

RESEARCH

Open Access



Development of a multiplicative, multi-attribute utility function and eight single-attribute utility functions for the Health Utilities Index Mark 3 in Japan

Shinichi Noto^{1*}, Takeru Shirowa², Makoto Kobayashi³, Tatsunori Murata³, Shunya Ikeda⁴ and Takashi Fukuda²

Abstract

Background: The Health Utilities Index Mark 3 (HUI3) is a generic multi-attribute, preference-based system for assessing health-related quality of life. It is widely used overseas as an outcome measure and for estimating quality-adjusted life years. We aimed to estimate a multi-attribute and eight single-attribute utility functions for the HUI3 system based on community preferences in Japan. We conducted two preference surveys in this study. The first survey was designed to estimate a model of utility function and collect preference scores, and the second survey was designed to evaluate predictive validity of the utility function and provide independent scores. Values obtained from the feeling thermometer and standard gamble scores obtained from using a chance board were included in the preference scale. We recruited 1043 respondents (age: 20–79 years) from five cities in Japan through the general population classified by sex and age groups. Respondents were further randomly divided into a modeling group ($n = 774$) and a direct group ($n = 263$).

Results: We acquired the estimation for eight single-attribute and a global multi-attribute utility function. The minimum expected multi-attribute utility score was -0.002 . The intraclass correlation coefficient between the directly measured utility score and the score generated by multi-attribute function in terms of 53 health conditions was 0.742.

Conclusions: The HUI3 scoring function developed in Japan has a strong theoretical and empirical basis. It will be useful in future to predict the directly measured score of health technology assessments in Japan.

Keywords: Health utilities index, Health utilities index mark 3, Preference-based measure, Standard gamble, Health-related quality of life, Quality-adjusted life year

Background

In Japan, a new decision-making process for determining the prices of medical technology has started from fiscal year 2019 [1]. The first officially approved guideline for the economic evaluation of drugs and medical devices has been developed for the analytical process in Japan [2]. The guideline describes that the preference-based

measure (PBM) with scoring algorithm developed in Japan should be used when assessing a new quality of life for economic evaluation. However, only few PBMs with scoring algorithm have been developed in Japan (e.g., the 3-level version of the EQ-5D and 5-level version of the EQ-5D [EQ-5D-5 L]). According to the guideline in many other countries (the United Kingdom [3], Canada [4], France [5], Netherlands [6], and Australia [7]), the usage of EQ-5D is mainly recommended as a PBM. However, the guideline in some other countries (Canada [4], France [5], and Australia [7])

* Correspondence: noto@nuhw.ac.jp

¹Department of Rehabilitation, Niigata University of Health and Welfare, 1398 Shimami, Kita-ku, Niigata City, Niigata, Japan
Full list of author information is available at the end of the article

recommends the use of the Health Utilities Index Mark 3 (HUI3) for estimating quality-adjusted life years.

The HUI3 is a generic multi-attribute, preference-based system for assessing health-related quality of life (HRQL) [8] and comprises two complementary components [8, 9]: the first is a multi-attribute health-status classification system used to describe health status and the second is a multi-attribute utility function used to value the health status assessed through the multi-attribute health-status classification system of the previous component. The system defines 972,000 unique health statuses, as it focuses on eight attributes (vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain or discomfort), with each stratified into five to six functional levels. A single-attribute scoring function generates scores for each attribute in the range of 0.00 (the most impaired) to 1.00 (no impairment). The original HUI3 multi-attribute function from Canada generates scores in the range of -0.36 (most impaired, the all-worst HUI3 health state) to 0.00 (being dead) or to 1.00 (perfect health) [9]. The HUI3 has been utilized to evaluate health conditions and HRQL of several patient groups with chronic diseases, and it is a reliable and relevant scale [10–15].

The health-status classification system represents an individual's health at a certain point based on eight attributes in health. There are five to six levels per attribute, ranging from a normal disability level to a severe disability level. For instance, there would be various ranges for pain from “free of pain and discomfort” to “severe pain that prevents most activities.” The scoring function based on a directly measured preference score (community preference) obtained from the random samples of respondents in a general population survey provides the utility score for all defined health states by the health-status classification system, from the conventional dead to perfect health scale. The HUI3 scoring function system is based on the multi-attribute utility theory (MAUT) [16, 17]. In the HUI3, the preference score is measured using standard gambling (SG) as the gold standard from MAUT. Precisely, visual analog scale (VAS)-measured values are converted to SG-measured preferences. The HUI3 health-status classification and preference scoring systems have been verified in various ways by researchers worldwide [10–13]. Direct evidence for the international generalizability of the HUI3 utility scoring function has been reported by LeGalès et al. [18], in that the HUI3 multi-attribute function from France is very similar to the original one from Canada.

This study aimed to develop one multi-attribute and eight single-attribute utility functions for the HUI3 system based on community preferences in Japan in order to use the Japanese economic evaluation in accordance with the unique method.

Methods

Study design and data collection

The design of the HUI3 preference measurement study included two complementary surveys: a survey to collect measurements required for fitting HUI3 multi-attribute utility functions—the HUI3 Modeling survey (HUI3-M)—and an associated survey to collect direct utility measures for 53 states, including states prevalent in the general population (HUI3-D). The HUI3-D provides a valuable commensurate data set for assessing inter-survey or external agreement of HUI3 utility scores. These surveys were conducted in a face-to-face interview.

Both surveys were conducted in five cities in Japan (Sapporo, Tokyo, Nagoya, Osaka, and Fukuoka). These cities are representative of various regions in Japan and are geographically dispersed. All respondents from 20 to 79 years of age were recruited based on snowball sampling by a research company (ANTERIO Inc.). The HUI3-M preference survey collected value and utility measurements from 774 respondents. Sets of health states were randomly allocated to respondents according to strata. Health state strata were defined as follows: scale anchor states (pits [V6, H6, S5, A6, D6, E5, C6, and P5], dead, and perfect health), methodological marker states (MA [V2, H1, S1, A1, D1, E1, C1, and P3], MB [V2, H1, S1, A3, D1, E2, C1, and P3], and MC [V2, H1, S1, A1, D1, E2, C3, and P5]), single-attribute states, and block states. The HUI3-D preference survey collected value and utility measurements from 263 respondents. As in the HUI3-M survey, sets of health states were randomly allocated to HUI3-D respondents according to strata. Health state strata for the HUI3-D survey were defined as follows: scale anchor states, methodological marker states, most prevalent states, and less prevalent states. For the HUI3-M survey, the number of respondents providing value and utility measures varied by health state strata; therefore, precision of the mean preference scores varied by strata.

Value scores were measured using the two-sided feeling thermometer developed by Furlong et al. [19], a prop for eliciting preference scores based on the VAS technique. Standard gamble questions were administered using a modified version of the original chance board prop as follows: in the first step, the interviewee was with certainty in the described health state; in the second step, they were in the best possible state with a certain probability or in the state they considered to be the worst possible one with complementary probability. Different probability values were proposed in an iterative manner until the interviewees stated that they felt indifferently toward both propositions. This last set of data enabled us to establish the function for transforming the values into utilities. Interviews were conducted by 50 interviewers whom we trained in the specific

field of preference elicitation in each region. Interviewers used specifications that included both instructions for managing the interviews and those to be read aloud to the interviewees.

Statistical analysis

Direct preference measures, both values and utilities, are summarized using various statistics: the 10% trimmed mean (5% trimmed off each end of the distribution), standard deviation, minimum, and maximum. The trimmed mean was selected, rather than the median or mode, to maintain most statistical properties associated with using mean-type estimates while reducing the effects of outlier scores on the estimates of central tendency for distributions of the health state preference score with skewed distributions. The person-mean score was defined as the trimmed mean for a specific health state.

The underlying theory of the multiplicative, multi-attribute utility function was described previously by Keeny and Raiffa [20]. The general form for an eight-attribute multiplicative function is as follows:

$$1 + c = \prod_{j=1}^8 (1 + c * c_j) \tag{1}$$

The subscripts (j) indicates a sub-group of attributes.

where $\prod_{j=1}^8$ is the product of all $(1 + c * c_j)$ from c_1 to c_8 .

Respondents were classified into two groups according to the state that each respondent selected as the lowest anchor state when using the feeling thermometer (group A respondents reported pits to be equally or less preferable to dead, and group B respondents reported dead less preferable than pits). Person-mean value scores were calculated for groups A (person-mean (A)) and B (person-mean (B)) scores. Overall person-mean disutility scores were used to fit a multi-attribute disutility function (MADUF), with the scale defined such that perfect health = 0.00 and pits = 1.00 (2), and the MADUF was converted into a multi-attribute utility function (MAUF), with the scale defined such that pits = 0.00 and perfect health = 1.00 (3). Then, the MAUF on the pits/PH scale was converted to a MAUF on the conventional dead = 0.00 to PH = 1.00 scale (4).

Formula (pits/PH scale)

MADUF:

$$\bar{u} = [1/c \left[\prod_{j=1}^8 (1 + c * c_j * \bar{u}_j) - 1 \right]] \tag{2}$$

MAUF:

$$u = 1 - \bar{u} \tag{3}$$

Conversion to the dead/PH scale

$$\begin{aligned} \bar{u}^* &= \bar{u} / \bar{u}_{Dead} \\ u^* &= 1 - \bar{u}^* \end{aligned} \tag{4}$$

Each respondent’s value scores for the single-attribute (including corner) states were normalized such that the least desirable health state was assigned a value score of 0.00 and the most desirable health state was assigned a value score of 1.00. Next, respondent preference measures (i.e., value and utility scores) were classified into one of two groups: person-mean (A) or person-mean (B). The person-mean single-attribute disutility scores provide the \bar{u}_j ‘s. The c_j ‘s are the disutility scores for each of the lowest attribute-level states (i.e., the corner states) on the pits/PH scale, and c was calculated by iteratively solving the equation.

Finally, external agreement (i.e., the extent to which each model can predict utility scores for a group respondents other than the group whose preference scores were used to develop the model) was assessed by comparing the utility scores calculated using the MAUF for each of the 53 health states (marker and 50 other states) to the mean of directly measured utility scores for these states, as reported by respondents in the HUI-D preference survey. Agreement between utility scores by SG and scores was assessed using a two-way mixed model intraclass correlation (ICC) in which the SG and HUI3 MAUF scores were treated as fixed effects and interactions between the participant and instrument were treated as random effects [21]. The ICC estimates the proportion of between-subject variation in relation to total variation, where 1 represents perfect agreement and 0 indicates no agreement at all. A coefficient < 0.40 was considered as poor agreement [22]. Statistical analyses were performed using SAS 9.4.

Results

Flowchart of the survey is shown in Fig. 1. From February to March 2019, 1043 respondents were interviewed by 50 interviewers. Of these, we could not obtain scores because of missing answers on VAS ($n = 6$). As a result, the analysis set included data from 1037 respondents: 774 in the HUI-M and 263 in the HUI-D. Table 1 presents basic characteristics of the respondents (sex, age, location, type of employment, marital status, educational background, and annual household income). We sampled the same number of respondents from each age category; therefore, the groups had almost the same sex ratio. Married and unmarried people accounted for 65.2% and 24.7% of the population in the HUI-M and 65.0% and 24.0% of the population in the

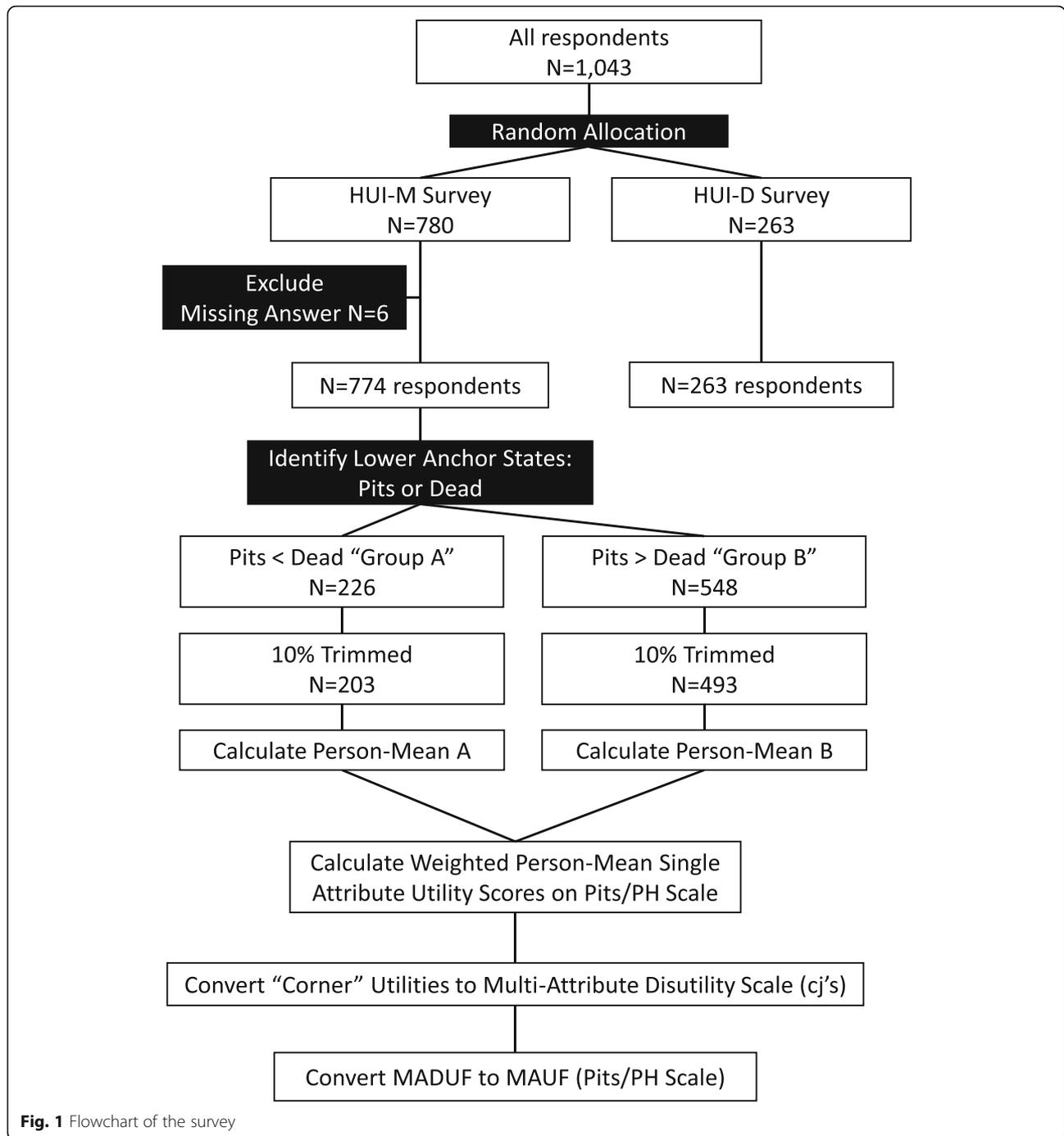


Fig. 1 Flowchart of the survey

HUI-D, respectively. The background of the respondents was comparable with that of the general population.

The person-mean (A) value/utility model was fitted using the person-mean (A) value and person-mean (A) utility scores for states MA, MB, and MC. Each person-mean score is reported in Table 2. The person-mean (A) value and utility scores for these states were each based on 226 observations. Scores for the MA, MB, and MC states were 0.725, 0.659, and 0.487, respectively. The

person-mean (A) fitted value/utility relationship was as follows:

$$u = v^{0.7467}$$

The fitting process used straight-line regression through the origin on the natural log transformations of person-mean (A) value and utility scores.

Table 1 Basic socio-demographic characteristics of the respondents

	HUI-M Population (N = 774)		HUI-D Population (N = 263)	
Gender				
Male	390	50.4%	129	49.0%
Female	384	49.6%	134	51.0%
Age				
20 <= <=29	127	16.4%	44	16.7%
30 <= <=39	128	16.5%	46	17.5%
40 <= <=49	127	16.4%	43	16.3%
50 <= <=59	130	16.8%	45	17.1%
60 <= <=69	130	16.8%	42	16.0%
70 <= <=79	132	17.1%	43	16.3%
Location				
Sapporo	156	20.2%	51	19.4%
Tokyo	154	19.9%	54	20.5%
Nagoya	153	19.8%	54	20.5%
Osaka	155	20.0%	53	20.2%
Fukuoka	156	20.2%	51	19.4%
Employment				
Employed or self-employed	468	60.5%	179	68.1%
Retired	50	6.5%	18	6.8%
Student	41	5.3%	6	2.3%
Homemaker	155	20.3%	41	15.6%
Leave	13	1.7%	4	1.5%
Others	47	6.1%	15	5.7%
Marital status				
Married	505	65.2%	163	62.0%
Unmarried	191	24.7%	63	24.0%
Divorced	47	6.1%	24	9.1%
Bereaved	29	3.7%	13	4.9%
Others	2	0.3%	0	0.0%
Education				
Junior Highschool	19	2.5%	10	3.8%
Highschool	292	37.7%	88	33.5%
College etc.	205	26.5%	74	28.1%
University	248	32.0%	88	33.5%
Graduate school	10	1.3%	3	1.1%
Household income				
<JPY 2mil	59	7.6%	27	10.3%
JPY 2 mil<= < 4 mil	170	22.0%	60	22.8%
JPY 4 mil<= < 6 mil	216	27.9%	61	23.2%
JPY 6 mil<= < 10 mil	194	25.1%	63	24.0%
JPY 10 mil<= < 15 mil	62	8.0%	26	9.9%
JPY 15 mil <=	15	1.9%	5	1.9%
Refused, Unknown	58	7.5%	21	8.0%

Table 2 Scores for three marker states in two groups

	10% trim mean (VAS)		10% trim mean (SG)	
	Person-mean (A) (N = 226)	Person-mean (B) (N = 548)	Person-mean (A) (N = 208)	Person-mean (B) (N = 489)
MA ^a	83.8	85.3	0.725	0.688
MB	62.4	64.1	0.659	0.610
MC	35.6	35.9	0.487	0.455
Dead	16.2	–	0.315	–
Pits	–	4.3	–	0.113

^aMA [V2, H1, S1, A1, D1, E1, C1, and P3], MB [V2, H1, S1, A3, D1, E2, C1, and P3], MC [V2, H1, S1, A1, D1, E2, C3, and P5] and Pits [V6, H6, S5, A6, D6, E5, C6, and P5]

The fit yielded an R² of 0.8687, not corrected for the mean.

The person-mean (B) value/utility model was also fitted using person-mean (B) value and person-mean (B) utility scores for states MA, MB, and MC. Person-mean (B) value and utility scores for these states were each based on 548 observations. Scores for the MA, MB, and MC states were 0.688, 0.610, and 0.455, respectively. The person-mean (B) fitted value/utility relationship was as follows:

$$u = v^{0.8437}$$

The R² was 0.8437, not corrected for the mean.

Table 3 presents the single-attribute utility functions on the pits = 0.00/PH = 1.00 scale and single-attribute utility functions on the lowest level = 0.00/highest level = 1.00 scale.

For measurement and computational convenience, the multiplicative, multi-attribute utility function (Eq. 1) was fitted using disutility scores [19] (disutility = 1 – utility). The disutility function was fitted on the perfect health = 0.00 to most disabled = 1.00 scale. The estimates on this scale were as follows: c1 = 0.58, c2 = 0.46, c3 = 0.54, c4 = 0.58, c5 = 0.57, c6 = 0.61, c7 = 0.63, c8 = 0.53, c = – 0.998, and Σci = 4.50 (the disutility parameters, ci, correspond to the utility parameters, kj, in Eq. (1); c corresponds to k). It is noteworthy that the sum of the ci was 4.50, and the value for c was – 0.998. A simplified format of the scoring function, converted from disutility to utility and transformed using the dead/perfect health scale, appears in Table 4. MAUF was as follows:

$$u = 1.003 \times (b1 \times b2 \times b3 \times b4 \times b5 \times b6 \times b7 \times b8) - 0.003.$$

Scores estimated using the multi-attribute utility function were defined on a scale such that the minimum score was – 0.002 (most disabled), dead was 0.00, and the maximum (perfect health) score was 1.00. This

Table 3 Single-attribute utility function for Japanese HUI3

Level	Vision	Hearing	Speech	Ambulation	Dexterity	Emotion	Cognition	Pain
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.96	0.89	0.87	0.89	0.90	0.91	0.89	0.92
3	0.77	0.75	0.56	0.73	0.71	0.63	0.83	0.77
4	0.58	0.53	0.33	0.44	0.42	0.28	0.58	0.35
5	0.31	0.35	0.00	0.24	0.17	0.00	0.29	0.00
6	0.00	0.00	–	0.00	0.00	–	0.00	–

approach to model fitting did not involve analysis of variance, and this distinction was an important design factor in selecting the appropriate methods for assessing the performance of the MAUF.

A high level of intra-survey (HUI-M) agreement was observed between the person-mean scores and scores used in the MAUF model (ICC = 0.756). The inter-survey agreement statics are presented in Fig. 2 for 53 health states reported to be prevalent in the general population. The ICC point between directly measured utility scores and scores generated by the multi-attribute function estimate was 0.742 (95%CI = 0.592–0.843).

Discussion

In this study, the preference measurement survey included two widely used preference elicitation instruments: a VAS known as the feeling thermometer and the standard gamble implemented using the chance board. Feeny et al. [23] reported that the use of props and face-to-face interviews by well-trained interviewers have important effects on eliciting the preference score. The use of these well-tested props also had a mutually beneficial effect on the face-to-face interview survey in our study.

The multiplicative, multi-attribute HUI3 scoring function is based on a well-established theory and on the gold standard method for measuring preferences—the standard gamble. The multiplicative functional form accommodates important interactions in the preferences among attributes. Our result was

almost consistent with that obtained in Canada for the HUI3 [9]. The similarities of our protocol with those chosen to construct the HUI3 multi-attribute utility function in Canada enabled us to make an international comparison of our results.

Scores estimated using the multi-attribute utility function are defined on a scale such that the minimum score is – 0.002, similar to that of the EQ-5D-5 L pits state [55555], which was – 0.035 for Japan [24]. Despite the existence of these similarities between the pits score by the Japanese HUI3 and that by the Japanese EQ-5D-5 L, the difference between the pits score by the Japanese HUI3 (minimum score = – 0.002) and one by the original Canadian function (minimum score = – 0.36) is very large probably due to the preference for worse than dead. However, at this time, it is difficult to make a definitive conclusion regarding this issue, so further research is needed.

The HUI3 has more attributes than the EQ-5D-5 L, which has five attributes (domains) [25]. The weakness of the EQ-5D-5 L, as reported by Brazier et al. [26], is a lack of domains for vision and hearing and cognition and dementia. The HUI3 has the potential to compensate for the weakness of the EQ-5D-5 L, and the development of the function in Japan using our work is sensible.

A high level of agreement was observed between scores for 53 health states generated by the function and directly measured scores for the same states from a direct survey. The multi-attribute function, which was developed in Japan, has strong theoretical and empirical

Table 4 Multi-attribute utility function for Japanese HUI3

Level	Vision	Hearing	Speech	Ambulation	Dexterity	Emotion	Cognition	Pain
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.97	0.95	0.93	0.94	0.94	0.94	0.93	0.96
3	0.87	0.88	0.76	0.84	0.84	0.77	0.90	0.88
4	0.76	0.78	0.64	0.67	0.67	0.56	0.74	0.66
5	0.60	0.70	0.46	0.55	0.53	0.39	0.55	0.47
6	0.42	0.54	–	0.42	0.43	–	0.37	–

$u = 1.003 \cdot (b1 \cdot b2 \cdot b3 \cdot b4 \cdot b5 \cdot b6 \cdot b7 \cdot b8) - 0.003$
 Pits score = – 0.002

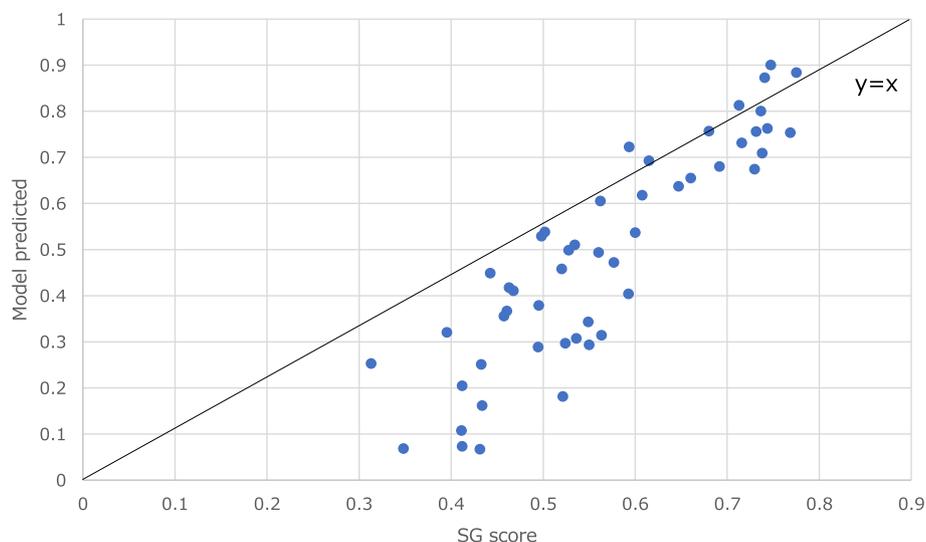


Fig. 2 External agreement (SG scores vs scores using MAUF)

foundations and performs well in predicting directly measured scores for health technology assessment in Japan.

To the best of our knowledge, the work presented herein is the first attempt at revealing individual preferences for health states using VAS and SG methods performed in a general Japanese population.

Conclusions

We developed the first multiplicative, multi-attribute utility function-based standard gamble in Japan. The Japanese HUI3 MAUF seems to perform very well. The level of intra- and inter-survey agreement scores might be interpreted as evidence of both validity and reliability. Our work will be useful in establishing cost-effectiveness analyses in Japan where new decision-making processes of the pricing of health technologies has started.

Abbreviations

HUI3: Health Utilities Index Mark 3; PBM: Preference-based measure; EQ-5D-5 L: 5-level version of the EQ-5D; HRQL: Health-related quality of life; MAUT: Multi-attribute utility theory; SG: Standard gambling; VAS: Visual analog scale; HUI3-M: HUI3 modeling survey; HUI3-D: HUI3 direct survey; MADUF: Multi-attribute disutility function; MAUF: Multi-attribute utility function; ICC: Intraclass correlation

Acknowledgements

We gratefully acknowledge the help of David Feeny and Bill Furlong in reviewing earlier drafts of this manuscript and providing their invaluable suggestions for improving it.

Authors' contributions

SN and TS were primarily responsible for the study conception and design including informed consent, negotiation with representatives in the community, acquisition of data, analysis and interpretation of data, and drafting and revising the manuscript. TS, MK, and TM were responsible for the analysis of data. SN was responsible for the study design, analysis and interpretation of data, and revising the manuscript. SI and TF were responsible for reviewing the manuscript. All authors have approved the

final version of this manuscript for publishing and agree to be held accountable for all aspects of the work.

Funding

This study was partly funded by Health and Labour Science Research Grants, Ministry of Health, Labour, and Welfare (H30-seisaku-shitei-003). In addition, this study was partly supported by the JSPS KAKENHI Grant of Ministry of Education, Culture, Sports, Science, and Technology, Research on the JSPS KAKENHI grant number JP 18H03031.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the Ethical Guidelines for Clinical Research of the Japanese Ministry of Health, Labour and Welfare, and with the 1964 Helsinki declaration and its later amendments. Written informed consent for responding to the surveys was obtained from all respondents after the aim of this study was explained to them. Ethics committee approval for this study was received from Niigata University of Health and Welfare (approval number: 17560–150205).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Rehabilitation, Niigata University of Health and Welfare, 1398 Shimami, Kita-ku, Niigata City, Niigata, Japan. ²Center for Outcomes Research and Economic Evaluation for Health (C2H), National Institute of Public Health, Wako, Saitama, Japan. ³CRECON Medical Assessment Inc., The Pharmaceutical Society of Japan, Nagai Memorial, 12-15 Shibuya 2-chome, Shibuya-ku, Tokyo 150-0002, Japan. ⁴Department of Public Health, School of Medicine, International University of Health and Welfare, 4-3 Kozunomori, Narita, Chiba 286-8686, Japan.

Received: 24 October 2019 Accepted: 12 March 2020

Published online: 03 April 2020

References

- Shiroiwa, T., Fukuda, T., Ikeda, S., & Takura, T. (2017). New decision-making processes for the pricing of health technologies in Japan: The FY 2016/2017 pilot phase for the introduction of economic evaluations. *Health Policy*, *121*, 836–841. <https://doi.org/10.1016/j.healthpol.2017.06.001>.
- Shiroiwa, T., Fukuda, T., Ikeda, S., Takura, T., & Moriwaki, K. (2017). Development of an official guideline for the economic evaluation of drugs/medical devices in Japan. *Value in Health*, *20*, 372–378. <https://doi.org/10.1016/j.jval.2016.08.726>.
- National Institute for Health and Care Excellence. (2013). *Guide to the methods of technology appraisal*. London: National Institute for Health and Care Excellence.
- Canadian Agency for Drugs and Technologies in Health. (2006). *Guidelines for the economic evaluation of health technologies* (3rd ed.). Ottawa: Canadian Agency for Drugs and Technologies in Health.
- Haute Autorité de Santé. (2002). *Choices in methods for economic evaluation*. Paris: Haute Autorité de Santé.
- National Health Care Institute. (2016). *Guideline for economic evaluations in healthcare*. Diemen: National Health Care Institute.
- Pharmaceutical Benefits Advisory Committee. (2008). *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (version 4.3)*. Canberra: Pharmaceutical Benefits Advisory Committee.
- Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The Health Utilities Index (HUI): Concepts, measurement properties and applications. *Health and Quality of Life Outcomes*, *1*, 54. <https://doi.org/10.1186/1477-7525-1-54>.
- Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., et al. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care*, *40*, 113–128.
- Szecket, N., Medin, G., Furlong, W. J., Feeny, D. H., Barr, R. D., & Depauw, S. (1999). Preliminary translation and cultural adaptation of Health Utilities Index questionnaires for application in Argentina. *International Journal of Cancer*, *83*(S12), 119–124.
- Felder-Puig, R., Frey, E., Sonleithner, G., Feeny, D. H., Gadner, H., Barr, R. D., et al. (2000). German cross-cultural adaptation of the Health Utilities Index and its application to a sample of childhood cancer survivors. *European Journal of Pediatrics*, *159*, 283–288.
- Wang, Q., & Chen, G. (1999). The health status of the Singaporean population as measured by a multi-attribute health status system. *Singapore Medical Journal*, *40*, 389–396.
- Glaser, A. W., Furlong, W. J., Walker, D. A., Fielding, K., Davies, K., Feeny, D. H., et al. (1999). Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumors in the U.K. *European Journal of Cancer*, *35*, 256–261. [https://doi.org/10.1016/S0959-8049\(98\)00367-0](https://doi.org/10.1016/S0959-8049(98)00367-0).
- Costet, N., Le Galès, C., Buron, C., Kinkor, F., Mesbah, M., Chwalow, J., et al. (1998). French cross-cultural adaptation of the Health Utilities Index Mark 2 (HUI2) and 3 (HUI3) classification systems. *Quality of Life Research*, *7*, 245–256 Clinical and Economic Working Groups.
- Feeny, D., Kaplan, M. S., Hugué, N., & McFarland, B. (2010). Comparing population health in the United States and Canada. *Population Health Metrics*, *8*, 8. <https://doi.org/10.1186/1478-7954-8-8>.
- Furlong, W. J., Feeny, D. H., Torrance, G. W., & Barr, R. D. (2001). The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Annals of Medicine*, *33*, 375–384.
- Torrance, G. W., Furlong, W., & Feeny, D. (2002). Health utility estimation. *Expert Review of Pharmacoeconomics & Outcomes Research*, *2*, 99–108. <https://doi.org/10.1586/14737167.2.2.99>.
- Le Galès, C., Buron, C., Costet, N., Rosman, S., & Slama, G. (2002). Development of a preference-weighted health status classification system in France: The Health Utilities Index 3. *Health Care Management Science*, *5*, 41–51.
- Furlong, W., Feeny, D., Torrance, G. W., Goldsmith, C. H., DePauw, S., Zhu, Z., et al. (1998). *Multiplicative multi-attribute utility function for the Health Utilities Index Mark 3 (HUI3) system: A technical report, working paper series no.98-11*. Hamilton: CHEPA, McMaster University.
- Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value tradeoffs* (2nd ed.). New York: Cambridge University Press.
- Shrout, E., & Fleiss, L. (1979). Inter class correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.
- Bland, J. M., & Altman, D. G. (1990). A note on the use of the intra class correlation coefficient in the evaluation of agreement between two methods of measurement. *Computers in Biology and Medicine*, *20*, 337–340.
- Feeny, D., Furlong, W., & Torrance, G. W. (2018). What were they thinking when providing preference measurements for generic health states? The evidence for HUI3. *Health and Quality of Life Outcomes*, *16*, 166. <https://doi.org/10.1186/s12955-018-0993-9>.
- Ikeda, S., Shiroiwa, T., Igarashi, I., Noto, S., Fukuda, T., Saito, S., et al. (2015). Developing a Japanese version of the EQ-5D-5L value set. *Journal of the National Institute of Public Health*, *64*, 47–55 (in Japanese).
- Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*, 1727–1736. <https://doi.org/10.1007/s11136-011-9903-x>.
- Brazier, J., Ara, R., Rowen, D., & Chevrou-Severac, H. (2017). A review of generic preference-based measures for use in cost-effectiveness models. *Pharmacoeconomics*, *35*(Suppl 1), 21–31. <https://doi.org/10.1007/s40273-017-0545-x>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)