Introduction

The assessment of HRQOL is an essential element of health care evaluation [1]. Generic HRQOL instruments are designed to be applicable across a wide range of populations and interventions. HUI3 is a generic approach to the measurement of health status and assessment of HRQOL. The HUI provides a comprehensive framework within which to measure health status and calculate HRQOL scores. HUI is comprised of two complementary components. The first component is a multiattribute health status classification system that is used to describe health status. The second component is a multiattribute utility function that is used to value health status as assessed within the corresponding multiattribute health classification system. The system defines 972000 unique health using five or six functional levels for each of the eight attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain or discomfort. HUI3 provides a global utility score and eight single attribute scores. A multiattribute preference function has been developed in Canada. Details are presented in [2].

The HUI3 system has been implemented in 4 large scale Canadian populations health survey: the 1990 Ontario Health Survey [3,4], the 1991 General Social Survey [5], the ongoing National Population Health Surveys [6,7], and the ongoing National Longitudinal Surveys of Children and Youth [8]. HUI measures have been shown to be reliable [9,10] and to capture the pertinent attributes of health status in the general population [11,12].

Generic instruments have been translated for use in general population samples in Japan. Fukuhara et al. reported the results of translation, adaptation and validation study of SF-36 Health Survey in Japan [13]. The Japanese EuroQol development committee reported on the official Japanese version of EuroQol [14] and Ikeda et al. reported on the health status in Japanese populations using it [15]. Tazaki et al reported the results of a qualitative and field study using WHOQOL instrument for cancer patients [16]. Tazaki et al translated the original English version of the question naive into Japanese according to the strict protocol required by the original developers. In order to perform international comparison of health status using HRQOL measurement instruments, linguistic differences in the questionnaire between two languages have to be minimized. There are three major steps. First is forward translation, second is backward translation and review by original developers, third is testing in focus groups. Because Japanese is not based on Indo-European languages and Western culture is not dominant in Japan, conceptual difficulties in translation for some words are not uncommon. The goal is to produce a conceptual rather than literal translation that is compatible with original meaning of the questionnaires.

HUI dose not provides a single official guideline for translation, so we followed several published protocols and worked closely with the developers of HUI. This article describes the process of translation, cultural adaptation and testing in focus group chronologically. Next the results of a community survey to provide evidence of validity of the translated and adapted HUI3-based instrument, test-retest reliability, conceptual validity, interpretability, and construct validity are described.

Material and methods

Translation and cultural adaptation

Forward translation and reconciliation version

An original English language questionnaire which includes HUI2 and HUI3 was obtained from the HUI developers. The questionnaire provides sufficient information to classify subjects in both the HUI2 and HUI3 systems. The health status classification system for HUI3 is summarized in Table1. (Results from applying HUI2 will be reported elsewhere.) Translators 1 and 2 who were Japanese and English bilingual (native speakers of Japanese) translated instruction items and questionnaire independently to produce two initial Japanese versions. Both translators and the Japanese HUI development team discussed the translation and conceptual problems of the two versions, to produce a single reconciled version.

Back-translation and revision by developers

The reconciled version was translated by two professional translator (translator 3 and 4) who are bilingual and bicultural native speakers of English, to produce two back translation versions. The Japanese HUI development team compared the back translation versions with the original English questionnaire. Critical review by the HUI developers was provided. Through discussions between Japanese HUI development team and the HUI developers, several linguistic, cultural and conceptual problems were found. These problems are described briefly.

General Directions

The original English version asks about "ability" in the general directions and for each attributes except emotion. Among several options, we chose literal translation "NOURYOKU" for Japanese to avoid misunderstanding "physical ability" for "usual condition". The focus of the questions is on what one's health status permits one to do or inhibits one from doing, not what you choose to do. HUI23SU15Q includes several similar questions, so in order to encourage respondents to think carefully about their response to each question, in the instruction we asked respondents to "please excuse the apparent overlap and answer each question independently."

Vision

The translated question naive emphasized that the questions were about the ability to see and, not about how well you read. The question is about vision, not literacy. This concept was reflected in both the question and the response options. The Japanese questionnaire about vision uses the concept of "being able to see or see to distinguish."

Hearing

Similar to concept of vision, the focus is on hearing, not comprehension. The Japanese questionnaire about hearing reflects this concept.

Speech

The concept and best translation of "when speaking with people who know you well" were discussed. "People who know you well" should convey that people who are very familiar with the respondent's speech. "Your own language" was deleted from original questionnaire because almost all respondents in Japan are native speakers of Japanese.

Ambulation

The cultural concept of "neighborhood" was difficult to translate. The wording of Japanese version were arranged to convey the respondents the ability to walk several hundred meters outdoors but in a non-challenging environment. Further, the intent of the question is not just walking but physically moving about. The Japanese questionnaire about ambulation conveys this concept.

Dexterity

Concept of "special tools" was discussed. The tools described in original questionnaire are not common in Japanese daily life and culture. "Limitation" were translated as "Not free to do" to convey the concept.

Emotion

As is often the case, the translation and cultural adaptation of emotional concepts were difficult. The concept of "somewhat" was discussed. Translation was designed to convey the degree of burden from Level 1 through Level 5. Magnitude estimation was employed among translator and the development team using visual analogue scale, to assist in selecting the best Japanese words.

Cognition

The translation focuses on the severity of the cognitive problems, remembering or thinking rather than frequency of such problems.

Pain

The original concept of the frequency and degree of severity of pain were discussed. Cultural differences for how to relieve the pain were taken into account.

Revision and Second back translation.

Close attention was paid to problems identified by the back translation and the reconciled version was corrected. Several lay panel session (different groups) provided evidence on how to convey original concept while being natural and acceptable in Japanese. Testing in small focus group was done with 158 community respondents. They were asked about difficulties in concept of questionnaire and response options. Very few problems were identified. The questionnaire was considered to be suitable for Japanese use. All procedure were reported to the HUI developers together with second back translation. The HUI developers and Japanese HUI development team were satisfied with the results. The Japanese HUI23SU15Q was deemed ready for use in Japan.

Test-retest validity

A community sample (n=112) was used with re-sampling of three weeks was used to estimate Cronbach's alpha (internal consistency) and the test-retest reliability. Multiattribute global utility scores, single-attribute utility scores and VAS (described below) scores were compared between the two surveys. Also, Cronbach's alpha and correlation coefficient test-retest reliability were calculated by the age-group and personal characteristics as follows: with or without chronic disease, excellent to very bad for interhuman-relationship in family and work site.

Japanese community survey

A large community survey was undertaken in fall of 1999. 3860 people who are employees of two large corporations and members of their families and nearby residents were surveyed. 200 local residents in Shizuoka (200km west of Tokyo) district were also included as respondents. HUI questionnaire were distributed to each respondents individually via local branches of two corporations or at the time of their visit to the company clinic for a routine health check up. Questionnaires were returned by mail. For respondents living outside of Tokyo, completed question naives were initially collected in local branches of the corporation and were mailed to author's office.

Respondents were asked about personal characteristics along with the HUI questionnaire. VAS (visual analogue scale) estimation was done after the HUI questionnaire. The VAS tasks involved the use of a visual aid consisting of a vertical thermometer shaped scale, 9cm long. The top was labeled "1.0" perfect health" and the bottom was labeled "0" dead". Respondents were asked to mark a point on this scale to represent their usual health status. VAS estimation is not part of the original HUI questionnaire. Respondents were asked about 20 personal characteristics potentially related to HRQOL, the same variables as surveyed in Ontario Health Survey [3,4]. The items included 1 name, 2 sex, 3 age, 4 BMI (height and weight), 5 survey date, 6 occupation, 7 educational attainment, 8 resident area, 9 family size, 10 marital status, 11 types of residence, 12 annual family income, 13 with or without debt load, 14 work schedule, 15 employment conditions, 16 job stability, 17 commuting time, 18 interpersonal relationship at work over the past three months, 19 interpersonal relationship at home over the past three months, and 20 number and type of chronic disease.

In order to calculate multi-attribute global utility scores and single-attribute utility scores, the HUI3 scoring function by Furlong et al., (1998) was adopted [17]. The single-attribute and global scoring functions for HUI3 are based on data collected in a preference survey of a random sample of the general population in Hamilton, Ontario, Canada. The HUI3 global utility scores are defined on the interval –0.36 to 1.00 (perfect health). The negative lower bound reflects the fact that respondents judged the health state with the lowest level of capacity in each of the eight attributes to be worse than dead [17].

Statistical analysis

This study was designed to validate the Japanese version HUI3 for use in Japan, so we mainly focused on describing the distribution of attribute levels and mean utility scores. We also examined relationships between personal characteristics and HUI3 and VAS scores. Further we examined the ability of personal characteristics to predict HUI3 scores in this community sample.

In order to assess the construct validity, the relationships between personal characteristics and HUI3 were examined. First, we compared the mean global utility and single-attribute utility scores among the groups by personal characteristic variables. Categories of variables are as follows; younger than 20, 20 to 29, 30 to 39, 40 to 49, 50 to 59, 60 to 69, 70 and older for age group, student, low, high for educational attainments, married, divorced, widowed, single for marital status, male, female for gender, 0-10000USD, 10000-50000USD, more than 50000USD for annual family income, seeking work or part time, student, housewife, others for employment, excellent, good, fair, bad, very bad for inter-human relationship in work site and in family, 0, 1, 2, 3 and over for number of chronic diseases.

The defamation to high education category included those with at least a college education, low education category includes elementary school, junior high school, high school, vocational school and others. Family income is defined as annual household income of respondents. If the respondents are students, income of their part time work and income from parents were included. If the respondents are housewife, income means their husband's income. For employment the categories were employed seeking work, part time job retired and no job. For inter-human relationship, we offered 5 options in responses: excellent, good, fair, bad, very bad, and respondents selected one of these. For the definition of number of chronic disease, it means the number of chronic diseases which respondents checked among the response options as follows; 1 allergies, 2 asthma, 3 arthritis or rheumatism, 4 back pain and back problem, 5 high blood pressure, 6 migraine headaches, 7 chronic bronchitis or emphysema, 8 sinusitis, 9 diabetes, 10 epilepsy, 11 heart disease, 12 cancer, 13 stomach or intestinal ulcers, 14 effects of stroke, 15 urinary incontinence, 16 liver dysfunction, 17 dermatitis requiring medication, 18 dementia, 19 cataract, 20 other chronic condition. The list of chronic disease was taken from Ontario Health Survey [3,4]. If our translation was successful and Japanese HUI3 system is valid, then respondents in the groups known to have, on average, lower HRQOL should have lower mean global utility score and single-attribute utility scores.

The 20 types of chronic disease were classified into 10 categories of chronic condition as follows; 1 allergy, 2 cardiopulmonary disease, 3 musculo-skeletal disorder, 4 hypertension, 5 hyper lipidemia, metabolic disease, 6 visual and hearing disorder, 7 central nervous disorder, 8 malignant tumor, 9 gastro-intestinal disorder, 10 no chronic disease, to clarify the relationship between HUI attributes and each disease or condition specific problem. Mean single-attribute scores, global utility scores, and VAS score were assessed for each of these 10 categories of chronic conditions. If the Japanese HUI3 system is valid, single utility scores should predict some disease specific problems, for example, the mean cognition utility score for central nervous disorder respondents should be lower etc.

To clarify the relationships between personal characteristic variables and HUI3 scores, linear regression models were also used to compare utility scores between groups while controlling for the effects of potentially confounding variables. Failure to control for confounding effects could result in biased results. For example, mean age was related to the number of chronic diseases, widowed status, and lower education. Thus, differences in health status between these groups was likely due to the effects both of age and personal characteristics variables rather than the effects of each personal characteristics alone, as would be implied by uncontrolled comparisons of utility scores. Categorical variables were captured as dummy variables for the multiple regression analysis. To accommodate the complex survey design, the statistical software package SPSS 10.0J was used.

The set of control variables included respondents' educational attainments, marital status, gender, annual family income, employment, inter-human relationship in work site, inter-human relationship in family, age, and number of chronic diseases. Groups and categories are same as described above in the assessment of mean utility scores, except age was included as continuous variables. After excluding respondents with non-response to the necessary questions, 2,960 subjects remained with which to estimate the models. Linear regression estimate for model of multiattribute global utility score, single-attribute utility scores and VAS scores as function of age and 10 categories of chronic condition were also done. The baseline category represents the respondents who did not have any a type of chronic disease (no chronic condition). Data used to estimate the regression models were 3,762 for single-attribute and global utility score, 3,576 for VAS score. If Japanese HUI3 has reasonable construct validity, respondents in the groups of possibly lower HRQOL related personal characteristics should have negative correlation coefficient between global utility score and variable categories. As well, in the regression model of the relationship between chronic condition and utility scores, respondents who have specific chronic condition should have negative correlation coefficient between single-attribute utility scores and a category usually associated with the disease specific problems, such as central nervous disorder, musclo-skeletal disorder and single-attribute utility scores for cognition and pain, respectively.

Kendall correlation between HUI3 single-attribute score were calculated to estimate independence of each of the eight attributes of Japanese HUI3. If independence of each attributes of Japanese HUI3 would be suggested, no substantial linear correlation will be found among 28 possible cross comparisons. Relationship between multiattribute global HUI3 utility score and self-rated health was estimated. We used the response to a self-rated health question: Overall how would you rate your usual health? The responses were excellent, very good, good, fair, and poor. We also estimated the distribution (percentage) of three categories of self-rated health (1 excellent and very good, 2 good, 3 fair and poor) among the 10 groups of respondents whose global utility score were as follows; less than 0.2, 0.2 to less than 0.3, 0.3 to less than 0.4, 0.4 to less than 0.5, 0.5 to less than 0.6, 0.6 to less than0.7, 0.7 to less than 0.8, 0.8 to less than 0.9, 0.9 to less than 1.0, 1.0. The relationship between multiattribute global utility scores and VAS scores was examined. These two approaches examine if the Japanese HUI3 material correlates with subjective (self- rated) health status.

Result

In pilot study with a small community sample, test-retest with a three week interval, reliability of Japanese HUI3 as assessed with Cronbach's alpha was 0.84 for global utility score and 0.79 for VAS score. With respect to single-attribute utility scores, Cronbach's alpha were 0.78, 0.93, 0.73, 0.96, 0.80, 0.44, 0.62 and 0.73 for vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain, respectively. With respect to personal characteristics, older age-group and respondent with 1 and more chronic conditions showed higher Cronbach's alpha.

A larger community survey was done using Japanese HUI3 questionnaire and VAS with 3,860 subjects. The mean age was 41±14.3 years old and ranged from 14-90 years, with a median of 39 and mode of 37. Male to female ratio was 2651:1209. Male gender were more than two times as frequent as female gender. Age distribution by 10-year age group (from 10's to 70 and over) was as follows:3.6, 17.6, 29.7, 23.2, 15.5, 5.7, 4.7%, respectively. Sixty percent of those surveyed lived in the greater Tokyo metropolitan area, while the remaining 40% were distributed throughout the nation. Distribution of single-attribute levels by respondents is shown in Table 2, no level 6 cases were found in, hearing, speech, dexterity, emotion, and pain. Approximately 100% belonged to level 1 with respect to hearing, speech, ambulation, and dexterity. Distribution of respondents in level 1 and level 2 were approximately same in vision, emotion, and pain. Table 3 gives the mean and standard deviation of utility score (single attribute, multiattribute global) and VAS score by 10-year age groups. There were no age-related decline in all attribute in age group younger than 70 years old. ANOVA revealed significant differences in mean utility score (single-attribute and multiattribute global) and VAS score, showing that the age-related decline starts from older than 70 years old. Significant lower utility scores were seen in younger groups in emotion. Global utility score was substantially lower in age group 40's and 50's. VAS score declined as respondents age-group become older.

Table 4 provides the mean and standard deviation of utility score and VAS score by the group of personal characteristics. For single-attribute utility score, several substantially lower score was seen. In hearing, ambulation, cognition and pain, widowed and seeking work or part time for employment showed lower single-attribute utility scores. As well, in cognition, lower levels of educational attainment showed lower scores. As for inter-human relationship in work site and family, and number of chronic conditions, single-attribute and global utility scores showed rank correlation between each other, such as worse categories of inter-human relationship associated with lower score in both single-attribute and global utility scores. With respect to global utility scores, lower scores were seen for low education, widowed, male gender, higher annual family income, seeking work or working part time.

Table 5 provides the results of linear regression estimates for model of multiattribute global utility score as function of personal characteristic variables to control the effect of confounding.

Baseline variables (omitted category) are noted for each variables in Table 5. Significant negative correlation coefficient were seen in lower educational attainment, male gender and higher number of chronic diseases. Significant positive correlations were seen with fair, good, excellent (with omitted category of very bad), in inter-human relationship in family and work site. The intercept is 0.67. The coefficient of determination is 0.19.

We also examined the mean utility and VAS scores by other personal characteristic variables, such as BMI (height and weight), occupation, residential area, family size, types of residence, debt load, work schedule, job stability, and commuting time. We did not reveal systematic associations. Reasons may include misunderstanding of some questions such as types of residence. Respondents answered ownership status but we were asking about space and comfort. In addition, the sample size was small.

Table 6 provides the mean single-attribute and global utility scores and VAS scores by type of chronic disease. Comparatively lower single-attribute utility scores in all attributes was seen in respondents with any type of chronic disease. Hyper lipidemia showed lower single-attribute utility score in vision, malignant tumor did in hearing and allergy and metabolic disease did in speech. As well, visual and hearing disorder showed lower single-attribute utility score in ambulation, cardiopulmonary disease, musculo-skeletal disorder and central nerve disorder did in emotion, central nervous disorder did in cognition and musculo-skeletal disorder did in pain. With respect to global utility and VAS score, the highest mean score were seen in groups with no chronic disease. Table 7 demonstrates the results of linear regression model of global, single-attribute utility scores and VAS scores as a function of age and type of chronic disease. The baseline category represents respondents with no chronic disease. With respect to single-attribute utility scores, allergy had significantly negative correlation in speech, emotion, cognition, and pain, cardiopulmonary disease did in hearing, speech, emotion and pain, musculo-skeletal disorder did in dexterity and pain, hyper lipidemia did only in emotion, metabolic disease did only in speech, visual and hearing disorder did in ambulation and pain, central nervous disorder did in vision, hearing, ambulation, dexterity, cognition and pain, malignant tumor did in vision, hearing, ambulation, dexterity and pain, and gastrointestinal disorder did only in pain. Age was used in regression estimates as continuous variable and showed significant negative correlation in global utility score and VAS score and all single-attribute utility scores except speech. With respect to global utility score, significant negative correlation was seen in allergy, cardiopulmonary disease, musculo-skeletal disorder, metabolic disease, visual and hearing disorder, central nerve disorder and gastrointestinal disorder. Number of subjects used in this regression estimates was 3,762 for single-attribute and global utility scores and 3,576 for VAS score.

Table 8 provides the Kendall correlation among HUI3 single-attribute utility scores. A substantial correlation (r>0.25) were seen in only 3 combinations (ambulation and dexterity, speech and cognition, cognition and pain) among the 28 possible comparisons.

Figure 1 shows the relationship between multiattribute global utility scores and self-rated health. Black area of bar-graph represents response frequency of fair and poor in self-rated health. Black space decreases gradually as multiattribute global utility scores become higher (with the exception of only one group global utility score 0.3 to less than 0.4). As well, white area of bar-graph which represents response frequency of excellent and very good shows a gradual increase as global utility scores became higher. With respect to gray area which represents good as middle level of self rated-health, was about same among groups of 0.3 to 0.8 of global utility scores and decreased at both the higher and lower ends of global utility scores.

Figure 2 shows the relationship between global utility score and VAS score in all respondents. The correlation coefficient was 0.44, which suggest moderate positive correlation between the two scores. We also calculated relationship of HUI3 and VAS score by 10 year age-groups. Correlation coefficients were 0.53, 0.52, 0.35, 0.47, 0.29, 0.28 and 0.49, respectively.

Discussion

Respondent and administrative burden was acceptable. Mean respondent time was 3.5 minutes and incomplete response rate was less than 1.0%.

For test-retest reliability, our results suggested that the reliability of Japanese HUI3 is approximately the same as in Canada. Boyle MH et al [9] reported that for the eight attributes, kappa estimates varied from 0.137 to 0.728. The interclass correlation for global utility score was 0.767. The translation of the HUI3 questionnaire into Japanese seems to have been successful.

In interpreting the result of community survey, we should note that the age and gender distribution of the sample was not fully represent Japanese general population. The community sample was relatively healthy, especially in older age groups.

With respect to personal characteristic variables, the mean global and single-attribute utility scores demonstrated a discriminative ability in agreement with our expectation, but with several exceptions. We emphasize that, especially in inter-human relationship in work site and in family, and number of chronic disease, the mean utility scores were lower. As well, lower educational attainment, male gender, widowed for marital status and seeking work or part time job for employment had lower utility score which was consistent with previous HRQOL investigations using generic instruments [19].

Single-attribute utility scores for vision, hearing, speech, cognition and pain were lower in older age groups. Chronic conditions demonstrated specific deficits in health status consistent with expectations. These results provide initial confirmation of the construct validity of the Japanese HUI3.

This conclusion is corroborated by the results of linear regression estimates of factors associated with multiattribute global utility scores. Controlling for potential of confounders, results reveal strong relationship between variables and utility scores, as expected. Almost all negative coefficients were significant between utility score and variables which were hypothesized to reduce HQQOL.

Lower single utility scores were associated with chronic conditions. Impairments of specific attributes were associated, as expected, with particular conditions. The results are similar to those of Grootendorst et al., who reported evidence of construct validity of HUI3 for stroke and arthritis in a n=77,663 population health survey in Canada [20]. Grootendorst et al. reported lower global utility scores in respondent with stroke, arthritis and both. The difference in mean utility score was –0.297, -0.084 and –0.712, respectively. Respondents with stroke were reported to have lower single-attribute utility scores for speech, ambulation, dexterity, emotion, and cognition. Results from this Japanese study include lower single-attribute utility scores for pain and cognition for musculo-skeletal disorder as well, emotion, cognition and pain for central nervous disorder.

In the regression results, 7 of 10 chronic disease categories showed significant negative coefficients for global utility score and significant age-related deterioration for attributes. With respect to particular chronic conditions, the expected relationships were, in general, observed. For instance, significant negative coefficients for emotion and pain were observed for cardiopulmonary disease. Indeed patients with ischemic heart disease often complain of chest pain and anxiety (risk of sudden death). Negative coefficients for dexterity and pain were observed for musclo-skeletal disorder. For central nervous disorder and malignant tumor, there were negative and significant coefficients for 5 or 6 of the 8 attributes, respectively. Negative coefficients for pain were observed for gastrointestinal disorder (perhaps due to stomachache or other abdominal pain). The use of generic instruments, in particular the HUI, has revealed underrecognized burdens caused by pain in a number of published clinical studies, including pediatric neuro-oncology [21,22], adult neuro-oncology [23], and survivors of extremely low birth weight [24-26]. Our results provide preliminary evidence that Japanese HUI3 system has cross-cultural validity, linguistic validity, and construct validity.

With respect to the Kendall correlations among HUI3 single-attribute attribute scores, only 3 of 28 possible cross comparisons showed substantial correlation (r>0.25). These results suggest acceptable independence among the attributes. This result is also compatible with the recent report from Houle et al. [27], in which only 2 of 28 comparison demonstrated substantial correlation.

There was a wide range of global utility scores among respondents reporting "good" for their health status. 30-40% of respondents reporting "good" had global utility scores lower than 0.4. On the other hand, 60-80% of respondents who reported "excellent" or "very good" had scores higher than 0.8. Similar results were reported by Gold et al., from the result of a survey of 14,407 US adult respondents [19]. Results in a Canadian survey are also similar [27]. VAS scores and global utility scores were positively correlated. The Japanese HUI3 appears to have discriminative validity and interpretability.

Although reliability, face validity, construct validity, discriminate validity were reported from several studies using original questionnaire and scoring function, the question remains as to whether the HUI3 scoring function developed in Canada can be adopted for Japanese use. To determine the international generalizability of the HUI3 scoring function, preference surveys of representative and appropriate sized samples of general populations in Japan should be performed. Results should be compared to results representing the ethnically heterogeneous Canadian population used in fitting the HUI3 scoring function.

In general, there is substantial heterogeneity of preferences for health states among individuals Furlong 1996 [28], Kaplan 1994 [29], Torrance et al 1996 [30]. Further, there is growing evidence that quantitative preferences (values and utilities) are robust when measured with the same procedures, regardless of the population or even the country where the measurement took place. For example, the scoring methods for the original Quality of Well-Being scale were undertaken on a general population sample in San Diego in early 1970s [31]. When this work was replicated on an arthritic population in the Northeastern United States in the early 1980s, similar results were found [32]. Yet many would argue that the population of the U.S. Northeast is different culturally relative to that of Southern California.

Direct support for the international robustness of quantitative preferences measured with the same procedures came with the early work of the EuroQol group. They found that EuroQol VAS scores were similar across three European countries [33]. More recently, LeGales and colleagues from INSERM have replicated in France the Canadian scoring procedures for the HUI3 and have found quite similar results [3,4]. All of this is consistent with the growing realization that demographic, societal or cultural characteristics of subjects are not consistent predictors of utilities. As we often say in lay terms, "poor quality health status is universally recognized and deemed undesirable; this is a constant of being a human being". In addition, Bosch et al assessed health status in patients with peripheral arterial disease using HUI2 and the EuroQol-5D. They concluded that results are highly similar even though HUI2 was developed in North America and the EQ-5D in Europe.

On the other hand, it has long been known that different measurement procedures lead to consistently different results i.e.; standard gamble, time trade-off, visual analog scale. Similarly, different multi-attribute systems also lead to different results one compared to the other [27,35].

The inescapable conclusion of the appropriateness for adaptation of HUI3 scoring function is that the procedure or instruments matters but the reference population that provided the scoring does not. This is reflected in the recent Australian report [36] that recommends that the Australian government standardize on one, or at most a small number, of instruments like the Health Utilities Index. The Australian report did not recommend that any of these instruments needed to be re-scored for use in Australia. The implications for Japan are that the existing scoring formula for the Health Utilities Index is suitable and appropriate for use in Japan.

Our results indicate that the translation and cultural adaptation of HUI into Japanese was successful. We have provided evidence of construct validity and discriminative validity. But some limitations of this study should be kept in mind.

First, our sample size is not large enough to cover the full range of health status. The sample was generally healthy. The age gender distributions were not representative of general population.

Second, personal characteristic variables are based solely on self-reports. The results depends on the extent to which self-report is accurate.

Third, although the HUI3 scoring function may be quite generalzible in the Western context, The scoring function may not generalize to Japan where Western culture is not dominant and the religious traditions differ.

Future studies are needed. Large scale of population health surveys should be conducted to cover a wide range of health status and include a variety of documented clinical conditions. Further survey is necessary to examine the accuracy of Canadian scoring function for use in Japan.

In spite of the study limitations, the Japanese HUI3 appears to be a useful measure of HRQOL for use in Japan.

Acknowledgement

The authors are thank Professor Dr. Chifumi Satoh, Professor Dr. Kazuyuki Omae for their scientific advice. Authors also very thank Dr. Matsuoka, Dr. Noriyoshi Uemura and Miss Chizuru Yamashita for their cooperation in field survey. This study was supported in part, by Health Sciences Research Grants of Ministry of Health and Welfare, Research on Health Services.

References

1. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993; 118: 622-629.

2. Feeny DH, Torrance GW, Furlong W. Health Utilities Index. In: Spilker B, ed. Quality of life and pharmacoeconomics, 2nd ed. Philadelphia: Lippincott-Raven Press, 1996:85-95.

3. Ontario Ministry of Health. Ontario Health Survey, 1990, User's Guide. Vol.1: Documentation. Toronto: Ontario Ministry of Health, 1992

4. Ontario Ministry of Health. Ontario Health Survey, 1990, User's Guide. Vol.2: Documentation. Toronto: Ontario Ministry of Health, 1992

5. Statistic Canada. The 1991 General Social Survey-Cycle6: Health-Public use microdata file documentation and user's guide. Ottawa: Statistic Canada, 1992

6. Statistic Canada. National Population Health Survey, 1994-1995, Public Use Microdata Files (82-F0001XCB). Ottawa: Statistic Canada, 1998.

7. Statistic Canada. National Population Health Survey, 1994-1995, Public Use Microdata Files (82-M0009XCB). Ottawa: Statistic Canada, 1998.

8. Statistic Canada and Human Resources Development Canada. 1994-95 National Longitudinal Survey of Children and Youth-User's handbook and microdata guide (89M0015GPE, 89M0015XDB). Ottawa: Statistic Canada, 1995.

9. Boyle MH, Furlong W, Feeny DH, Torrance GW, Hatcher J. Reliability of Health Utilities Index Mark Ⅲ used in the 1991 cycle 6 Canadian General Social Survey Health Questionnaire. Qual Life Res 1995;4:249-257.

10. Gemke RJ, Bonsel GJ. Reliability and validity of comprehensive health status measure in a heterogeneous population of children admitted to intensive care. J Clin Epidemiol 1996;49:327-333.

11. Furlong W, Torrance GW, Feeny DH. Properties of Health Utilities Index: Preliminary evidence. Quality of life Newsletter 1995:13-14.

12. Torrance GW, Furlong W, Feeny DH, Boyle M. Multi-attribute preference functions: Health Utilities Index. PharmacoEconomics 1995; 7: 503-520.

13. Fukuhara S, Bito S, Green J, Hsiao A, Kurokawa K. Translation, Adaptation, and Validation of the SF-36 Health Survey for Use in Japan. J Clin Epidemiol 1998; 51: 1037-1044.

14. Japanese EuroQol Translation Team. The development of the Japanese EuroQol Instrument. J Health Care and Society 1998; 8: 109-123.

15. Ikeda S, Ikegami N, Japanese EuroQol Tariff Project Team. Health Status in Japanese Population: Results from Japanese EuroQol Study. J Health Care and Society 1999; 9: 83-92.

16. Tazaki M, Nakane Y, Endo T, Kakikawa F, Kano K, Kawano H, Kuriyama K, Kuroko K, Miyaoka E, Ohta H, Okamoto N, Shiratori S, Takamiya S, Tanemura K, Tsuchiya R. Results of Qualitative and Field Study Using the WHOQOL Instrument for Cancer Patients. Jpn J Clin Oncol 1998; 28: 134-141.

17. Furlong W, Feeny DH, Torrance GW, Goldsmith C, DePauw S, Boyle m, Denton M, Zhu Z, "Multiplicative Multi-Attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report, McMaster University, Centre for Health Economics and Policy Analysis Working Paper. 1998; No. 98-11

18. Gabriel SE, Kneeland TS, Melton J III, Moncur MM, Bruce Ettinger B, Tosteson ANA, Health-Related Quality of Life in Economic Evaluations for Osteoporosis: Whose Values Should We Use? Medical Decision Making, 1999; 19: 141-148.

19. Gold M, Franks P, Erickson P. Assessing the Health of the Nation: The Predictive Validity of a Preference-Based Measure and Self-Rated Health. Medical Care. 1996; 34(2): 163-177.

20. Grootendorst P, Feeny D, Furlong W. Health Utilities Index Mark 3: Evidence of Construct Validity for Stroke and Arthritis in a Population Health Survey. Medical Care. 2000; 38(3): 290-299.

21. Whitton C, Rhydderch H, Furlong W, Feeny D, Barr D. Self-Reported Comprehensive Health Status of Adult Brain Tumor Patients Using the Health Utilities Index. Cancer. 1997; 80(2): 258-265.

22. Barr D, Simpson T, Whitton A, Rush B, Furlong W, Feeny D. Health-Related Quality of Life in Survivors of Tumours of the Central Nervous System in Childhood - A Preference-Based Approach to Measurement in a Cross-Sectional Study. European Journal of Cancer. 1999; 35(2): 248-255.

23. Whitton C, Rhydderch H, Furlong W, Feeny D, Barr D. Self-Reported Comprehensive Health Status of Adult Brain Tumor Patients Using the Health Utilities Index. Cancer. 1997; 80(2): 258-265.

24. Saroj S, Rosenbaum P, Stoskopf B, Hoult L, Furlong W, Feeny D, Burrows E, Torrance G. Comprehensive Assessment of the Health Status of Extremely Low Birthweight Children at Eight Years of Age: Comparison with a Reference Group. Journal of Pediatrics. 1994; 125(3): 411-417.

25. Saroj S, Feeny D, Furlong W, Rosenbaum P, Burrows E, Torrance G. Comparison of the Health-Related Quality of Life of Extremely Low Birthweight Children and a Reference Group of Children at Age Eight Years. Journal of Pediatrics. 1994; 125(3): 418-425.

26. Saroj S, Feeny D, Rosenbaum P, Furlong W, Burrows E, Stoskopf B, Hoult L. Self-Perceived Health Status and Health-Related Quality of Life of Extremely Low Birthweight Infants at Adolescence. Journal of the American Medical Association. 1996; 276(6): 453-459.

27. Houle C, Berthelot JM. A Head-to-Head Comparison of the Health Utilities Mark 3 and the EQ-5D for the Population Living in Private Households in Canada. Quality of Life Newsletter. 2000; 24: 5-6.

28. Furlong W. Variability of utility scores for health states among general population groups. MSc thesis, McMaster University. 1996.

29. Kaplan M, Value judgment in the Oregon medical experiment. Medical Care. 1994; 32(10): 975-988.

30. Torrance G, Feeny D, Furlong W, Barr D, Zhang Y, Wang Q. Multi-Attribute Preference Functions for A Comprehensive Health Status Classification System: Health Utilities Index Mark 2. Medical Care. 1996; 34(7): 702-722.

31. Patrick D, Bush J, Chen M. Methods for measuring levels of well-being for a health status index. Health Services Research. 1973; 8(3): 228-245.

32. Balaban D, Sagi P, Goldfarb N, Nettler S. Weights for scoring the quality of well-being instrument among rheumatoid arthritics: A comparison to general population weights. Medical Care. 1986; 24(11): 973-980.

33. EuroQoL Group. EuroQoL: A new facility for measurement of health-related quality of life. Health Policy. 1990; 16: 199-208.

34. LeGales C, Buron C, Costet N, Rosman S, Slama G. The French Health Index Mark 3. Value in Health. 2000; 3: 103-103.

35. Bosch L, van Wijck E, Baum L, Donaldson C, van den Dungen M, Hunink M. The McMaster Health Utility Index (II) and the EuroQol 5D Assessed in Patients with Peripheral Arterial Disease in the United States and the Netherlands. Medical Decision Making. 1996; 16(4): 450.

36. Richardson J, Olsen A, Hawthorne G, Mortimer D, Smith R. The measurement and validation of utility based quality of life: Report to the Pharmaceutical Evaluation Section, Pharmaceutical Benefit Branch, Commonwealth Department of Health and Family Services. 1999; Tender No: 22/98.