ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Preference-Based Assessments

Generic Health-Related Quality of Life Utility Measure for Preschool Children (Health Utilities Preschool): Design, Development, and Properties

William Furlong, MSc, Charlene Rae, PhD, David Feeny, PhD, Satvinder Ghotra, MD, Vicky R. Breakey, MD, Teresa Carter, MD, Nikhil Pai, MD, Eleanor Pullenayegum, PhD, Feng Xie, PhD, Ronald Barr, MD*

ABSTRACT

Objectives: Health Utilities Preschool (HuPS) was developed to fill the need for a generic preference-based measure (GPM) applicable in early childhood. A GPM has all the properties for higher-order summary measures, such as quality-adjusted life-years, required to inform important policy decisions regarding health and healthcare services.

Methods: Development was in accordance with published standards for a GPM, statistical procedures, and modeling. HuPS incorporates key components of 2 existing measurement systems: Health Status Classification System for Preschool Children and Health Utilities Index Mark 3 (HUI3). The study included a series of 4 measurement surveys: definitional, adaptational, quantificational, and evaluational health-related quality of life (HRQL). HuPS measurements were evaluated for reliability, validity, interpretability, and acceptability.

Results: Definitional measurements identified 8 Health Status Classification System for Preschool Children attributes in common with HUI3 (vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain and discomfort), making the HUI3 scoring equation commensurate with HuPS health states. Adaptational measurements informed the content of attribute-level descriptions (n = 35). Quantificational measurements determined level scoring coefficients. HRQL scoring inter-rater reliability (intraclass correlation coefficient = 0.79) was excellent. Continuity of HRQL scoring with HUI3 was reliable (intraclass correlation coefficient = 0.80, $P < .001$) and valid (mean absolute difference = 0.016, $P = .396$).

Conclusions: HuPS is an acceptable, reliable, and valid GPM. HRQL scoring is continuous with HUI3. Continuity expands the applicability of GPM (HUI3) scoring to include subjects as young as 2 years of age. Widespread applications of HuPS would inform important health policy and management decisions as HUI3 does for older subjects.

Keywords: generic, health measure, health-related quality of life, multiattribute, preschool children, utilities.

VALUE HEALTH. 2022; ■(■):■–■

Introduction

Health Utilities Preschool (HuPS) was conceived to fill a void in coverage of the human lifespan among existing high-quality, comprehensive, individual-level preference-based health measurement systems used to inform important policy decisions regarding the health of, and alternative healthcare services for, clinical and general populations. The void was in early childhood, preschool, considered the most important developmental phase of the human lifespan.[1]

High-quality health measurement systems were considered the leading generic preference-based measures (GPM). Existing GPM lifetime coverage was limited by a lack of construct and content validity for preschool children. A GPM consists of a health status classification system defining comprehensive health states with a matching scoring function for determining preference scores of health-related quality of life (HRQL). Generic refers to being applicable to all, or most, types of people and health issues. The health status classification system should consist of a set of health attributes (ie, constructs or dimensions or domains) with each attribute having a set of descriptive levels of disability (ie, content). GPM scores of HRQL should represent preferences of a general population for the universe of comprehensive states defined by the classification system. There are 2 types of HRQL preference scores: values and utilities. Values are measured under conditions of certainty. Utilities are measured under conditions of uncertainty.[2] The general relationship for converting values to utilities is well described.[3] HRQL scores should have interval scale properties on a scale in which being dead = 0.00 and perfect health = 1.00 and other properties required of higher-order summary measures such as quality-adjusted life-years and health-adjusted life expectancies.[4-6] The importance of such

*Ronald Barr considered as senior author.

measures for assessing HRQL and quality-adjusted survival in children has been demonstrated by systematic reviews of published studies.[7-9]

HuPS was designed to be a stand-alone GPM for children at 2 to 4 years of age commensurate and continuous with an appropriate existing GPM for older subjects referred to as the foundational GPM. Commensurate refers to commonality of attributes for health status classification.[10] Continuous refers to HRQL scoring from HuPS and the foundational GPM measurements being in agreement with mean difference close to zero for subjects at the age at which the measurement systems intersect.[11] The foundational GPM was to have been used frequently in clinical studies and acceptable to national health technology assessment agencies with attributes relevant for young children. Attribute-level descriptions were to be age appropriate and developmentally relevant for content validity.

There were 2 contenders for being the HuPS foundational GPM: EQ-5D and Health Utilities Index (HUI) Mark 3 (HUI3) that, combined, accounted for 73% of all relevant published studies.[12,13] Both EQ-5D and HUI3 are preferred by national health technology assessment agencies for use in economic evaluation, especially cost-utility analysis.[14] EQ-5D is not recommended for assessing children.[15,16] The youth version of EQ-5D, EQ-5D-Y, has a minimum recommended assessment age of 4 years, and adult EQ-5D value sets (either 3-level version or 5-level version) are not continuous with EQ-5D-Y value sets.[17,18] Alternatively, HUI3 is appropriate for assessing children, adolescents, and adults; recommended for subjects at $\geq$ 5 years of age; and the most frequently used (40%) GPM for assessments of children.[8,19,20] Furthermore, the Health Status Classification System for Preschool Children (HSCS-PS) is based on Health Utilities Index Mark 2 and HUI3 classification systems with level descriptions adapted for preschool children.[21] A subset of the HSCS-PS attributes is entirely commensurate with the full set of HUI3 attributes. The HUI3 scoring function (ie, model) is a multiplicative function based on a multiattribute utility theory (MAUT) model that allows for calibration of attribute-level coefficients without estimation of the overall scoring equation structure with constants.[22,23] Therefore, the commensurate attribute effect includes the HUI3 scoring function structural form and constants, leaving a relatively small number ($n < 30$) of coefficients to be estimated for age-adapted attribute-level descriptions in a HuPS scoring function. For these reasons, HUI3 should be the foundational GPM for HuPS.

GPM instrument development should be rigorous and reported in accordance with relevant standards for development and evaluation of health measurement instruments, health measurement properties, guideline criteria for quality-of-life instruments, multiattribute utility-based instrument valuation elements, and ISPOR-Society for Medical Decision Making best practices.[24-30] The standards define important criteria and terminology and demand a methodical approach to reporting. This article provides detailed descriptions of HuPS development steps, including evaluations of results from each step. The steps are transparent and in logical order. The article includes highly technical terms, methods, and results to satisfy demanding methodologists. Less technical summary statements should satisfy the interests of other readers. Measurement results are evaluated for reliability, validity, interpretability, and burden. Reliability refers to the extent to which a measure is free from random error and consistent or reproducible.[28] Validity refers to the extent to which a measure has appropriate content and is associated with theoretical constructs.[28] Evaluations of scores from the final measurement system focus on continuity between HuPS and HUI3 HRQL scores.

Methods

The study was approved by the Hamilton Health Sciences/McMaster University Health Sciences Research Ethics Board and Izaak Walton Killam Health Centre Research Ethics Board.

Design

The overall design components and process flow are shown in Figure 1. Previously published resources were the HSCS-PS and HUI3 multiattribute utility function. The HUI3 multiattribute utility function consists of a scoring equation with constants and 8 sets of level scoring coefficient variables, 1 set for each attribute.[4] HuPS was developed in 4 major steps. In steps 1 and 2, the HuPS attributes and level descriptions were derived from HSCS-PS in 2 measurement surveys: definitional and adaptational. The resulting set of attribute and level descriptions defined the HuPS health status classification system. Step 3 determined the level scoring coefficients for the HuPS health status classification system as being the mean of value measurements collected in the quantificational survey. Step 3 also involved converting the mean values to utilities. These utilities are attribute-level coefficients of the HuPS multiattribute utility function. The finalized HuPS multiattribute utility function consists of the HUI3 scoring equation with constants and the 8 sets of HuPS utility coefficients. HuPS is the combination of the HuPS classification system and HuPS multiattribute utility function. Step 4 collected HuPS and HUI3 measurements to determine continuity of HRQL scores between the 2 systems.

Participants

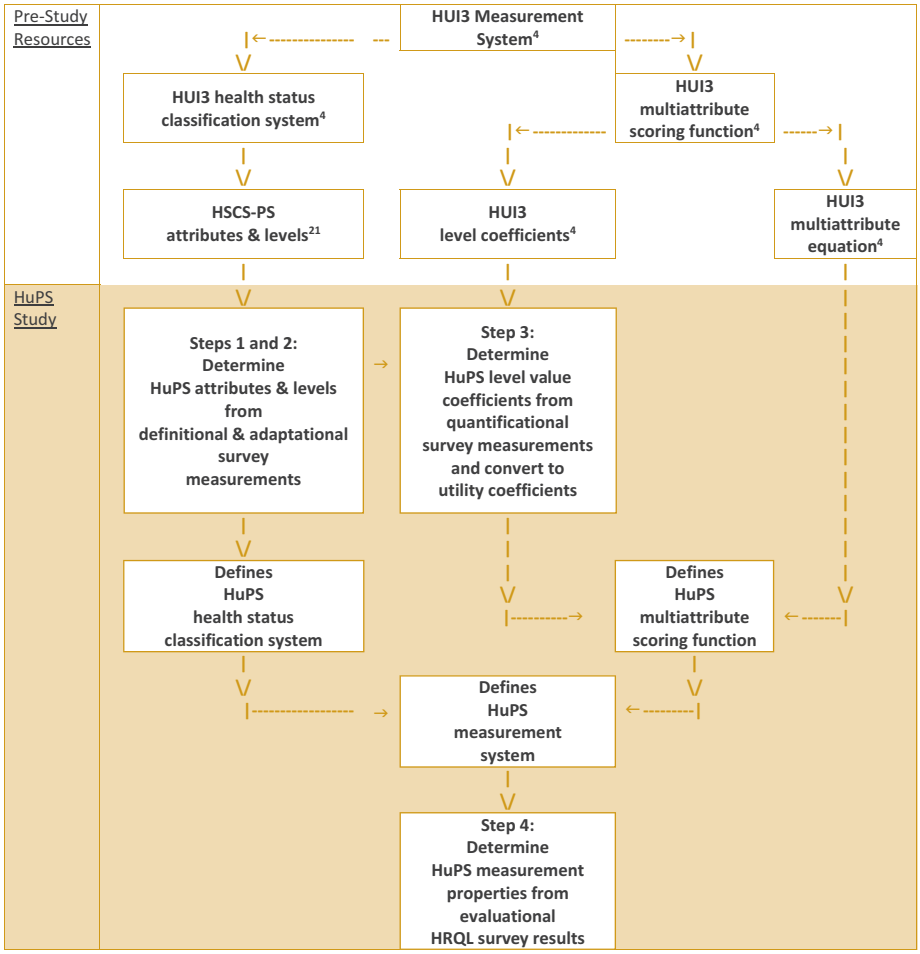
Respondent and subject sample populations were selected to be most appropriate for survey objectives. Definitional survey respondents were methodologists with expertise in assessing health status and utility measurement. Adaptational survey respondents were parents and pediatricians representing heterogeneity of expertise in assessing health status, as recommended.[31] Quantificational survey respondents were parents of children at 2 to 6 years of age in the general population, representing the community perspective in assessing health status as required for a GPM. Evaluational HRQL survey respondents were parents of subjects and subjects' healthcare providers to assess inter-rater reliability between 2 important perspectives. Evaluational HRQL survey subjects were 2 to 6 years of age sampled from general and clinical populations, to provide a sample with heterogeneity of disability type and severity. Five- and 6-year-old subjects were included to evaluate continuity of HuPS and HUI3 scoring.

Clinical survey subjects were patients of McMaster Children's Hospital and Izaak Walton Killam Health Centre. Subjects in the general population survey sample were from the McMaster Family Medicine Research List or local community sports and recreational clubs. Clinician survey respondents included family doctors, developmental pediatricians, pediatric gastroenterologists, pediatric oncologists, nurses, and nurse practitioners.

Measures and Measurements

HSCS-PS has 12 attributes, each with 3 to 5 levels, and a total of 49 levels: vision, hearing, speech, mobility, dexterity, self-care, emotion, learning and remembering, thinking and problem solving, pain and discomfort, general health, and behavior.[21] HUI3 has 8 attributes, each with 5 or 6 levels and a total of 45 levels: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain.[4] The HUI3 multiattribute utility function is a mathematical

Figure 1. HuPS study design flow diagram: resources, activities, and results.



HRQL indicates health-related quality of life; HSCS-PS, Health Status Classification System for Preschool Children; HUI3, Health Utilities Index Mark 3; HuPS, Health Utilities Preschool.

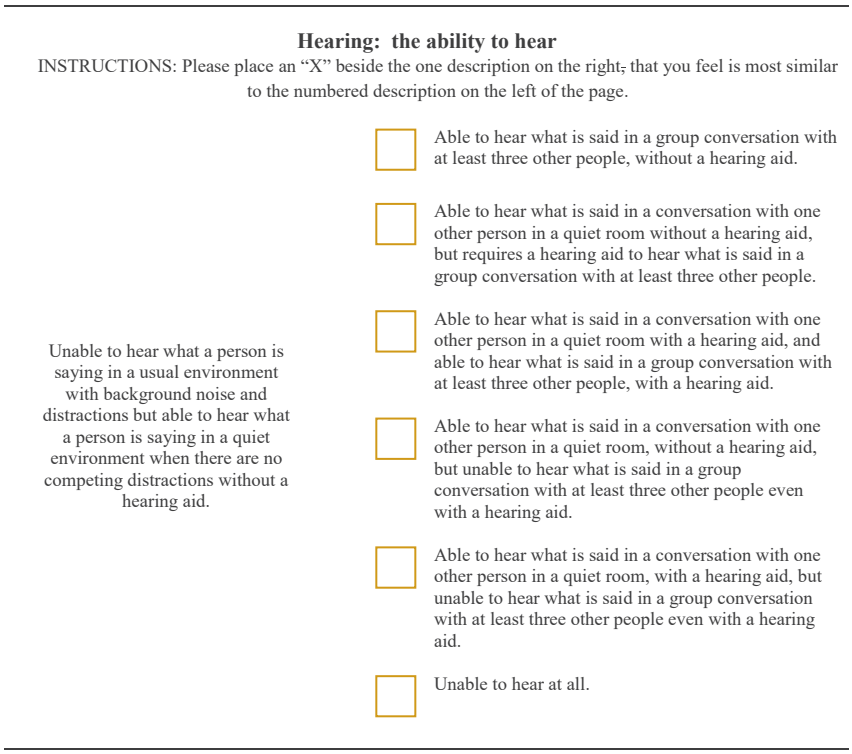
formula, derived from preference measurements in the general population, for calculating a HRQL utility score of each unique 8-element health state defined by the HUI3 classification system (N = 972 000).

Measurements were from 4 cross-sectional surveys.

1. Definitional. Respondents paired attributes and attribute levels in common between HSCS-PS and HUI3 to avoid unnecessary measurements in the adaptational and quantificational surveys. HSCS-PS levels unpaired by definition were measured in the subsequent surveys.
2. Adaptational. Respondents independently paired a HSCS-PS level description to 1 level description from the set of level descriptions for the paired HUI3 attribute, using a customized pairing questionnaire formatted for self-completion (Fig. 2), to inform decisions about adaptation of levels for scoring level coefficients. Each page of the questionnaire presented 1 HSCS-PS level description on the left side of the page. The right side of the page listed the full set of level descriptions for the paired HUI3 attribute. Respondents were asked to place an "X" beside the one description on the right, which was most similar to the description on the left of the page. Adaptational survey results were reviewed by the investigators to adapt level descriptions for utility scoring. Adaptation was minimized by requiring full agreement among investigators to conserve previously established HSCS-PS measurement properties. The adapted version of a HSCS-PS level description was defined as the HuPS level description.
3. Quantificational. Respondents independently paired, using the same process as the adaptational survey, a HuPS level description to a HUI3 level description with a pre-existing utility coefficient from the HUI3 multiattribute utility function. The HUI3 utility coefficient provided a coefficient measurement for the HuPS level description.
4. Evaluational HRQL. Respondents completed 2 HRQL questionnaires for subject children: a custom 8-item multiple-choice self-complete health status questionnaire based in part on the HuPS health status classification system and the standard HUI multiple-choice self-complete questionnaire for proxy-assessment with a "past 1-week" recall assessment duration. The HUI questionnaire collected Health Utilities Index Mark 2 and HUI3 health status measurements.[20]

Survey measurements were evaluated for reliability, validity, interpretability, and acceptability. Research study staff were available to survey respondents for clarification of measurement tasks.

Figure 2. Pairing question example: for Health Status Classification System for Preschool Children hearing level 3.



Analysis

Categorical scale measurements were summarized by proportions (percent) and modes for central tendency. Interval scale measurement distributions were described by summary statistics (eg, mean, SD, and 95% confidence bounds). Consistency was evaluated by modal percent for categorical measurements, and reliability was evaluated by single measurement intraclass correlation coefficient (ICC) for interval scale measurements. A measure has interval scale properties if all the intervals on the scale are equal in size and any zero is arbitrary. Mode is the most frequently occurring observation. Hypotheses were tested using chi-square or analysis of variance. The standard for interpreting consistency and reliability were as follows: poor $<$ 0.40; moderate = 0.40-0.59; good = 0.60-0.75; and excellent $\geq$ 0.75.[32] Differences in mean HRQL scores of $\geq$ 0.03 were interpreted as clinically important.[20] Acceptability of the HuPS classification system questionnaire was measured by Flesch-Kincaid Grade Level readability, with a grade of approximately 8 being appropriate for the general public.[33]

Adaptational survey level pairing consistency was measured by modal percent. A mode less than excellent ($\leq$ 75%) initiated a structural review of the number and descriptions for levels within an HuPS attribute. Structural review evaluated 6 factors: disability severity concepts, number and ordinality of levels, best level being absolute lack of disability, worst level being absolute disability, intermediate levels describing precise severities of disability, and consistency of phrasing across the set of levels.

Quantificational survey measurements were value type pairings of each HuPS level to an HUI3 level and, by extension, to an HUI3 level scoring coefficient. For example, a respondent pairing HuPS hearing level 3 and HUI3 hearing level 4 descriptions would, by extension, be pairing the HuPS hearing level 3 to HUI3 hearing level 4 scoring coefficient = 0.80. The mean of the HUI3 level scoring coefficients across respondents was the value coefficient for the specific HuPS descriptive level. Three types of validity for HuPS value coefficients were evaluated by hypothesis testing: anchor validity for level 1 coefficients = 1.00, rank validity for level coefficients within attributes in the order listed by Saigal et al,[21] and ordinal validity for significant differences among level coefficients within attributes. The value coefficients were used to calculate HRQL value scores of HuPS comprehensive health states from the evaluational survey to calculate the value-to-utility conversion function.

Both value and utility scales are anchored by 0.00 = dead and 1.00 = perfect health. For intermediate states, measured utility scores tend to be higher than measured value scores. The difference is most pronounced midscale. The precise relationship between value and utility measurements is context dependent (ie, study specific) and described by a power function.[3] HuPS value scores of HRQL and HUI3 utility scores of HRQL from the evaluational survey were paired, by subject, to calculate the study-specific value-to-utility conversion power function. The conversion function was applied to all HuPS value coefficients to calculate a full set of HuPS utility scoring coefficients. HuPS utility scores of HRQL were calculated using the HuPS utility coefficients and HUI3 scoring equation with constants.

HuPS scoring performance was assessed between 2 types of raters (ie, parents vs clinicians) and relative to HUI3 scores. Parent and clinician HuPS scores of HRQL were hypothesized to have excellent inter-rater reliability and validity (ie, no significant mean difference). Continuity (ie, predictive) reliability and validity between HuPS and HUI3 scores of HRQL were expected to be excellent.

Table 1. Characteristics of survey respondents and subjects.

Sample characteristics		Level pairing surveys and respondents			Evaluational HRQL survey Parent respondents	
		Adaptational		Quantificational		
Type	Category	Parents (n = 10)	Clinicians (n = 22)	Parents (n = 32)	Primary (n = 111)*	Continuity (n = 67)†
Group (%)	Autism spectrum disorder				6.3	7.5
	Cancer	10.0	59.1		27.0	20.9
	Cerebral palsy	20.0	13.6			
	Neuromuscular				14.4	17.9
	Preterm				29.7	32.8
	Short bowel syndrome				5.4	4.5
	Family practice	70.0	27.3			
	General population			100.0	17.1	16.4
Sex (%)	Female	90.0	72.7	78.1	38.0	40.0
	Male	10.0	27.3	21.9	62.0	60.0

HRQL indicates health-related quality of life.
*For primary reliability and validity analyses.
†For continuity primary reliability and validity analyses.

Results

Definitional survey respondents were 3 authors. Adaptational, quantificational, and evaluational survey subject characteristics are presented in Table 1. Evaluational survey subjects had a mean age of 4.1 (SD = 1.3) years and were assessed by parents (n = 111) and clinicians providing treatment to children with cancer, neuromuscular diseases, prematurity, and short bowel syndrome (n = 48). Parental assessments of health status were collected by HuPS (n = 111) and HUI (n = 67) questionnaires.

Definitional survey respondents identified 8 of the 12 HSCS-PS attributes as pairing to HUI3 attributes: vision, hearing, speech, mobility (ambulation), dexterity, emotion, thinking and problem solving (cognition), and pain and discomfort (Table 2, columns 1 and 2). According to MAUT, having attributes fully in common makes it appropriate to use the existing equation structure and attribute constant terms with age-adjusted level coefficient variables for scoring health states defined by age-adjusted level descriptions.[22,23] Respondents also identified 11 HSCS-PS attribute-level descriptions that each paired, by content definition, with an HUI3 classification system level description (Table 2, column 3, capitalized letters of HUI3 attributes with level codes). The remaining 22 HSCS-PS level descriptions required measurement by adaptational survey respondents.

Adaptational survey modes of HSCS-PS to HUI3 level pairing consistency were moderate or better for all measured levels (Table 2, column 3). HSCS-PS was then further adapted for quantificational measurements as follows. Mobility was relabeled as ambulation. Two attribute levels were added, 1 level to each of vision (equivalent in rank to HUI3 vision level 5) and emotion (equivalent in rank to HUI3 emotion level 2). Other adaptations improved consistency of wording syntax of level descriptions within attributes (eg, HSCS-PS vision levels 2 and 3) and descriptive focus (eg, HSCS-PS vision level 4). The adapted HSCS-PS, the HuPS classification system, had 8 attributes with each attribute having 3 to 6 levels. HuPS has a total of 35 levels (Table 3). Twenty-three HuPS attribute levels required measurement by quantificational survey respondents (Table 2, column 4).

Quantificational survey statistics for HuPS value coefficient measurements were calculated (Table 2, columns 4-7). The set of value coefficients within each attribute were ordered as hypothesized for rank validity. There were important differences among mean coefficients within all attributes as hypothesized for ordinal validity ($P < .001$) (Table 2, column 8). The value-to-utility conversion function was calculated (n = 67 pairs) to be $u = v^{0.65}$, where u is utility and v is value. Value coefficients, determined by definitional or quantificational survey results, were converted to utility coefficients using the power function, and the HuPS utility coefficients combined with the HUI3 scoring equation form the HuPS multiattribute utility function for calculating HuPS scores of HRQL (Table 4). Level 1 coefficients were 1.00 for each of the 8 attributes, as required by the conventional HRQL scoring scale maximum of perfect health = 1.00.

Evaluational HRQL survey scoring inter-rater reliability (parent vs clinician) was excellent (n = 48, ICC = 0.79, 95% confidence interval [CI] = 0.65-0.88), and inter-rater validity was confirmed with no significant difference of paired scores (n = 48, mean difference = 0.012, SD = 0.145, $P = .582$). Further analyses focused on measurements from the assessment perspective having the largest and most clinically diverse sample of subjects: parental scores of HRQL (n = 111, mean = 0.790, SD = 0.214, minimum = 0.01, maximum = 1.00). Continuity reliability between HuPS and HUI3 scores of HRQL, paired by subjects, was excellent (n = 67, ICC = 0.80, 95% CI = 0.70-0.87, $P < .001$), as hypothesized. Continuity validity was confirmed with no significant differences between HuPS and HUI3 mean HRQL scores (n = 67 pairs, mean HuPS = 0.779, SD = 0.212, mean HUI3 = 0.763, SD = 0.256, mean absolute difference = 0.016, SD of differences = 0.147, $P = .396$). There were important size differences between pairs of HuPS and HUI3 scores within some individual subjects (SD of differences > 0.03). Evaluational HRQL survey respondent burden for the HuPS questionnaire was low for adults, as hypothesized, with a completion time of < 5 minutes and a Flesch-Kincaid Grade Level readability of 6.9 (easily understood by 13- to 14-year-olds). Administrative burden was low: self-complete questionnaires with no specialized resources required for scoring.

Table 2. HuPS attribute and level pairing survey results.

HuPS		Developmental survey					
Attribute	Level	Definitional HUI3 level criteria and adaptational mode (n = 32)	Quantificational HuPS value coefficients				
			n	Mean	SD	95% CB	*P* value
Vision	1	V1(c)		1.00*			
	2	86 (E)	32	0.89	0.04	0.88-0.91	< .001
	3	86 (E)	32	0.85	0.03	0.84-0.86	
	4	93 (E)	32	0.77	0.05	0.74-0.78	
	5	new - V5(c)		0.75*			
	6	V6(e)		0.61*			
Hearing	1	H1(c)		1.00*			
	2	50 (M)	31	0.88	0.09	0.85-0.91	< .001
	3	79 (E)	32	0.81	0.07	0.78-0.83	
	4	100 (E)	32	0.63	0.06	0.61-0.65	
	5	H6(e)		0.61*			
Speech	1	S1(c)		1.00*			
	2	43 (M)	32	0.93	0.04	0.91-0.94	< .001
	3	71 (G)	32	0.89	0.03	0.88-0.90	
	4	93 (E)	32	0.71	0.06	0.69-0.73	
Ambulation	1	100 (E)	31	1.00	0.01	0.99-1.00	< .001
	2	100 (E)	32	0.93	0.02	0.93-0.94	
	3	100 (E)	32	0.85	0.03	0.84-0.86	
	4	71 (G)	32	0.65	0.05	0.63-0.67	
Dexterity	1	D1(e)		1.00*			
	2	D2(e)		0.95*			
	3	D4(e)		0.76*			
	4	D6(e)		0.56*			
Emotion	1	50 (M)	32	1.00	0.01	0.99-1.00	< .001
	2	new	32	0.92	0.02	0.90-0.94	
	3	71 (G)	32	0.76	0.10	0.73-0.80	
	4	86 (E)	32	0.61	0.08	0.58-0.64	
Cognition	1	C1(c)		1.00*			
	2	86 (E)	32	0.91	0.06	0.89-0.93	< .001
	3	100 (E)	32	0.66	0.10	0.61-0.68	
	4	100 (E)	32	0.45	0.11	0.42-0.49	
Pain and discomfort	1	P1(c)		1.00*			
	2	93 (E)	32	0.94	0.08	0.91-0.97	< .001
	3	64 (G)	32	0.66	0.12	0.62-0.71	
	4	93 (E)	32	0.55	0.00	0.55-0.55	
Total (n)	35						

Note. Mode is percent of most frequently paired HUI3 level.
c indicates consensus (100%); CB, confidence bounds; C1, HUI3 cognition level 1; D1, HUI3 dexterity level 1; D2, HUI3 dexterity level 2; D4, HUI3 dexterity level 4; D6, HUI3 dexterity level 6; H1, HUI3 hearing level 1; H6, HUI3 hearing level 6; e, exact wording; (E), excellent pairing agreement; (G), good pairing agreement; HuPS, Health Utilities Preschool; HUI3, Health Utilities Index Mark 3; (M), moderate pairing agreement; *P* value, probability from one-way analysis of variance; P1, HUI3 pain level 1; S1, HUI3 speech level 1; V1, HUI3 vision level 1; V6, HUI3 vision level 6.
*Level not measured and equals HUI3 coefficient by definition.

Discussion

HuPS has all the properties of a stand-alone GPM: valid health status classification system with age-appropriate level descriptions for preference modeling (Table 3), valid level scoring coefficients measured with precision, and commensurate scoring equation for calculating valid HRQL utility scores on the interval scale in which being dead = 0.00 and perfect health = 1.00 (Table 4). HuPS has construct validity of health status attributes and content validity for level descriptions. Level coefficients have

Table 3. HuPS health status classification system.

Attribute	Measurement concepts (bolded) and level descriptions (numbered)
Vision	**Ability to see. Seeing "close to oneself" is at arm's length; seeing "at a distance" means across the street. "Small objects" means as small as a thumb nail.**
	1. Sees normally without glasses (eg, able to see well enough to recognize small objects close to oneself and familiar people across the street)
	2. Sees normally with glasses (eg, able to see well enough to recognize small objects close to oneself and familiar people across the street)
	3. Sees well enough to recognize small objects close to oneself but unable to recognize familiar people across the street even with glasses
	4. Sees well enough to recognize familiar people across the street but unable to recognize small objects close to oneself even with glasses
	5. Unable to see well enough to recognize small objects close to oneself or familiar people across the street, even with glasses
	6. Unable to see at all
Hearing	**Ability to hear**
	1. Hears what a person is saying in a usual environment with background noise and distractions, without a hearing aid
	2. Unable to hear what a person is saying in a usual environment with background noise and distractions, but able to hear what a person is saying in a quiet environment when there are no competing distractions without a hearing aid
	3. Hears what a person is saying in a quiet environment when there are no competing distractions, but requires a hearing aid to do so
	4. Unable to hear what a person is saying, even in a quiet environment and with a hearing aid
	5. Unable to hear at all
Speech	**Ability to articulate clearly (not a measure of ability to use language)**
	1. Speaks clearly and is understood by everyone
	2. Speech understood by parents, but only partially by others
	3. Speech understood only partially by both parents and others
	4. Does not speak or makes only monosyllabic/unintelligible sounds (includes sounds understood by caregivers)
Ambulation	**Ability to get around. Mechanical equipment: canes, crutches, braces, wheelchair/stroller, push toy/walker, includes holding onto wall/furniture for support—does not include ankle-foot orthotics**
	1. Walks, bends, lifts, jumps, and runs as well as others the same age
	2. Walks, bends, lifts, jumps, or runs with some limitations but does not require mechanical equipment or the help of another person to get around independently (eg, a clumsy but independent walker)
	3. Walks or gets around without any help from another person, but requires mechanical equipment (this level includes independent crawlers)
	4. Unable to walk or get around without the help of another person and may also require mechanical equipment (eg, stroller)
Dexterity	**Ability to perform fine motor tasks. Ability to use both hands to play, feed self, and assist with dressing and undressing. "Special tools" include special plate with a rim; adapted spoon, fork, or cup; rings in zippers; and foam grips on markers/paintbrushes**
	1. Full use of both hands and 10 fingers
	2. Limitations in the use of hands or fingers, but does not require special tools or help of another person (includes slow, awkward but independent)
	3. Limitations in the use of hands or fingers, requires the help of another person for some tasks (not independent even with use of special tools)
	4. Limitations in the use of hands or fingers, requires the help of another person for all tasks (not independent even with use of special tools)
Emotion	**Frequency of emotional distress**
	1. Usually cheerful and interested in everyday activities
	2. Occasionally irritable, fretful, unhappy, or uninterested in everyday activities
	3. Often irritable, fretful, unhappy, or uninterested in everyday activities
	4. Always or almost always irritable, fretful, or unhappy and usually uninterested in everyday activities
Cognition	**Ability to think and understand how to solve everyday problems. Everyday problems, eg, getting a cookie, spilled drink—needs to recognize a problem and generate possible solutions; follows a 2-step instruction**

continued on next page

Table 3. Continued

Attribute	Measurement concepts (bolded) and level descriptions (numbered)
	1. Able to think and understand how to solve everyday problems as well as others the same age
	2. Has a little more difficulty than others the same age when trying to think and understand how to solve everyday problems
	3. Has a great deal more difficulty than others the same age when trying to think and understand how to solve everyday problems
	4. Unable to think and understand how to solve everyday problems
Pain and discomfort	**Frequency of pain and disruption of activities**
	1. Free of pain and discomfort (eg, usually does not have pain from earaches, constipation, toothache)
	2. Occasional pain, with no interference of normal activities
	3. Frequent pain which interferes with most normal activities
	4. Constant pain which interferes with all normal activities

Note. Notes about use of HuPS classification system in quantificational and evaluational HRQL surveys. (1) The wording of level descriptions here is exactly as presented to respondents of the Quantificational Survey. (2) The evaluational HRQL survey did not include measurement concept descriptions, and 5 vision level descriptions were modified to include important conceptual details. For example, the vision level 1 description was modified to read "Sees normally without glasses (able to see well enough to recognize small objects such as a thumb nail within arm's length and recognize familiar people across the street)." Very minor modifications, to improve readability, were made to descriptions of vision levels 1-5, hearing level 4, speech level 4, ambulation levels 1-4, dexterity levels 3-4, and pain and discomfort level 2. HRQL indicates health-related quality of life; HuPS, Health Utilities Preschool.

rank, ordinal, and anchor validity. HRQL scores had inter-rater reliability, continuity reliability, and inter-subject validity with HUI3 scores of HRQL. Large intrasubject differences between HuPS and HUI3 scores indicate that continuity may be poor for HRQL measurements of some individuals collected at < 5 and $5+$ years of age and should be interpreted with caution.

HuPS has a well-documented theoretical and empirical pedigree providing a strong foundational structure of important attributes, level severity, and HRQL scoring. Much evidence of methodological rigor and measurement properties for the major foundational components of HuPS (eg, from HUI3 and HSCS-PS) has been published and is directly applicable.[4,20,21,34,35] Two major factors facilitated the success of HuPS: applicability of HUI3 attributes to preschool children and ability to integrate utility coefficients for age-appropriate attribute levels with the HUI3 scoring equation. Applicability of the attributes was due in part to pediatric interests being central throughout the development of HUI3.[34] Applicability was also enhanced by attributes defined "within the skin" and levels defined by functional capacity.[35]

Continuity (ie, predictive) validity is the most important measurement property of models for multiple-use service after publication.[30] The HuPS classification system defines 122 880 unique comprehensive health states (n = 6 vision levels × 5 hearing levels × 4 speech levels × 4 ambulation levels × 4 dexterity levels × 4 emotion levels × 4 cognition levels × 4 pain and discomfort levels). The number of unique health states defined a large 8-attribute space for modeling HRQL scores, representing a challenge for achieving acceptable levels of HuPS to HUI3 continuity. Calibration of HuPS level scoring coefficients, 1 at a time, for use within the HUI3 scoring equation offered a practical and efficient modeling solution. The HuPS scoring function is a MAUT (not regression) model fully specified by 10 parameters: 2 multiattribute constants and 8 sets of level coefficient variables. The number and constructs of HuPS attributes are in common with HUI3, so the multiattribute utility model structure and constants are defined by the HUI3 scoring equation. In other words, health status attribute effects are fully controlled between the HuPS and HUI3 multiattribute utility functions. Health status level effects are controlled between the HuPS and HUI3 multiattribute utility functions by calibration of HuPS scoring coefficients on single-attribute utility scales with HUI3 anchoring. The number of HuPS level coefficients requiring measurement was practical (n = 23 level descriptions). Control of both the attribute and level effects facilitates reliable and valid continuity of HRQL scoring between HuPS and HUI3. A model fitted to predict HuPS scores from HUI3 scores would lack these control mechanisms, leading to less reliable and valid continuity scoring. Continuity validity of HuPS and HUI3 scores of HRQL across subjects is excellent for analyses of groups of subjects to estimate effectiveness of treatments in clinical trials and economic evaluations and calculate higher-order health measures (eg, quality-adjusted life-years and health-adjusted life expectancy).

The outcome of this instrument development provides opportunities for hypothesis generation. Accordingly, a small sample (n = 5) of clinician coauthors and other colleagues were asked to rank the attributes of HuPS most affected, on the basis of a combination of prevalence and severity, in populations of preschool children in the care of whom they have particular expertise. This exercise yielded the following data, to inform testable hypotheses. To refine eventual targets, the top 3 attributes were captured in each instance: cancer (pain and discomfort, emotion, and dexterity), short gut syndrome (emotion, ambulation, pain and discomfort), cerebral palsy (ambulation, dexterity, and cognition), autism spectrum disorder (speech, emotion, and cognition), and children in the general population (pain and discomfort, emotion, and cognition). Examples of the usefulness of the supplemental survey results include studies of the developmental trajectory of preschool aged children with autism and the impact of therapeutic interventions in this age group, as in the treatment of cancer.

It is also important that evidence on the responsiveness and longitudinal construct validity of HuPS be accumulated. Important health policy and management decisions would be well informed by HuPS measurements of effectiveness from trials for clinical and public health interventions with cost-utility analyses of efficiency for the interventions. Reliable and valid instrumentation (eg, questionnaires and coding algorithms) should be developed and distributed to support widespread applications of the HuPS measurement system.

Conclusions

HuPS is a high-quality GPM system for preschool children. HuPS development objectives, design criteria, measurements,

Table 4. HuPS multiattribute utility scoring function.

HuPS level	HuPS attribute utility scoring coefficients							
	Vision b_1	Hearing b_2	Speech b_3	Ambulation b_4	Dexterity b_5	Emotion b_6	Cognition b_7	P&D b_8
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.93	0.92	0.95	0.95	0.97	0.95	0.94	0.96
3	0.90	0.87	0.93	0.90	0.84	0.84	0.76	0.76
4	0.84	0.74	0.80	0.76	0.69	0.73	0.60	0.68
5	0.83	0.73	n/a	n/a	n/a	n/a	n/a	n/a
6	0.73	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Note. Summary statistics of value (before conversion to utility) coefficients in Table 2, columns 4-7.
Formula (on 0.00 = dead to 1.00 = perfect health scale)
$u^* = 1.371\ (b_1{*}b_2{*}b_3{*}b_4{*}b_5{*}b_6{*}b_7{*}b_8) - 0.371$.
Such that u* is the utility of a chronic health state on the utility scale where dead has a utility of 0.00 and perfect health has a utility of 1.00.
b_1 indicates coefficient for level of vision; b_2, coefficient for level of hearing; b_3, coefficient for level of speech; b_4, coefficient for level of ambulation; b_5, coefficient for level of dexterity; b_6, coefficient for level of emotion; b_7, coefficient for level of cognition; b_8, coefficient for level of pain and discomfort; HuPS, Health Utilities Preschool; n/a, not applicable; P&D, pain and discomfort.

respondents, subjects, analyses, and evaluation were well specified. The development process was credible and effective in accordance with recommendations of the ISPOR-SMDM Modeling Good Research Practices Task Force–7.[30]

HuPS scores of HRQL are reliable, valid, and interpretable. HuPS provides scores to inform important policy decisions considering the effectiveness and cost-effectiveness of healthcare services and the comprehensive health of populations. HuPS measurements are commensurate and continuous with those of HUI3 for group analyses. HuPS, in combination with HUI3, makes HUI3-based scoring applicable to the greatest range of age applicability among leading GPM systems: ≥ 2 years of age. HUI3 is among the most frequently used GPM systems worldwide. HuPS has the potential to be widely used in dozens of countries and languages as HUI3 does for older subjects.

Article and Author Information

Accepted for Publication: July 14, 2022

Published Online: xxxx

doi: https://doi.org/10.1016/j.jval.2022.07.015

Author Affiliations: Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, Ontario, Canada (Furlong, Feeny); Division of Hematology and Oncology, Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada (Rae, Breakey, Barr); Division of Neonatal-Perinatal Medicine, Department of Pediatrics, Dalhousie University, Halifax, Ontario, Canada (Ghotra); Division of Developmental Pediatrics, Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada (Carter); Division of Gastroenterology, Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada (Pai); Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada (Pullenayegum); Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada (Xie).

Correspondence: William Furlong, MSc, Centre for Health Economics and Policy Analysis, McMaster University, 88 Sydenham St, Dundas, Ontario L9H-2V3, Canada. Email: furlongb@mcmaster.ca

Author Contributions: *Concept and design:* Furlong, Rae, Feeny, Pai, Pullenayegum, Xie, Barr
Acquisition of data: Rae, Ghotra, Carter, Pai
Analysis and interpretation of data: Furlong, Rae, Feeny, Pai, Pullenayegum, Xie
Drafting of the manuscript: Furlong, Rae, Feeny, Carter, Barr
Critical revision of the paper for important intellectual content: Furlong, Rae, Feeny, Breaky, Pai, Pullenayegum, Xie, Barr
Statistical analysis: Furlong, Feeny
Provision of study materials or patients: Ghotra, Breaky, Carter, Pai
Obtaining funding: Rae
Administrative, technical, or logistic support: Furlong, Rae, Ghotra, Breaky
Supervision: Feeny, Barr

Conflict of Interest Disclosures: Mr Furlong and Dr Barr reported receiving grants from the Canadian Institutes of Health Research during the conduct of the study. Mr Furlong reported receiving personal fees from Health Utilities Inc (HUInc), during the conduct of the study, personal fees and other from HUInc, outside the submitted work. Mr Furlong and Dr Feeny reported a proprietary interest in HUInc, Dundas, Ontario, Canada. HUInc distributes copyrighted Health Utilities Index (HUI) materials and provides methodological advice on the use of the HUI Mark 2 and Mark 3. It should also be noted that HUInc received no payments for the use of the HUI Mark 2 and Mark 3 in the study reported here and a patent copyright is planned. No other disclosures were reported.

Funding/Support: The Canadian Institutes of Health Research, Canada funding reference number 120310. October 2011. PI RD Barr. Title - Development of a health-related quality of life (HRQL) preference-based measure for preschoolers. Mapping the Health Status Classification System Preschool (HSCS-PS) to the HUI2/3 scoring algorithm.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgment: The authors thank George W. Torrance and anonymous reviewers who contributed critical appraisals of a previous version of the manuscript.

REFERENCES

1. Social determinants of health. Early childhood development. World Health Organization. https://www.who.int/social_determinants/themes/earlychilddevelopment/en/. Accessed October 10, 2020.
2. von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. 3rd ed. New York, NY: John Wiley; 1953.
3. Torrance GW, Feeny D, Furlong W. Visual analogue scales: do they have a role in the measurement of preferences for health states. *Med Decis Making*. 2001;21(4):329–334.
4. Feeny D, Furlong W, Torrance GW, et al. Multi-attribute and single-attribute utility functions for the Health Utilities Index Mark 3 System. *Med Care*. 2002;40(2):113–128.
5. Sullivan DF. A single index of mortality and morbidity. *HSMHA Health Rep*. 1971;86(4):347–354.
6. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. *Value Health*. 2009;12(suppl 1):S5–S9.
7. Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *Eur J Health Econ*. 2018;19(4):557–570.

8. Kwon J, Kim SW, Ungar WJ, Tsiplova K, Madan J, Petrou S. A systematic review and meta-analysis of childhood health utilities. *Med Decis Making*. 2018;38(3):277–305.
9. Kwon J, Kim SW, Ungar WJ, Tsiplova K, Madan J, Petrou S. Patterns, trends and methodological associations in the measurement and valuation of childhood health utilities. *Qual Life Res*. 2019;28(7):1705–1724.
10. Commensurate. Merriam-Webster dictionary. www.merriam-webster.com/dictionary/commensurate. Accessed November 13, 2021.
11. Continuous. Merriam-Webster dictionary. www.merriam-webster.com/dictionary/continuous. Accessed May 11, 2021.
12. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 2001;33(5):337–343.
13. Richardson JRJ, Mckie JR, Bariola EJ. Multiattribute utility instruments and their use. In: Culyer AJ, ed. *Encyclopedia of Health Economics*. 1st ed. Vol. 2. San Diego, CA: Elsevier; 2014:341–357.
14. Kennedy-Martin M, Slaap B, Herdman M, et al. Which multiattribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *Eur J Health Econ*. 2020;21(8):1245–1257.
15. Guide to the methods of technology appraisal 2013. Process and methods. National Institute for Health and Care Excellence. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781. Accessed April 20, 2020.
16. Methodological guidance: choices in methods for economic evaluation. Haute Autorité de Santé (France). https://www.has-sante.fr/upload/docs/application/pdf/2020-11/methodological_guidance_2020_-choices_in_methods_for_economic_evaluation.pdf. Accessed January 9, 2021.
17. Rowen D, Keetharuth AD, Poku E, Wong R, Pennington B, Wailoo A. A review of the psychometric performance of selected child and adolescent preference-based measures used to produce utilities for child and adolescent health. *Value Health*. 2021;24(3):443–460.
18. Ramos-Goñi JM, Oppe M, Stolk E, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics*. 2020;38(7):653–663.
19. Rowen D, Rivero-Arias O, Devlin N, Ratcliffe J. 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? *Pharmacoeconomics*. 2020;38(4):325–340.
20. Horsman JR, Furlong WJ, Feeny DH, Torrance GW. The Health Utilities Index (HUI®): concepts, measurement properties and applications. *Health Qual Life Outcomes*. 2003;1:54.
21. Saigal S, Rosenbaum P, Stoskopf B, et al. Development, reliability and validity of a new measure of overall health for pre-school children. *Qual Life Res*. 2005;14(1):243–257.
22. Keeney RL, Raiffa H. *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*. 2nd ed. New York, NY: Cambridge University Press; 1993.
23. Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Oper Res*. 1982;30(6):1043–1069.
24. Chan EKH. Standards and guidelines for validation practices: development and evaluation of measurement instruments. In: Zumbo BD, Chan EKH, eds. *Validity and Validation in Social, Behavioural, and Health Sciences*. Cham, Switzerland: Springer; 2014.
25. Mokkink LB, Terwee CB, Gibbons E, et al. Inter-rater agreement and reliability of the COSMIN (Consensus-based Standards for the selection of health status Measurement Instruments) Checklist. *BMC Med Res Methodol*. 2010;10:82.
26. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737–745.
27. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539–549.
28. Scientific Advisory Committee of the Medical Outcomes Trust, J Alonso, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002;11(3):193–205.
29. Xie F, Pickard AS, Krabbe PFM, et al. A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). *Pharmacoeconomics*. 2015;33(8):867–877.
30. Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value Health*. 2012;15(6):843–850.
31. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*. 2009;11(2):215–225.
32. Nunnally JC. *Psychometric Theory*. 2nd ed. New York, NY: McGraw-Hill; 1978.
33. Readable.com. https://readable.com/blog/the-flesch-reading-ease-and-flesch-kincaid-grade-level/. Accessed April 20, 2021.
34. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med*. 2001;33(5):375–384.
35. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems: health Utilities Index. *Pharmacoeconomics*. 1995;7(6):490–502.