3L study[7], tested for overall differences in value on a QALY scale by age (7 and 10 years) and duration of loss in child HRQoL (1 and 2 years). They found that a loss in HRQoL at age 7 is 13.1% more detrimental than a loss at age 10 and that a 2-year loss in HRQoL is not twice as detrimental as a 1-year loss, increasing the burden by about a third. Instead of eliciting the preferences of adults on child HRQol, Ramos-Goñi et al[40] asked adults in the U.K. and the U.S. to imagine the preferences of children aged "5-7 years old," "8-10 years old," "11-13 years old," and "14-15 years old." However, they found no statistically significant differences in main effects by child age in either country.

To the best of our knowledge, this is the first study to examine differences in main-effect estimates by child age and problem duration. Upon closer examination, Figure 3 illustrates that U.S. adults seem to place a lower value on 'looking after myself' at 4 years old compared to ages 7, 10, and 13. Interestingly, the remaining attributes demonstrated consistent values across different ages, emphasizing their stability regardless of the child's age. This finding on self-care is not unexpected, and to some extent, an indication of the internal validity of our methods, that adults assigned less importance to 'taking a bath or shower by myself or getting dressed by myself' for a 4-year-old child. While it is anticipated that a 4-year-old can perform certain washing and dressing tasks, it is not necessarily expected that they can do so independently[41]. According to the U.S. Centers for Disease Control and Prevention, having difficulty with self-care at age 4 is developmentally appropriate[42]; therefore, improvements in washing and dressing at age 4 may seem less important to U.S. adults compared to improvements at age 7, 10, and 13. This observation highlights the need for age-appropriate items and attributes in child HRQoL measurement and valuation, as recommended for an adapted version of the Y-3L designed for ages 2 to 4[43].

Although the focus of this study is on child age and problem duration, its approach builds upon the work conducted by Jumamyradov et al.[20] introducing the use of experience scaling and kaizen tasks for the purpose of avoiding discrimination against vulnerable populations and making the DCE less burdensome for its respondents. This paper further demonstrates how this approach can mitigate key issues in health preference research, particularly in regard to producing HRQoL value sets.

As described in the introduction, this is not a standard Y-3L value set on a QALY scale[7,29]. Instead, we produced a Y-3L value set on an experience scale. This scale may be used to compare gains in HRQoL (apart from life extension) in economic evaluations, and the elicitation of the values on this scale does not require choosing between quality of life and life extensions or the consideration of challenging outcomes (e.g., immediate death). For detailed description of methodological choices regarding the derivation of child health state utilities for use in economic evaluation, see the ISPOR Task Force Report[44].

We recognize that the QALY scale may be used to summarize gains in HRQoL and life extensions in some contexts, but the QALY scale has faced various criticisms regarding its potential discriminatory aspects toward persons with disabilities, older adults, and the most vulnerable individuals in society. As a result, various interest groups have lobbied against the use of the QALY scale in federal programs, and legislation addressing this, such as the Inflation Reduction Act (IRA) of 2022, has been enacted to limit their use.

Furthermore, this DCE survey included kaizen tasks, which were developed in response to respondent feedback that making choices involving the lives of children evoked feelings of guilt at the idea of abandoning a child[7,8,9,21]. Instead of choosing between HRQoL profiles for a child, each respondent in a kaizen task promotes the sequential relief of child health problems, which is more appealing to respondents and more similar to what happens in clinical practice.

This study has some limitations that we would like to mention. First, although the recruitment strategy attempted to accord with U.S. Census evidence, some categories of respondent characteristics such as “some other race,” “two or more races” (Table 1), “never married,” and “some college, no degree” (Appendix C) fell outside of the 1/3 tolerance interval (i.e., being within 66% - 133% of the ACS estimates). Second, this study does not use a standard QALY scale to produce Y-3L values. While QALYs incorporate tradeoffs between gains in HRQoL and life extensions, experience scaling is a novel approach that omits life extensions from its calculations. This distinction is relevant in the context of recent U.S. regulations that restrict the use of QALY-based scaling in value assessment; by anchoring values to 'being in a coma' rather than 'dying immediately,' experience scaling avoids one of the specific features that regulators have flagged. However, it is important to note that experience scaling is not a wholesale solution to the broader concerns surrounding QALYs. The approach remains relatively new, lacks the extensive validation that QALYs have accumulated over decades, and its omission of life-extension tradeoffs may itself be viewed as a limitation depending on the decision context. More research is required before experience scaling can be considered mainstream in measuring gains in HRQoL, and its alignment with regulatory requirements should not be interpreted as evidence of its superiority over QALY-based methods in all respects. The approach could in principle be used to value any child and adult HRQoL instrument in a manner that avoids anchoring at dead = 0.

5. CONCLUSION

In this paper, we estimate the value of child HRQOL on an experience scale, which ranges from 0 (i.e., "no experience"/unconscious) to 1 (i.e., "experience full health"). Our results suggest that child age affects the value that U.S. adults place on self-care and relieving issues such as "problems taking a bath or showering by myself or getting dressed by myself." However, the results did not indicate differences in the main effects by problem duration. Based on these findings, future child health valuation studies may consider DCE with kaizen tasks, presenting values on an experience scale, and setting the minimum age in concordance with developmental norms.

In alignment with current U.S. guidelines (via the Inflation Reduction Act of 2022), we produced U.S. Y-3L values that summarize the gains in child HRQoL by child age and problem duration. Furthermore, we describe the importance of alleviating a child's pain and discomfort and the relevance of problems in performing self-care depends on a child's age.

REFERENCES

1. Shah K. EuroQol Working Papers Series. A brief review of concepts: health, quality of life, health-related quality of life and well-being. Number 17001. Accessed June7, 2024. Available at https://euroqol.org/wp-content/uploads/2024/01/EuroQol-Working-Paper-Series-Manuscript-17001-Koonal-Shah.pdf
2. Oppe M, Craig BM. EuroQol Working Papers Series. Conceptual issues with Health Valuation and its Underlying Assumptions. Number 17002. Accessed June7, 2024. Available at https://euroqol.org/wp-content/uploads/2024/01/EuroQol-Working-Paper-Series-Manuscript-17002-Mark-Oppe.pdf
3. Karimi M, Brazier J. Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? Pharmacoeconomics. 2016;34(7):645-9. https://doi.org/10.1007/s40273-016-0389-9
4. Charlton V, DiStefano M, Mitchell P, et al. We need to talk about values: a proposed framework for the articulation of normative reasoning in health technology assessment. Health Economics, Policy and Law. Published online 2023:1-21. https://doi.org/10.1017/s1744133123000038
5. Stolk EA, Craig BM, Mulhern B, et al. Health Valuation: Demonstrating the Value of Health and Lifespan. Patient. 2017;10(4):515-517. https://doi.org/10.1007/s40271-017-0252-x
6. Thurstone LL. The measurement of values. Psychological Review. 1954;61(1):47-58. https://doi.org/10.1037/h0060035
7. Craig BM, Greiner W, Brown DS, et al. Valuation of Child Health-Related Quality of Life in the United States. Health Economics. 2016a;25(6):768-777. https://doi.org/10.1002/hec.3184
8. Powell PA, Rowen D, Rivero-Arias O, et al. Valuing child and adolescent health: a qualitative study on different perspectives and priorities taken by the adult general public. Health and Quality of Life Outcomes. 2021;19(1):1-14. https://doi.org/10.1186/s12955-021-01858-x
9. Dewilde S, Janssen MF, Lloyd AJ, et al. Exploration of the Reasons Why Health State Valuation Differs for Children Compared With Adults: A Mixed Methods Approach. Value in Health. 2022;25(7):1185-1195. https://doi.org/10.1016/j.jval.2021.11.1377
10. Sampson C, Parkin D, and Devlin N. Is anchoring at 'dead' a theoretical requirement for health state valuation? Health Economics. 2024;33(9):1929-1935. https://doi.org/10.1002/hec.4863
11. Schneider P. The QALY is ableist: on the unethical implications of health states worse than dead. Quality of Life Research. 2022;(31):1545-1542. https://doi.org/10.1007/s11136-021-03052-4
12. Devlin NJ, Drummond MF, Mullins CD. QALYs, QALY-like measures, or neither? The debate continues. Value in Health. Published online May 1, 2024. https://doi.org/10.1016/j.jval.2024.04.021
13. Rand LZ, Kesselheim AS. Controversy Over Using Quality-Adjusted Life-Years In Cost-Effectiveness Analyses: A Systematic Literature Review. Health Affairs. 2021;40(9):1402-1410. https://doi.org/10.1377/hlthaff.2021.00343
14. Lakdawalla DN, Doshi JA, Garrison LP, et al. Defining Elements of Value in Health Care—A Health Economics Approach: An ISPOR Special Task Force Report [3]. Value in Health. 2018;21(2):131-139. https://doi.org/10.1016/j.jval.2017.12.007
15. Garrison LP, Kamal-Bahl S, Towse A. Toward a Broader Concept of Value: Identifying and Defining Elements for an Expanded Cost-Effectiveness Analysis. Value in Health. 2017;20(2):213-216. https://doi.org/10.1016/j.jval.2016.12.005
16. Dolan P, Shaw R, Tsuchiya A, et al. QALY maximisation and people's preferences: a methodological review of the literature. Health Economics. 2005;14(2):197-208. https://doi.org/10.1002/hec.924
17. H.R.5376 - 117th Congress (2021-2022): Inflation Reduction Act of 2022. (2022, August 16). https://www.congress.gov/bill/117th-congress/house-bill/5376
18. Fox-Rushby J, Selai C. (2003). What concepts does the EQ-5D measure? Intentions and interpretations. In: Brooks R, Rabin R, de Charro F, eds. The Measurement and Valuation of Health Status Using EQ-

5D: A European Perspective: Evidence from the EuroQol BIOMED Research Programme. The Netherlands: Springer 2003;167-182. https://link.springer.com/chapter/10.1007/978-94-017-0233-1_10
19. Jumamyradov M, Craig BM, Rivero-Arias O, et al. Child health valuation protocol for a discrete choice experiment comparing paired comparison and kaizen tasks and estimating US EQ-5D-Y-3L values on an experience scale. BMJ open. 2023;13(10):e077256. https://doi.org/10.1136/bmjopen-2023-077256
20. Jumamyradov M, Craig BM, Jakubczyk M. Re-visiting the Valuation of Child Health-Related Quality of Life: Replacing paired comparisons with kaizen tasks and QALY scaling with experience scaling. Medical Care. 2025;63(10):771-778 https://doi.org/10.1097/MLR.0000000000002200
21. Craig BM, Rand K, Hartman JD. Preference paths and their Kaizen tasks for small samples. Patient. 2022;15:187-196. https://doi.org/10.1007/s40271-021-00541-z
22. Williams A. The EuroQol instrument. In: Kind P, Brooks R, Rabin R, eds. EQ-5D Concepts and Methods: A developmental history. The Netherlands: Springer 2005;1-17. https://link.springer.com/content/pdf/10.1007/1-4020-3712-0.pdf#page=21
23. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208. https://doi.org/10.1016/0168-8510(90)90421-9
24. National Institute of Health. Coma. Available: https://www.ninds.nih.gov/health-information/disorders/glossary-neurological-terms#coma [Accessed March 2026]
25. Posner et al. Plum and Posner's Diagnosis of Stupor and Coma, 4 edn, Contemporary Neurology Series (2008; online edn, Oxford Academic, 1 Apr. 2013), https://doi.org/10.1093/med/9780195321319.001.0001
26. Hux, K., Schram, C. D., and Goeken, T. Misconceptions about brain injury: A survey replication study. Brain Injury. 2006;20(5), 547–553. https://doi.org/10.1080/02699050600676784
27. Young, G.B. Coma. Annals of the New York Academy of Sciences. 2009;1157: 32-47. https://doi.org/10.1111/j.1749-6632.2009.04471.x
28. Patel S and Hirsch N. Coma. Continuing Education in Anaesthesia Critical Care & Pain. 2014;14(5):220-223. https://doi.org/10.1093/bjaceaccp/mkt061
29. Ramos-Goñi, J.M., Oppe, M., Stolk, E. et al. International Valuation Protocol for the EQ-5D-Y-3L. PharmacoEconomics. 2020;38:653–663. https://doi.org/10.1007/s40273-020-00909-3
30. Reckers-Droog et al. Why do adults value EQ-5D-Y-3L health states differently for themselves than for children and adolescents: a think-aloud study. Value Health. 2022; 25(7):1174-1184. https://doi.org/10.1016/j.jval.2021.12.014
31. Dewilde et al. Exploration of the reasons why health state valuation differs for children compared with adults: a mixed methods approach. Value Health. 2022; 25(7):185-1195. https://doi.org/10.1016/j.jval.2021.11.1377
32. Craig BM, Brown DS, Reeve BB. The Value Adults Place on Child Health and Functional Status. Value in Health. 2015;18(4):449-456. https://doi.org/10.1016/j.jval.2015.02.012
33. Craig BM, Brown DS, Reeve BB. Valuation of Child Behavioral Problems from the Perspective of US Adults. Medical Decision Making. 2016b;36(2):199-209. https://doi.org/10.1177/0272989x15594370
34. Pickard S et al. Valuing Child Health: An EQ-5D-Y-3L Value Set for the United States. Value in Health. 2025. https://doi.org/10.1016/j.jval.2025.11.003.
35. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Quality of Life Research. 2010;19(6):875-886. https://doi.org/10.1007/s11136-010-9648-y
36. Brooks R, Rabin R, F. de Charro. The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective. Springer Science & Business Media; 2013.
37. American Community Survey [database online]. United States Census Bureau; 2021. Updated August 16, 2023.

38. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, eds. Frontiers in Econometrics. New York: Academic Press 1974;105–142.
39. Henningsen A, Toomet O. maxLik: A package for maximum likelihood estimation in R. Computational Statistics. 2011;26(3):443-458. https://doi.org/10.1007/s00180-010-0217-1
40. Ramos-Goñi JM, Estévez-Carrillo A, Rivero-Arias O, et al. Does Changing the Age of a Child to be Considered in 3-Level Version of EQ-5D-Y Discrete Choice Experiment–Based Valuation Studies Affect Health Preferences? Value in health. 2022;25(7):1196-1204. https://doi.org/10.1016/j.jval.2022.03.001
41. Hayton J, Wall K, Dimitriou D. Get your coat: examining the development of independent dressing skills in young children with visual impairment, Down syndrome and typically developing children. International Journal of Inclusive Education. 2020;24(3):235-250. https://doi.org/10.1080/13603116.2018.1456568
42. Zubler JM, Wiggins LD, Macias MM, et al. Evidence-Informed Milestones for Developmental Surveillance Tools. Pediatrics. 2022;149(3). https://doi.org/10.1542/peds.2021-052138
43. Dalziel K, Alexander van Heusden, Janani Sarvananthar, et al. A Qualitative Investigation to Develop an Adapted Version of the EQ-5D-Y-3L for Use in Children Aged 2-4 Years. Value in health. 2023;26(10):1525-1534. https://doi.org/10.1016/j.jval.2023.06.004
44. Matza LS, Rowen D, Chandler F, et al.. Valuing child and adolescent health states for use in economic evaluation: A good practices report of an ISPOR task force. 2026. Value in Health. https://doi.org/10.1016/j.jval.2025.12.016
45. Kreimeier S, Åström M, Burström K, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. Quality of Life Research. 2019;28(7):1951-61. https://doi.org/10.1007/s11136-019-02115-x