# University of Groningen

## Valuing Health Status in the First Year of Life

Jabrayilov, Ruslan; Vermeulen, Karin M; Detzel, Patrick; Dainelli, Livia; van Asselt, Antoinette D I; Krabbe, Paul F M

*Document Version*
Publisher's PDF, also known as Version of record

# Valuing Health Status in the First Year of Life: The Infant Health-Related Quality of Life Instrument

Ruslan Jabrayilov, PhD,[1] Karin M. Vermeulen, PhD,[1] Patrick Detzel, PhD,[2] Livia Dainelli, PhD,[2] Antoinette D.I. van Asselt, PhD,[1] Paul F.M. Krabbe, PhD[1,*]

*[1]Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; [2]Nestlé Research Center, Lausanne, Switzerland*

## A B S T R A C T

*Objectives:* Efforts to evaluate HRQoL and calculate quality-adjusted life years (QALYs) for infants less than 12 months of age are hampered by the lack of preference-based HRQoL instruments for this group. To fill this gap, we developed the Infant Quality of life Instrument (IQI), which is administered through a mobile application. This article explains how weights were derived for the 4 levels of each health item.

*Methods:* The IQI includes 7 health items: sleeping, feeding, breathing, stooling/poo, mood, skin, and interaction. In an online survey, respondents from the general population (n = 1409) and primary caregivers (n = 1229) from China, the United Kingdom, and the United States were presented with 10 discrete choice scenarios. Coefficients for the item levels were obtained with a conditional logit model.

*Results:* The highest coefficients were found for sleeping, feeding, and breathing. All coefficients for these items were negative and logically ordered, meaning that more extreme levels were less preferred. Stooling, mood, skin, and interaction showed some irregularities in the ordering of coefficients. Results for caregivers and the general population were about the same.

*Conclusions:* The IQI is the first generic instrument to assess overall HRQoL in infants up to 1 year of age. It is short and easy to administer through a mobile application. We demonstrated how to derive values for infant health states with a discrete choice methodology. Our next step will be to normalize these values into utilities ranging from 0 (dead) to 1 (best health state) and to collect IQI values in a clinical population.

*Keywords:* infants, health-related quality of life, health states, value, measurement

## Introduction

In the past decades, the conceptualization of health has expanded beyond clinical indicators of physical well-being. In line with the World Health Organization (WHO) definition of health, assessment of health status now also includes inferences about the impact of health on people's social and emotional lives.[1] Broadening the scope has led to the conceptualization of constructs such as "health-related quality of life" (HRQoL). Therefore, it is no coincidence that regulatory bodies such as the Food and Drug Administration (FDA) and National Institute for Health and Care Excellence actively encourage qualitative assessments in addition to traditional clinical assessments of health.[2,3]

There are several widely used HRQoL instruments for adults. Nevertheless, much less progress has been made in developing and measuring HRQoL in younger age groups. Some generic instruments, such as the Infant and Toddler Quality of Life Questionnaire[4,5] and the Pre-school Children Quality of Life Questionnaire,[6,7] can be used to measure HRQoL in young children, and the Infant and Toddler Quality of Life Questionnaire even in infants under 12 months. Nevertheless, these are conventional HRQoL instruments; they consist of sections that yield separate measures for various health domains rather than a single score capturing overall HRQoL.

Measuring the overall impact of a health condition requires preference-based methods. Instead of measuring the level of the reported complaints (ie, their frequency and intensity), these methods express the quality of health (or of specific health conditions) by generating a single number that reflects the patient's health status as a whole.[8] Respondents are asked to formulate a

value judgment about a specific health phenomenon, condition, or outcome by making trade-offs between health items or attributes. For this reason, preference-based methods do not generate "scores" but "values." At present a few preference-based instruments, such as the Child Health Utility 9D (CHU-9D) and the EuroQol five-dimensional questionnaire youth version,[9,10] are available or under development for children around 7 years and older.[11] Although proxy versions of these instruments can be used for children below this age, the relevance of some of the included health items regarding social and emotional aspects (eg, school work, being annoyed, joining in activities) would be questionable for the first years of life. In addition, the first version of the Health Utilities Index Mark 1 was developed for use in evaluating outcomes of neonatal intensive care for infants with very low birth weight.[12] This instrument consists of four items (physical function, role functioning, social-emotional function, and health problems). Responses with this instrument were collected as part of a follow-up study for two cohorts: infants from 9 to 14 years and from 1.5 to 6 years.[13,14] The Health Utilities Index Mark 1 can be considered as an infant instrument, but has not been widely used. The successor of this initial 4-item instrument was the Health Utilities Index Mark 2. Although widely used in adults, it was originally developed with an application to evaluate the long-term effect of childhood cancer. Because of this explicit long-term goal, including a fertility item next to 6 generic health items, the Health Utilities Index Mark 2 seems not fully tailored for children.[15] Also, a recent review by Thorrington and Eames[16] shows that these child-specific instruments have not been used widely in practice. They conclude that the regular (adult) versions of instruments, such as the EuroQol five-dimensional questionnaire youth version and Health Utilities Index Mark 3, were used most frequently to obtain health utilities (normalized values with a lower anchor of 0 = dead and an upper anchor of 1.0 = full health) from pediatric populations. A meta-analysis of childhood health utilities by Kwon et al.[17] also found that although child-specific instruments are applied quite frequently, the most commonly used indirect valuation method was the Health Utilities Index Mark 3. The same observation was made by Montgomery and Kusel,[18] who found that in most published National Institute for Health and Care Excellence appraisals concerning child populations, adult utilities were used to inform the decision model. Nevertheless, as they also argue, children are not small adults and we cannot assume that their preferences are the same as in the adult population. Summarizing, it can be said that although HRQoL preference-based instruments for children are available, they are not commonly used and are rarely suitable to generate utilities, in particular where it concerns the infant population.

In an effort to fill this gap, we have developed the Infant Quality of life Instrument (IQI), which aims to measure health status in the first year of life as perceived by caregivers.[19] Based on two extensive searches of the current HRQoL literature for infants, a comprehensive list of all health items that were observable and applicable to each time point up to 1 year of age was compiled. Subsequently, three international expert meetings were held in which the items were reviewed and excluded from the list in case they were deemed unequivocally irrelevant for HRQoL of the infant population. Also, based on the input from the experts, items could be rephrased. The final step consisted of two international surveys with primary caregivers to obtain feedback on the importance and relevance of the candidate items proposed and to identify additional parent-generated items not previously considered. A second survey was conducted to test the usability of the mobile application that is used to administer the IQI. A detailed report of the process of selecting the health items and the levels to include in the IQI can be found elsewhere.[19]

**Figure 1.** Infant Quality of Life Instrument (IQI) health items and their levels (left: screenshot of the app for the IQI).
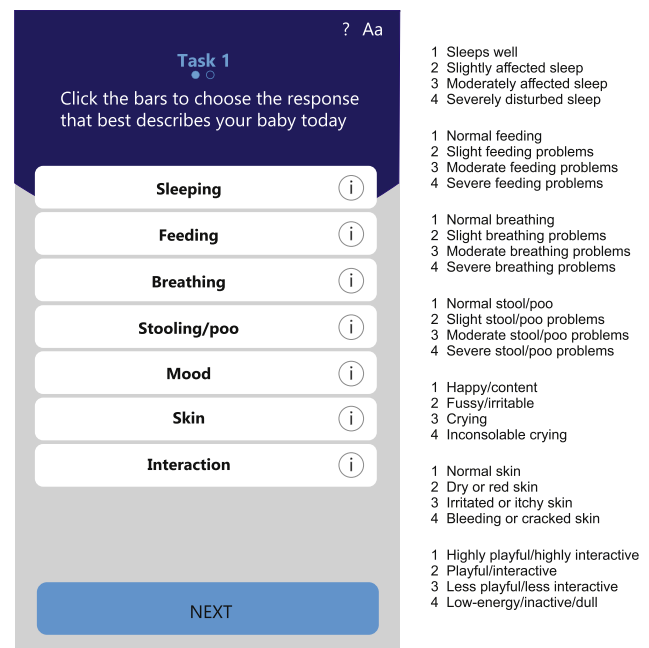


Building on that work, the aim of the present study is to explain how we derived the weights for the different levels, which are necessary to calculate the final values for infants' health status.

## Methods

### Instrument

The IQI includes 7 health items. These are sleeping, feeding, breathing, stooling/poo, mood, skin, and interaction. Each item consists of 4 levels, most of which are ranked by severity. For instance, the levels for sleeping are 1, sleeps well; 2, slightly affected sleep; 3, moderately affected sleep; and 4, severely disturbed sleep. We developed a mobile application to administer IQI; its usability was tested on parents and further improved in light of their opinions (Figure 1). For each health item, parents or other primary caregivers can select the level that best applies to their infant. In this way they "construct" an IQI health state that forms an overall health description expressed in 7 digits (eg, 3231421).

### Samples

Participants were recruited through a market research company (Survey Sampling International, SSI). They were members of the general population or primary caregivers of infants aged 0 to 3 years from China (only Hong Kong), the United Kingdom, and the United States. Clear instructions were given to all participants, and those who fully completed the survey received a small financial compensation from SSI. The rewards were defined by the company's (SSI) internal agreements with the groups of respondents. Whereas the instrument targets infants up to 1 year, in the survey we chose to include primary caregivers of 2- and 3-year-olds as well to be able to recruit a larger sample. We assumed that the caregivers could recollect their experiences of the first year of their infant's life quite easily.

**Figure 2.** Example of a discrete choice task (screenshot).



## Please indicate which state is better

Suppose that an infant's first year of life is spent mainly in either State A or State B and that its health is uncertain afterwards.

(1 of 10)

| | State A | State B |
|---|---|---|
| Feeding | normal feeding | normal feeding |
| Skin | normal skin | normal skin |
| Interaction | less playful/less interactive | playful/interactive |
| Breathing | moderate breathing problems | slight breathing problems |
| Stool/poo | normal stool/poo | normal stool/poo |
| Sleeping | slightly affected sleep | severely disturbed sleep |
| Mood | fussy/irritable | inconsolable crying |

To gain a better understanding of the extreme IQI health states, a separate study was conducted for which a smaller sample was recruited from the general population of the United States. The Medical Ethics Review Committee at the University Medical Center of Groningen issued a waiver for this study because the pertinent Dutch Legislation (the Medical Research Involving Human Subjects Act) does not apply to noninterventional studies (METc2017.115).

### Valuation Task

During the online survey, 10 discrete choice (DC) scenarios, each consisting of a pair of hypothetical IQI health states, were displayed on a computer screen and respondents were requested to indicate which one they thought was better (Figure 2). The order of the items (eg, sleeping, breathing, and interaction) was randomized for every respondent. Before each task, respondents were instructed about two assumptions: the health states presented in the task would occur in the first year of life and what would happen after that year was uncertain.

### Design

Based on the IQI classification system, a total of $4^7$ (16 384; 7 items with 4 levels) health states were possible. Consequently 134 209 536 ([{16 384 × 16 384} − 16 384]/2) unique pairs of IQI health states can be generated. In view of the results of an initially created efficient design and a previously conducted discrete choice study,[20] we determined that 200 pairs would be sufficient for the current study. In discrete choice modeling, a total of 50 to 60 observations per response task would generally be considered sufficient. Therefore, the minimum number of observations for 200 response tasks would be 10 000. Because every respondent would be presented with 10 response tasks (pairs), the required number of respondents is calculated at 1000. The selection of the 200 pairs from this large pool was based on three criteria. First, comparisons containing a dominant health state, that is, one with all items at a better level than the comparator state (eg, 222222 vs 3333333), were excluded from the task because they would not yield relevant information. Second, to facilitate the comparison of

the health states, pairs with some overlap were selected. Specifically, we included pairs that varied on 4 items and overlapped on 3 (Figure 2: only 4 of the items vary between infants A and B). Of the 4 items that varied, 2 represented better-off item levels in alternative A than alternative B, and 2 represented worse-off item levels in alternative A than in alternative B. The third criterion was that, at least in half of the tasks, the maximum difference in item levels between the health states was set to 1. For example, level 2 could be compared with levels 1 and 3 but not to level 4. The remaining set of tasks could comprise differences greater than 1. In this way, we reduced the number of comparisons containing health states that were very different from each other. No checks were built into the experimental design to identify respondents whose choices suggested attentional failures or a poor level of engagement or understanding. If such response behavior were present in the data, it would reduce the design's statistical efficiency (variability of parameter estimates rises; standard errors increase) rather than bias the results of the analysis.

A small additional study was conducted to elicit responses to extreme worse states. Respondents were asked to compare two IQI health states whereby all items were described in terms of the worst levels (ie, severe problems), except for one item with level 3. In total, 21 such health pairs were possible. The design for the main and the additional study was prepared in MATLAB.[21]

### Analyses

The coefficients for the IQI item levels were estimated with a conditional logit model (Stata, clogit). The first level (ie, no problems) of each health item was taken as the reference category. The coefficients for the remaining 3 levels were estimated using 21 dummy variables (7 × 3).

The value of a health state $j$ for individual $i$ is denoted by $V_{ij}$. It is assumed that $V_{ij}$ is a linear combination of the levels on the health items plus an error term $\varepsilon_{ij}$ for the individual. The model specification is

$$V_{ij} = \sum_{j=1}^{n} \beta x_{ij} + \varepsilon_{ij} \qquad (1)$$

**Table 1.** Demographics of the main study sample

| Demographic characteristics | General population (n = 1409) | Primary caregivers (n = 1229) | Total sample (n = 2638) |
|---|---|---|---|
| Country | | | |
| China | 421 | 407 | 828 (31%) |
| UK | 516 | 404 | 920 (35%) |
| USA | 472 | 418 | 890 (34%) |
| Male sex | | | |
| China | 38.5% | 17.2% | 28.0% |
| UK | 49.0% | 1.5% | 28.3% |
| USA | 46.0% | 1.2% | 24.9% |
| Age, years | | | |
| Min | 18 | 18 | 18 |
| Max | 65 | 60 | 65 |
| Mean | 41 | 33 | 37 |
| Median | 39 | 33 | 35 |
| China | 34 | 33 | 33 |
| UK | 44 | 33 | 36 |
| USA | 43 | 32 | 35 |

where $\beta$s represent a vector of 21 regression coefficients and $x_{ij}$ a vector of 21 binary dummy explanatory variables ($x^{\delta\lambda}$), where $\lambda$ = 2, 3, and 4 indicate levels 2, 3 and 4 of each of the 7 items ($\delta$ = 1, 2, …, 7) of a health state. In the case of IQI, $x^{32}$, for example, represents the second level (*slight problems*) of the third item (*breathing*). For a given health state, $x^{\delta\lambda}$ = 1 if item $\delta$ is at level $\gamma$ and $x^{\delta\lambda}$ = 0 otherwise.[22] To assess whether the models produced for each of the study samples (general population, primary caregivers) were comparable, we used the heteroscedastic conditional logit model (Stata, clogithet) to test the null hypothesis that parameters are equal across the two groups and to estimate a scale parameter.

All computations and the visualization of the results were carried out using a combination of the following software: Stata, R programming language,[23] and SigmaPlot.

## Results

### Samples

In total, 2638 respondents were recruited from China (n = 818), the United Kingdom (n = 920), and the United States (n = 890) in the main study. Slightly more than half (n = 1409) of this sample consisted of members of the general population; the remainder (n = 1229) comprised primary caregivers of an infant. The mean age of the respondents was 37 years (median 35 years) with 73% of the total sample consisting of women (Table 1). Both the average age and the proportion of men were substantially higher in the general population subsample; this was expected because a young mother would be the typical primary caregiver. The representativeness of the general population sample can be considered good for the United Kingdom and the United States, and reasonable for China. To conclude this, we have used national statistics and census data for the United Kingdom, the United States, and Hong Kong on median age and sex. For the additional study, a total of 1027 respondents were recruited among members of the general population in the United States. In the latter sample, 49% of the respondents were female. The mean age was 32 years (median: 33 years).

### Coefficients for the Levels of the IQI Items

An analysis was performed on the combined data from the main study (general population part) and the subsequent additional (extreme states) study. It showed that the estimated coefficients after adding the "extreme states" data were no different from the coefficients based on the initial general population study. Therefore, it was decided to merge the data from these two studies.

The items with the highest coefficients were sleeping, feeding, and breathing for both the general population and primary caregivers (Table 2). Coefficients were negative for most of the levels of these items and followed a logical order (ie, *slight problems* < *moderate problems* < *severe problems*). Negative coefficients implied that a particular level was worse than the baseline, which in our study was the first level of each health item. Moreover, the less preferable an item was considered, the higher its coefficient was in a negative direction. Conversely, a positive coefficient implied that a level was considered better than the baseline. In our study, the respondents preferred "better" levels to "worse" levels, as expected, for the items sleeping, feeding, and breathing. In the remaining 4 items (ie, stooling, mood, skin, and interaction), the order of the coefficients was not strictly monotonously decreasing. Although not significant, the coefficient for the third level (*moderate problems*) had a positive coefficient for stooling in the overall sample, indicating that it was more preferable than the baseline level (*no problems*) and also than the second level (*slight problems*). Similarly, for interaction, the second level (*playful/interactive*) had a significant positive coefficient, meaning that it was more preferable than the baseline level (*highly playful/interactive*). No positive coefficients were observed for mood and skin. The response levels for these two health items were qualitative (ie, no logical ordering) in nature, possibly explaining the lack of monotonously decreasing regression coefficients for levels 2-4.

The results were comparable between the general population and the primary caregivers (Table 2). Visual inspection of the coefficients for the two groups (Figure 3) showed that the fitted regression line is close to a slope of 1, which means that the regression coefficients for both groups are rather comparable. This figure also shows that for some items (eg, interaction, breathing) different weights were given by the two groups. The likelihood ratio statistic (clogithet) showed that overall there were no statistically significant differences (LR = 0.06, P = 0.80) between the caregivers and the general population samples, and the scale parameter was small (−0.012). Between the different countries, some minor differences were observed (see Appendix). In China, for example, sleeping was considered the most important item, whereas it was less important in the United Kingdom and the United States. Moreover, compared with the United Kingdom and China, feeding was more important in the United States.

### Values for the IQI Health States

The predicted values of all possible IQI health states (n = 16 384) were calculated separately for the general population (Table 2, column 2) and the primary caregivers (Table 2, column 5; Figure 4). The values above 0 were due to the positive coefficients for health items such as stooling and interaction (Table 2). The IQI health states were valued slightly more negatively by the caregivers (−1.89) than the general population (−1.51) (Figure 5).

## Discussion

In previous work, we described the selection of the health items included in the IQI, a generic instrument for assessing HRQoL in infants.[19] For the present study, we built on this work and explained how the weights for the different levels of the health items were derived using a discrete choice methodology. In a series of tasks, primary caregivers from China, the United

**Table 2.** Parameter estimates for the levels of the 7 IQI health items separately for the general population and primary caregivers

| Item level | General population | | | Primary caregivers | | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | Significance | Coefficient | SE | Significance |
| Sleeping (2) | −0.289 | 0.04 | 0.000 | −0.246 | 0.04 | 0.000 |
| Sleeping (3) | −0.328 | 0.04 | 0.000 | −0.403 | 0.05 | 0.000 |
| Sleeping (4) | −0.868 | 0.05 | 0.000 | −0.774 | 0.06 | 0.000 |
| Feeding (2) | −0.221 | 0.04 | 0.000 | −0.158 | 0.05 | 0.001 |
| Feeding (3) | −0.225 | 0.04 | 0.000 | −0.162 | 0.05 | 0.001 |
| Feeding (4) | −0.713 | 0.05 | 0.000 | −0.683 | 0.06 | 0.000 |
| Breathing (2) | −0.173 | 0.05 | 0.000 | −0.395 | 0.05 | 0.000 |
| Breathing (3) | −0.374 | 0.04 | 0.000 | −0.585 | 0.06 | 0.000 |
| Breathing (4) | −0.946 | 0.05 | 0.000 | −1.046 | 0.06 | 0.000 |
| Stooling (2) | −0.015 | 0.04 | 0.725* | −0.100 | 0.04 | 0.025 |
| Stooling (3) | 0.076 | 0.05 | 0.106* | −0.039 | 0.05 | 0.449* |
| Stooling (4) | −0.248 | 0.05 | 0.000 | −0.268 | 0.07 | 0.000 |
| Mood (2) | −0.501 | 0.04 | 0.000 | −0.509 | 0.05 | 0.000 |
| Mood (3) | −0.391 | 0.04 | 0.000 | −0.380 | 0.05 | 0.000 |
| Mood (4) | −0.672 | 0.05 | 0.000 | −0.613 | 0.06 | 0.000 |
| Skin (2) | −0.146 | 0.04 | 0.001 | −0.166 | 0.05 | 0.000 |
| Skin (3) | −0.194 | 0.04 | 0.000 | −0.120 | 0.05 | 0.021 |
| Skin (4) | −0.422 | 0.05 | 0.000 | −0.416 | 0.06 | 0.000 |
| Interaction (2) | 0.113 | 0.04 | 0.011 | 0.170 | 0.05 | 0.000 |
| Interaction (3) | −0.118 | 0.04 | 0.005 | −0.190 | 0.05 | 0.000 |
| Interaction (4) | −0.185 | 0.04 | 0.000 | −0.361 | 0.06 | 0.000 |

IQI indicates Infant Quality of life Instrument.
*Coefficients not significantly different from the baseline category (level 1).

Kingdom, and the United States and members of the general population from the United States were presented with two hypothetical IQI health states and asked to choose the one they thought was better.

Among the 7 IQI health items, sleeping, feeding, and breathing were found to have the highest impact on the health-state values, followed by mood and skin. Although the coefficients for mood and skin were not ordered as initially hypothesized (level 3 was more preferable to level 2), this irregularity is possibly due to the qualitative nature of their levels (eg, mood: *fussy/irritable*, *crying*; skin: *dry or red skin, irritated or itchy skin*). The unexpected ordering of the coefficients for the levels of stooling should be viewed with some reservations because these are not statistically significant. The positive coefficient for the second level of interaction (*playful/interactive*) suggests that respondents associated the first level (*highly playful/highly interactive*) with hyperactivity. The results were comparable among the three countries. Nevertheless, when compared with the United Kingdom and the United States, China's results were less generalizable to the whole country population because the study was conducted online and in English. That format could indeed lead to a selection bias by excluding those who do not speak English or do not have access to the internet, resulting in an overrepresentation of individuals with higher socioeconomic status. Nonetheless, it should be kept in mind that the present study was largely intended as a proof of principle to demonstrate how the process of valuation can take place and to provide a first value set for the IQI. From that perspective, the generalizability of the results is of minor importance at this stage.

The values generated with the IQI are on an interval scale whereby the relative differences between two values for two different health states are meaningful, irrespective of their location on the value scale. For example, if an infant's value increases from −1.4 to −1.2, this increase is identical to an increase from −2.8 to −2.6. Nevertheless, quality-adjusted life years (QALYs), which are necessary inputs in cost-effectiveness analyses, cannot be calculated with raw values. Those values should first be transformed into utilities ranging from 0.0 (dead) to 1.0 (full health).[24–27] To achieve this, another discrete choice study including the option "worse than dead" is planned for the near future. In this next study, the DC results (regression coefficients) from caregivers will be normalized by anchoring them on dead = 0. The anchor point will be derived from a sample of the general population that will perform an identical DC, but supplemented with a "dead" preference option. Moreover, the tool is planned to be used early next year, together with another tool, in a clinical trial involving a pediatric population. Even if the comparison of results will be not be straightforward because the two tools were built in different scales, it will give a first indication of how the IQI performs and on the eventual adjustments that have to be implemented.

In this study, the DC methodology was used to derive values for the IQI health states. Nevertheless, we have also developed an innovative new measurement model[28,29] for deriving health-state values by using a different value judgment task. Instead of presenting pairs of hypothetical IQI states, primary caregivers are first asked to use the IQI to classify the health condition of their own infant. Then, they are asked to compare this condition with a small

**Figure 3.** Plot of the coefficients for the general population and primary caregivers (thick line is estimated slope of the line of best fit; dashed line is with slope of 1).
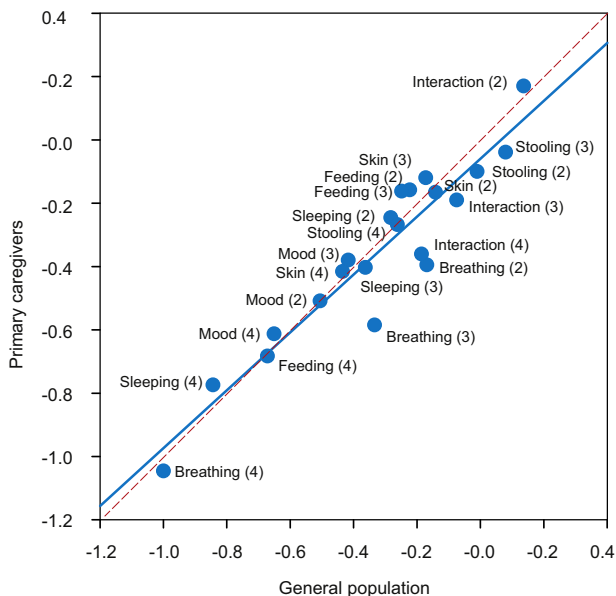


number of IQI health states, which are slightly different from the state of their infant, and to indicate whether these states are worse or better. In this measurement model, the descriptive content and the preference tasks are integrated. Moreover, given

**Figure 4.** Predicted health-state values (discrete choice analysis) for the IQI (Infant Quality of Life Instrument) in the general population and among primary caregivers for all possible health states (gray dots) and for those used in the study (blue dots). Note: Health-state values above 0 are due to the positive coefficients for some item levels.
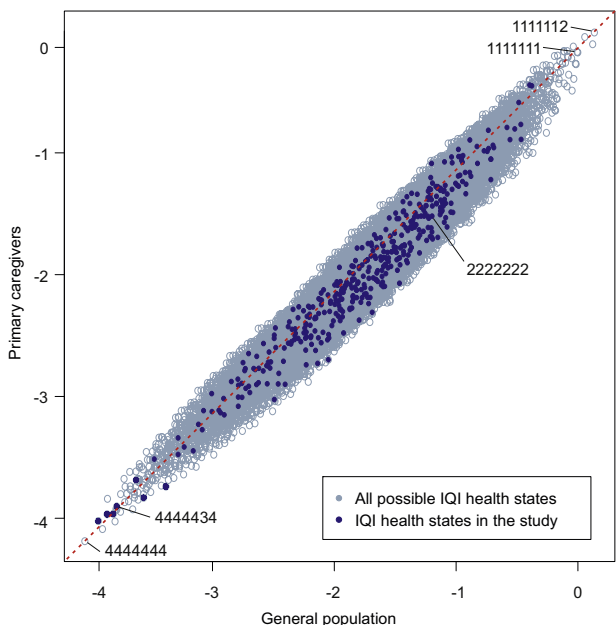


**Figure 5.** Kernel density plots of predicted IQI (Infant Quality of Life Instrument) health states for the general population and the primary caregivers.
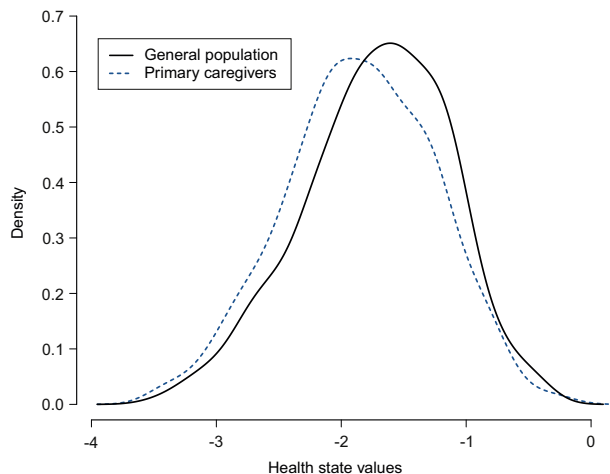


the interactive nature of this framework, an online administration (www.healthsnapp.info)—as opposed to paper-and-pencil testing—is practically a necessity and would make data collection easier and most likely more valid.

The development of a new measurement tool requires various steps. So far, we have explained how the items to be included in the tool were selected and how the health-state values were generated with a DC methodology. The next steps will be the normalization of these values to utilities (0-1) and the test of the IQI in a clinical population of infants.

## Conclusions

Although in the past attempts have been made to value health status in pediatric populations, the existing measures still suffer from considerable limitations because their content is not necessarily relevant for 0- to 1-year-old infants, they do not produce a single score capturing overall HRQoL, and they are not preference based. To our knowledge, the HRQoL instrument described in this study, the IQI, is the first generic preference-based tool to value health states that are relevant for 0- to 1-year-old infants. Its development marks an important step toward a substantiated approach for obtaining health state utilities that are relevant in the first year of life.

## Acknowledgment

## Supplementary Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2018.12.009.

# REFERENCES

1. World Health Organization. (n.d.). Frequently asked questions. http://www.who.int/suggestions/faq/en/. Accessed December 7, 2018.

2. Food and Drug Administration. *Guidance for Industry Use in Medical Product Development to Support Labeling Claims Guidance for Industry, Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.* Silver Spring, MD: FDA; 2009.

3. NICE National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2013. https://www.nice.org.uk/process/pmg9/chapter/foreword. Accessed December 7, 2018.

4. Raat H, Landgraf J, Oostenbrink R, Moll H, Essink-Bot M. Reliability and validity of the Infant and Toddler Quality of Life Questionnaire (ITQOL) in a general population and respiratory disease sample. *Qual Life Res.* 2006;16(3):445–460.

5. Landgraf J, Vogel I, Oostenbrink R, van Baar BM, Raat H. Parent-reported health outcomes in infants/toddlers: measurement properties and clinical validity of the ITQOL-SF47. *Qual Life Res.* 2013;22(3):635–646.

6. Fekkes M, Theunissen NC, Brugman E, et al. Development and psychometric evaluation of the TAPQOL: a health-related quality of life instrument for 1–5-year-old children. *Qual Life Res.* 2000;9(8):961–972.

7. Bunge EM, Essink-Bot ML, Kobussen MP, et al. Reliability and validity of health status measurement by the TAPQOL. *Arch Dis Child.* 2005;90(4):351–358.

8. Krabbe PFM. *The Measurement of Health and Health Status: Concepts, Methods and Applications from a Multidisciplinary Perspective.* San Diego: Elsevier/Academic Press; 2016.

9. Stevens K. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics.* 2012;30(8):729–747.

10. 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(6):875–886.

11. Stevens K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Qual Life Res.* 2009;18(8):1105–1113.

12. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions: Health Utilities Index. *Pharmacoeconomics.* 1995;7(6):503–520.

13. Torrance GW, Boyle MH, Horwood SP. *Application of Multi-Attribute Utility Theory to Measure Social Preferences for Health States.* Research and Working Paper Series No. 189. Hamilton: McMaster University; 1982.

14. Boyle MH, Torrance GW, Sinclair JC, Horwood SP. Economic evaluation of neonatal intensive care of very-low-birth-weight infants. *N Engl J Med.* 1983;308(22):330–331.

15. Torrance GW, Feeny D, Furlong W, et al. Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2. *Med Care.* 1996;34(7):702–722.

16. Thorrington D, Eames K. Measuring health utilities in children and adolescents: a systematic review of the literature. *PLoS One.* 2015;10(8):e013567.

17. Kwon J, Kim SW, Ungar WJ, et al. A systematic review and meta-analysis of childhood utility values. *Med Decis Making.* 2017;38(3):277–305.

18. Montgomery S, Kusel J. The prevalence of child-specific utilities in NICE appraisals for paediatric indications: rise of the economic orphans? *Expert Rev Pharmacoecon Outcomes Res.* 2016;16(3):347–350.

19. Jabrayilov R, Van Asselt ADI, Vermeulen KM, et al. A descriptive system for the Infant health-related Quality of life Instrument (IQI): Measuring health with a mobile app. *PLoS One.* 2018;13(8):e0203276.

20. Selivanova A, Groothuis-Oudshoorn KGM, Krabbe PFM. Does inclusion of interactions result in higher precision of estimated health state values? *Value Health.* 2018;21(12):1437–1444.

21. The MathWorks. *MATLAB User's Guide.* Natick, MA: The MathWorks, Inc; 1993.

22. McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. *J Health Econ.* 2006;25:418–431.

23. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing; 2013.

24. Fanshel S, Bush JW. A health-status index and its application to health-services outcomes. *Oper Res.* 1970;18:1021–1066.

25. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med.* 1977;296(13):716–721.

26. Froberg DG, Kane RL. Methodology for measuring health-state preferences-II: Scaling methods. *J Clin Epidemiol.* 1989;42(5):459–471.

27. Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health.* 2000;21:587–611.

28. Krabbe PFM. A generalized measurement model to quantify health: the multi-attribute preference response model. *PLoS One.* 2013;8(11):e79494.

29. Groothuis-Oudshoorn CGM, van der Heuvel E, Krabbe PFM. A preference-based item response theory model to measure health: concept and mathematics of the multi-attribute preference response model. *BMC Med Res Methodol.* 2018;18:62.