Evaluating the discriminatory power of EQ-5D, HUI2 and HUI3 in a US general population survey using Shannon's indices

Mathieu F. (Bas) Janssen · Erwin Birnie · Gouke J. Bonsel

Abstract

Objectives To compare quantitatively the discriminatory power of the EQ-5D, HUI2 and HUI3 in terms of absolute and relative informativity, using Shannon's indices.

Methods EQ-5D and HUI2/3 data completed by a sample of the general adult US population ($N = 3,691$) were used. Five dimensions allowed head-to-head comparison of informativity: Mobility/Ambulation; Anxiety/Depression/Emotion; Pain/Discomfort (EQ-5D; HUI2; HUI3); Self-Care (EQ-5D; HUI2); and Cognition (HUI2; HUI3). Shannon's index and Shannon's Evenness index were used to assess absolute and relative informativity, both by dimension and by instrument as a whole.

Results Absolute informativity was highest for HUI3, with the largest differences in Pain/Discomfort and Cognition. Relative informativity was highest for EQ-5D, with the largest differences in Mobility/Ambulation and Anxiety/Depression/Emotion. Absolute informativity by instrument was consistently highest for HUI3 and lowest for EQ-5D, and relative informativity was highest for EQ-5D and lowest for HUI3.

Discussion Performance in terms of absolute and relative informativity of the common dimensions of the three instruments varies over dimensions. Several dimensions are suboptimal: Pain/Discomfort (EQ-5D) seems too crude with only 3 levels, and the level descriptions of Ambulation (HUI3) and Self-Care (HUI2) could be improved. In absence of a formal measure, Shannon's indices provide useful measures for assessing discriminatory power of utility instruments.

Keywords Health status · Methodology · Psychometrics · Population health · Health-related quality-of-life

Introduction

The need for assessing health-related quality of life (HRQL) has brought forth hundreds of HRQL instruments, both generic and disease-specific [1, 2]. Generic instruments fall into two main categories: (1) preference-based health classification systems, and (2) non-preference based measures, sometimes referred to as health profile or psychometric measures [1, 3, 4]. Preference-based classification systems, also referred to as multi-attribute utility instruments (MAUIs) are standardized health state classifications that can be used to obtain a single summary index (utility score) or so-called preference weight for different health states. At the core of any MAUI is a classification system consisting of multiple attributes (dimensions) with ordered levels for each dimension. Most MAUIs are generic and aim to cover the full spectrum of disease and disability. MAUIs are widely used as measures of health outcome and are applied in clinical and economic evaluation (to calculate QALYs) and in population health surveys. Three widely used MAUIs are the EQ-5D, the Health Utilities Index Mark 2 (HUI2)

M. F. (Bas) Janssen (✉) · E. Birnie · G. J. Bonsel
Public Health Epidemiology, Department of Social Medicine, Academic Medical Center, 22660, 1100 DD Amsterdam, The Netherlands
e-mail: m.f.janssen@amc.uva.nl

and the Health Utilities Index Mark 3 (HUI3) [5–7]. All three instruments have shown acceptable psychometric properties as established by conventional measures [8, 9].

Feasibility, reliability, validity and responsiveness are important measurement properties in MAUIs, just as they are in non-preference based HRQL and health status measures such as the SF-36. However, these properties may be operationalized differently in MAUIs compared to non-preference based measures [4, 10, 11]. An underlying property to the concepts of reliability, validity and responsiveness is the ability of an instrument to discriminate between ('true') different levels of health. This requires a MAUI to define the full range of potential health states, and to be sensitive over this range. A necessary measurement property for any health status measure (including MAUIs) is the ability to discriminate among people at a single point in time. This property is sometimes referred to as sensitivity or, more accurately: "discriminatory power" [12–14].

Guyatt et al. (1992) proposed a reliability coefficient as a suitable statistic to express discriminatory power [15]. Reliability essentially reflects two different concepts: (1) consistency, e.g. between raters (inter-rater reliability) or over time (test-retest reliability), and (2) discriminatory power: the ability of an instrument to discriminate among people [16]. We propose Shannon's indices of informativity as suitable measures that solely reflect discriminatory power [17].

Discriminatory power of MAUIs is usually investigated in an informal and partial manner by examining the frequency distributions, e.g. for floor or ceiling effects [12, 13, 18–20]. Shannon's indices are suitable to assess discriminatory power in MAUIs for two reasons: first, they are theoretically based and second, they incorporate the frequency distribution across all categories of a MAUI's health status classification system (not just the highest and lowest categories, as is the case with ceiling and floor effects).

Our aim is to investigate the discriminatory power of the EQ-5D, HUI2 and HUI3 in a general population sample, as expressed by Shannon's indices. Informativity was assessed separately by dimension and by MAUI as a whole.

Methods

Data

A publicly available dataset was used (at http://www.ahrq.gov/rice/), resulting from the US EQ-5D valuation study [21, 22]. Collected data consisted of self-completed EQ-5D and HUI2/3 data from a sample of the general adult US population, with an over-sampling of Hispanics and non-Hispanic Blacks. The HUI2/3 data were collected using a standardized 15-item questionnaire, from which HUI2 and HUI3 health profiles were extracted using available recoding algorithms [23]. Only the responses of 3,691 respondents who had no missing data on any of the three instruments were included in this study (91.2% of the total number of respondents).

Instruments

The EQ-5D descriptive system consists of 5 dimensions (items) with 3 levels each, logically defining 243 unique health states (permutations). The HUI2 was originally developed to assess outcomes in survivors of cancer in childhood and contains 6 dimensions (excluding the original HUI2 dimension of fertility) with 4–5 levels per dimension. The HUI3, originally developed for a general population health survey in Canada, has 8 dimensions with 5–6 levels per dimension. The HUI2 and HUI3 descriptive systems define 8,000 and 972,000 unique health states, respectively [6]. Table 1 compares the 5 dimensions common to at least two of the classification systems: Mobility/Ambulation; Anxiety/Depression/Emotion; Pain/Discomfort (EQ-5D; HUI2; HUI3); Self-Care (EQ-5D; HUI2); and Cognition (HUI2; HUI3).

Shannon's indices: background and properties

The Shannon index, named after Claude Shannon who is considered to be the founder of information theory, was initially developed to separate noise from information carrying signals in telecommunication systems [17]. The Shannon index is also known as the Shannon–Weaver index because of Warren Weaver's contribution to Shannon's original paper, and as the Shannon–Wiener index named after Norbert Wiener who independently developed a concept similar to Shannon's [24, 25]. The Shannon index has been applied in a variety of fields, ranging from ecology (as a measure of biodiversity) to psychology, record linkage and molecular biology (genetic diversity) [26–30].

In information theory, the information of a signal is distinguished from the meaning or the semantic content of a signal. Rather, the information is quantified and is identified with uncertainty. Informativity is dependent on the number of classes (e.g. bits or response options) and the distribution of the observations (the 'signal') among classes. For classifications, this implies that if one

Table 1 Level descriptions for common dimensions between EQ-5D, HUI2 and HUI3

EQ-5D	HUI2	HUI3
Mobility	*Mobility*	*Ambulation*
No problems in walking about	Able to walk, bend, lift, jump, and run normally for age	Able to walk around the neighbourhood without difficulty, and without walking equipment
Some problems in walking about	Walks, bends, lifts, jumps, or runs with some limitations but does not require help	Able to walk around the neighbourhood with difficulty; but does not require walking equipment or the help of another person
Confined to bed	Requires mechanical equipment (such as canes, crutches, braces, or wheelchair) to walk or get around independently	Able to walk around the neighbourhood with walking equipment, but without the help of another person
	Requires the help of another person to walk or get around and requires mechanical equipment as well	Able to walk only short distances with walking equipment, and requires a wheelchair to get around the neighbourhood
	Unable to control or use arms and legs	Unable to walk alone, even with walking equipment. Able to walk short distances with the help of another person, and requires a wheelchair to get around the neighbourhood
		Cannot walk at all
Self-care	*Self-care*	
No problems with self-care	Eats, bathes, dresses, and uses the toilet normally for age	
Some problems washing or dressing self	Eats, bathes, dresses, or uses the toilet independently with difficulty	
Unable to wash or dress self	Requires mechanical equipment to eat, bathe, dress, or use the toilet independently	
	Requires the help of another person to eat, bathe, dress, or use the toilet	
Pain/Discomfort	*Pain*	*Pain*
No pain or discomfort	Free of pain and discomfort	Free of pain and discomfort
Moderate pain or discomfort	Occasional pain. Discomfort relieved by non-prescription drugs or self-control activity without disruption of normal activities	Mild to moderate pain that prevents no activities
Extreme pain or discomfort	Frequent pain. Discomfort relieved by oral medicines with occasional disruption of normal activities	Moderate pain that prevents a few activities
	Frequent pain; frequent disruption of normalactivities. Discomfort requires prescription narcotics for relief	Moderate to severe pain that prevents some activities
	Severe pain. Pain not relieved by drugs and constantly disrupts normal activities	Severe pain that prevents most activities
Anxiety/ Depression	*Emotion*	*Emotion*
Not anxious or depressed	Generally happy and free from worry	Happy and interested in life
Moderately anxious or depressed	Occasionally fretful, angry, irritable, anxious, depressed, or suffering "night terrors"	Somewhat happy
Extremely anxious or depressed	Often fretful, angry, irritable, anxious, depressed, or suffering "night terrors"	Somewhat unhappy
	Almost always fretful, angry, irritable, anxious, depressed	Very unhappy
	Extremely fretful, angry, irritable, anxious, or depressed usually requiring hospitalization or psychiatric institutional care	So unhappy that life is not worthwhile
	Cognition	*Cognition*
	Learns and remembers school work normally for age	Able to remember most things, think clearly and solve day to day problems
	Learns and remembers school work more slowly than classmates as judged by parents and/or teachers	Able to remember most things, but have a little difficulty when trying to think and solve day to day problems
	Learns and remembers very slowly and usually requires special educational assistance	Somewhat forgetful, but able to think clearly and solve day to day problems

Table 1 continued

EQ-5D	HUI2	HUI3
	Unable to learn and remember	Somewhat forgetful, and have a little difficulty when trying to think or solve day to day problems
		Very forgetful, and have great difficulty when trying to think or solve day to day problems
		Unable to remember anything at all, and unable to think or solve day to day problems

would want to develop a useful (informative) distinction between, say, European countries, distinguishing between Scandinavian and non-Scandinavian countries would be far less informative than distinguishing between Northern, Western, Eastern and Southern European countries. Note that the latter classification not only contains more categories but the countries are also more evenly distributed among categories.

The Shannon index is defined as:

$$H' = -\sum_{i=1}^{C} p_i \log_2 p_i$$

where H' represents the absolute amount of informativity captured, C is the number of possible categories (levels or permutations in this study), and $p_i = n_i/N$, the proportion of observations in the ith category ($i = 1,...,C$), where n_i is the observed number of scores (responses) in category i and N is the total sample size [17]. Any log base can be used, as long as one is consistent. Using log base 2, as did Shannon, allows the interpretation of the resulting units as bits per individual. The higher the index H' is, the more information is captured by the system. In case of a homogeneous (rectangular) distribution, i.e. ratings are evenly distributed among categories ($p_i = p^*$ for all i), the optimal amount of information is captured and H' has reached its maximum (H'_{max}) which equals $\log_2 C$. If the number of categories (C) is increased, H'_{max} increases accordingly but H' will only increase if the newly added categories are actually used. The variance of the Shannon index is defined as [31]:

$$\text{var } H' = \frac{\sum_{i=1}^{C} p_i (\log_2 p_i)^2 - \left(\sum_{i=1}^{C} p_i \log_2 p_i \right)^2}{N}$$

Accordingly, standard errors and 95% confidence intervals can be calculated.

The Shannon index combines the absolute information content as expressed by the number of categories with the extent to which the information is evenly spread over these categories. Shannon's Evenness index (J') exclusively reflects the latter component, i.e.

the rectangularity of a distribution. This measure was first proposed by Lloyd and Ghelardi [32]; Shannon already referred to it as relative entropy and Pielou termed the concept 'evenness' [17, 33]. Shannon's Evenness index (J') is defined as: $J' = H'/H'_{max}$, which expresses the use of the system (H') given its potential (H'_{max}). Shannon's index H' can be considered as an expression of the absolute informativity of a system whereas Shannon's Evenness index J' expresses the relative informativity of a system or 'evenness' of a distribution, regardless the number of categories.

Two alternative measures of (bio)diversity are the Simpson and the Brillouin index. We used the Shannon index, since the Brillouin index is dependent on sample size and the Simpson index gives very little weight to categories that are rarely occupied [26, 34, 35].

Shannon indices applied to MAUIs

The basic characteristics of Shannon's indices which make them suitable to reflect discriminatory power have been documented and are explained as follows. In an item where a response option has a very high (or low) endorsement, e.g. p is over 0.95 (or under 0.05), one learns very little because one can predict with more than 95% certainty what the answer will be. In other words, there is very little information being transmitted. Conversely, the maximum amount of information (uncertainty) is being transmitted when, in an item with two response options, p is 0.50 for each response option. As described above, this characteristic of an even distribution underlies the Shannon indices. In case of an even distribution, the item (dimension) is being most efficiently used, which means that the discriminant ability of the level descriptors is maximal.

The Shannon indices can be calculated by dimension separately or by MAUI as a whole. To calculate Shannon's indices by dimension, levels are treated as categories, so C represents the number of levels (L), p_i is the proportion of responses of the ith level, and H'_{max} equals $\log_2 L$. Suppose the EQ-5D Mobility dimension is scored by 10 respondents: no problems ($n = 6$), some problems ($n = 3$) and confined to bed ($n = 1$). Shannon's index for Mobility is calculated as
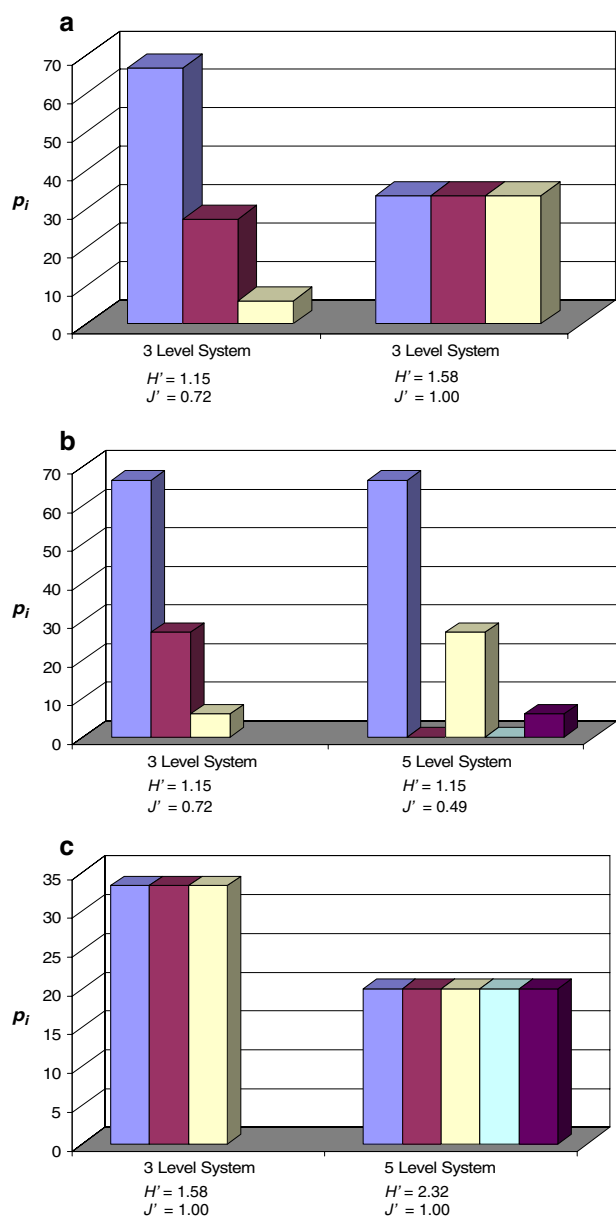
Fig. 1 Examples of Absolute Informativity (H') and Relative Informativity (J') with Skewed and Rectangular Distributions in a 3 Level System and 5 Level System

$H' = -((0.6 \quad \log_2 \quad 0.6) + (0.3 \quad \log_2 \quad 0.3) + (0.1 \quad \log_2 \quad 0.1)) = \mathbf{1.30}$ and $H'_{max} = \log_2 3 = 1.58$, so $\boldsymbol{J'} = 1.30/1.58 = \mathbf{0.82}$.

Figure 1 illustrates the difference between absolute and relative informativity (H', evenness J') relative to the number of levels (L) in a series of hypothetical health classification systems designed to describe the same underlying dimension. For illustrative purposes we consider only one dimension. Figure 1a shows two distributions of responses corresponding to two different classification systems, both of which have 3

levels; one system results in a skewed distribution while the other results in a rectangular distribution. Assuming these responses are obtained within the same population, the system that yields the rectangular distribution is superior in discriminating between patients and the Shannon indices have both reached their maximum values. Figure 1b illustrates the concept of relative informativity. The left panel shows the same skewed distribution as depicted in Figure 1a, the right panel shows the same distribution of responses but now as it results from a 5 level classification system in which levels 2 and 4 are unused. Absolute informativity (Shannon's H') remains unchanged but J' decreases, expressing lower relative informativity. Clearly, adding 2 extra levels that do not represent anyone in the population (no individual shifts from a current level to any of the new levels) does not lead to a gain in absolute informativity (H') while the potential of a 5 level system is underutilized, compared to a 3 level system, which is expressed by a lower J'. So why not use just the Shannon Evenness index? Figure 1c shows the added value of absolute informativity (the H' index). If the 3 and 5 level systems both yield rectangular distributions, evenness J' will be the same but obviously H' increases since the 5 level system is much more refined in discriminating between patients.

To calculate Shannon's indices by instrument as a whole, permutations are treated as unique categories (e.g. 243 categories for EQ-5D), so C is the number of permutations (P_{max}), p_i is the proportion of the ith permutation, and H'_{max} now equals $\log_2 P_{max}$.

Since the number of observations in our study ($N = 3{,}691$) is lower than the number of theoretically possible permutations in HUI2 (8,000) and HUI3 (972,000), maximum informativity (H'_{max}) in HUI2 and HUI3, and consequently maximum relative informativity J' cannot be reached a priori. Therefore, Shannon's indices by MAUI as a whole were calculated using an estimation approach. Assuming that the current sample is representative, subsamples of the original set of observed health states were drawn in order to estimate the number of different health states in hypothetical populations of 1, 10, and 100 million respondents, by means of extrapolation. This procedure was repeated for different proportions of the population in relation to the number of health states (e.g. 11 different EQ-5D health states accounted for a 90% proportion of the respondents), in order to estimate the shape of the frequency distribution in the hypothetical populations of 1, 10, and 100 million respondents. Finally, Shannon's H' and J' could be calculated (details can be obtained from the authors).

Table 2 Frequency distribution (%) of responses to the EQ-5D, HUI2 and HUI3 instruments (N = 3,691)

	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
EQ-5D						
Mobility	82.17	17.53	0.30	–	–	–
Self care	95.58	4.01	0.41	–	–	–
Usual activities	84.88	13.57	1.54	–	–	–
Pain/Discomfort	61.28	34.71	4.01	–	–	–
Anxiety/Depression	73.86	23.57	2.57	–	–	–
HUI2						
Sensation	44.54	43.54	10.76	1.16		–
Mobility	87.24	8.48	3.60	0.68	0.00	–
Emotion	69.20	27.85	1.82	0.65	0.49	–
Cognition	68.36	29.94	1.63	0.08	–	–
Self-care	96.64	2.95	0.19	0.22	–	–
Pain	48.17	40.94	7.10	2.98	0.81	–
HUI3						
Vision	48.50	47.87	1.00	2.47	0.03	0.14
Hearing	94.58	0.92	1.52	1.65	0.30	1.03
Speech	92.68	4.82	2.03	0.43	0.03	–
Ambulation	87.24	8.48	2.55	1.06	0.51	0.16
Dexterity	92.44	5.82	0.79	0.70	0.14	0.11
Emotion	72.50	22.32	3.74	1.16	0.27	–
Cognition	68.36	4.15	17.85	7.37	2.19	0.08
Pain	49.34	33.73	11.46	4.01	1.46	–

Results

The mean age of the respondents was 42.9 years (range: 18.0–99.3 years), with 42.2% of the respondents being male. White (non-Hispanic) respondents were 1,435 (38.9%), non-Hispanic blacks were 1,018 (27.6%) and Hispanic were 1,100 (29.8%).

Table 2 shows the frequencies of responses to the EQ-5D, HUI2 and HUI3 dimensions. The dominant response was 'no problems' (level 1) for all dimensions in all instruments, with a proportion larger than 90% for 1 out of 5 dimensions in EQ-5D (Self-Care), 1 out of 6 in HUI2 (Self-Care) and 3 out of 8 in HUI3 (Hearing, Speech, Dexterity). In all EQ-5D and HUI2 dimensions, frequencies decreased with increasing level severity. In the HUI3 Cognition dimension however, more respondents reported problems at level 3 (17.9%) and level 4 (7.4%) than at level 2 (4.1%). Although small, these differences also occurred in the HUI3 Vision and HUI3 Hearing dimensions.

Figure 2 shows absolute informativity (Shannon's H') and relative informativity (Shannon's Evenness J') of the common dimensions among the three instruments. Absolute informativity (H') was highest for HUI3 in all common dimensions, with largest differences between HUI3 and the other two instruments in the dimensions Pain/Discomfort (0.52 compared to EQ-5D; 0.15 compared to HUI2) and Cognition (0.41 compared to HUI2).

Relative informativity (J') was highest for EQ-5D in all common dimensions, with largest differences with the other two instruments in the dimensions Mobility/Ambulation (0.14 compared to HUI2; 0.16 compared to HUI3) and Anxiety/Depression/Emotion (0.14 compared to HUI2; 0.13 compared to HUI3).

Table 3 shows Shannon's indices by classification system as a whole. The EQ-5D, HUI2 and HUI3 descriptive systems distinguished 91, 322, and 694 observed different unique health states, accounting for 37.4%, 4.0%, and 0.07% of all possible permutations, respectively. The estimation procedure indicated that absolute informativity was highest for HUI3 (range 10.96–13.36), followed by HUI2 (range 8.57–9.48), and lowest for EQ-5D (range 6.24–6.41). Relative informativity was highest in EQ-5D (range 0.79–0.81), followed by HUI2 (range 0.66–0.73), and lowest for HUI3 (range 0.55–0.67).

Discussion

We compared the discriminatory power of the EQ-5D, HUI2 and HUI3 in the general population, using Shannon's indices of absolute and relative informativity, for each dimension separately and by MAUI as a whole.

As might be expected in a general population sample, most respondents reported no problems on all
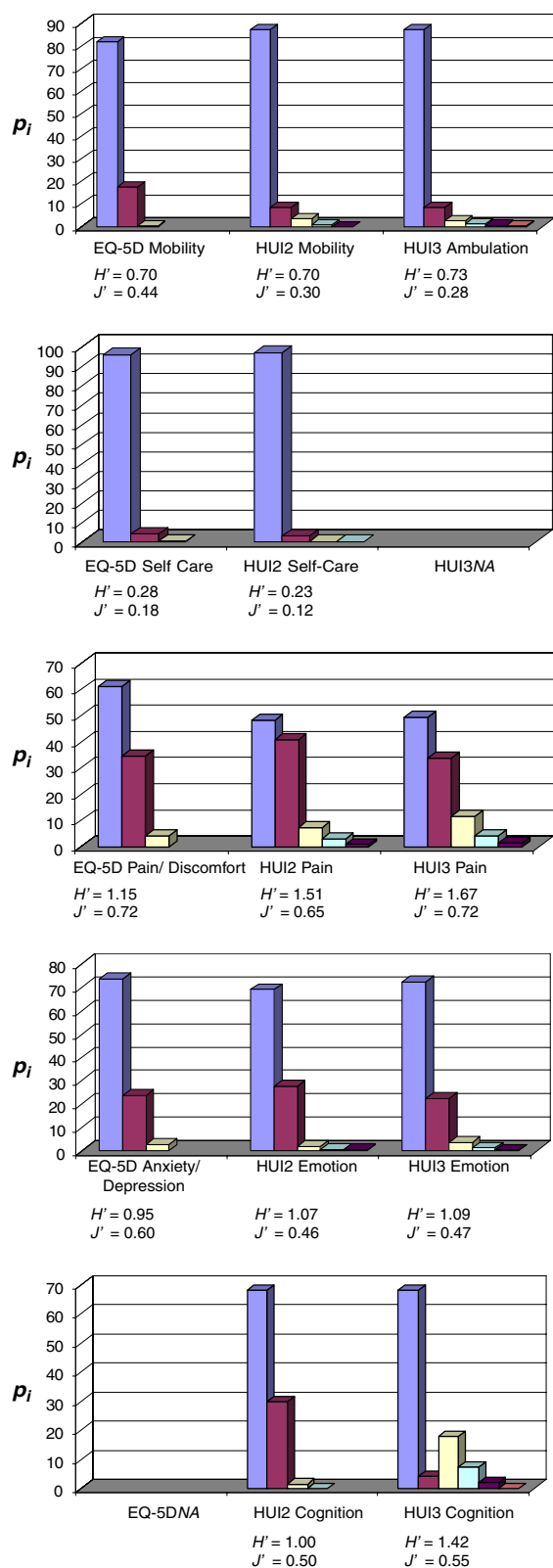
Fig. 2 The Shannon Index (H') and the Shannon Evenness Index (J')[†] for the common dimensions between EQ-5D, HUI2 and HUI3: comparison by dimension. NA = not available; [†] Confidence intervals had an average range of 0.0012 (H') and of 0.00081 (J')

dimensions and there were fewer responses with increasing level severity. An exception is HUI3 Cognition, where respondents reported more problems on levels 3 and 4 than on level 2. This is probably due to the fact that this dimension is not unidimensional, and levels 2 and 3 are conceptualized parallel rather than ordinal. That is, HUI3 Cognition level 2 focuses on problems in thinking and problem solving, level 3 addresses problems in remembering, whereas level 4 combines the problems mentioned in levels 2 and 3.

Absolute informativity by dimension was highest for the HUI3 descriptive system. EQ-5D appears to underperform in the Pain/Discomfort dimension. Moreover, EQ-5D appears to miss a considerable 'amount' of disability: 61.3% of the population indicated to have no problems on EQ-5D, against 48.2% on HUI2 and 49.3% on HUI3 (Table 2). Shannon's H' 'translated' this difference adequately (Fig. 2). Apparently, for this population, the EQ-5D would benefit from more levels on the Pain/Discomfort dimension. Regarding the Cognition dimension, the difference in absolute informativity between HUI2 and HUI3 might be explained by the 2 extra levels in HUI3, but the higher J' value in HUI3 suggests an alternative contributive factor. One explanation may be that HUI3 Cognition is not unidimensional and more sensitive to mild problems (levels 2–4) than HUI2 Cognition (level 2). Another explanation could be that the difference is due to currently suboptimal recoding algorithms.

For relative informativity by dimension, the EQ-5D descriptive system showed superior results in Mobility/Ambulation, Self-Care and Anxiety/Depression/Emotion. The large differences in Mobility/Ambulation could be due to a relatively large leap in the grading of the level descriptions in HUI3 Ambulation, where the difference between level 1 ('without difficulty') and level 2 ('with difficulty') can be considered disproportionately large in a 6 level dimension. The same leap from level 1 ('normal') to level 2 ('with difficulty') occurs in HUI2 Self-Care. We found that the 3 level EQ-5D Self Care outperformed the 4 level HUI2 Self-Care in both absolute and relative informativity (Fig. 2), which is probably due to the severe grading of level 2 in HUI2. The difference in relative informativity between EQ-5D and the HUI instruments in Anxiety/Depression/Emotion is probably due to the 2 extra levels in HUI2 and HUI3 that are rarely endorsed.

Overall, performance in terms of informativity of EQ-5D, HUI2 and HUI3 of the common dimensions varies over dimensions. The Pain/Discomfort dimension of EQ-5D, but perhaps also other dimensions, might benefit from an extension to 4 or 5 levels. HUI2

Table 3 Shannon's index (H') and Shannon's evenness index (J') for EQ-5D, HUI2, and HUI3: Comparison by instrument

	EQ-5D		HUI2		HUI3	
P_{max} (permutations)	243		8000		972,000	
Observed health states	91		322		694	
H'_{max}	7.92		12.97		19.89	
Estimation	H'	J'	H'	J'	H'	J'
$N = 1,000,000$	6.24	0.79	8.57	0.66	10.96	0.55
$N = 10,000,000$	6.37	0.80	9.12	0.70	12.29	0.62
$N = 100,000,000$	6.41	0.81	9.48	0.73	13.36	0.67

and HUI3 might benefit from more sensitive grading terms in their level descriptions, especially the 'threshold' level 2, in Ambulation (HUI3) and Self-Care (HUI2).

When assessing informativity by instrument, HUI3 shows the best results on absolute informativity but the lowest on relative informativity while EQ-5D shows highest relative informativity and lowest absolute informativity. HUI2 seems to be the optimal compromise. The importance of differences in the Shannon indices ultimately requires empirical evidence over a wider range of populations, conditions and instruments, including evidence on discriminant validity.

As Shannon's indices are new in the field of health status measurement, some methodological issues need to be addressed, taking into account that their principal focus is on classifications with mutually exclusive categories, rather than conventional (health status) measures which by design contain multiple partially overlapping items.

The Shannon indices share some properties with reliability coefficients. Like reliability indices, they express discriminatory power. Furthermore, they are also non-dimensional, i.e. they have no relation to the content, meaning or clinical relevance of what the instrument aims to measure, which make them suitable for comparability, between instruments as well as between populations. However, reliability reflects two different concepts: discriminatory power as such, and consistency, e.g. consistency between raters (inter-rater reliability) or consistency over time (test-retest reliability). This requires a repeated measurement (repetition 'over raters' or over time) which introduces an error component in case of a difference among the repeated measurements. Shannon's indices solely reflect discriminatory power, and need only a single measurement. Furthermore, the Shannon indices are non-parametric measures and therefore highly suitable for nominal or ordinal measurement scales.

Since Shannon's indices have no dimension and are independent of any external standard, a rectangular distribution is always the ideal from the perspective of informativity. When comparing the discriminatory power of similar dimensions of different MAUIs, rectangularity is always optimal as it reflects which MAUI is the most sensitive in discriminating between different persons *in that particular population*. This implies that one MAUI cannot be superior in varying populations (e.g. a general population and a diseased population sample). Furthermore, MAUIs are bound to score rather low on discriminatory power in a general population sample, as the extreme categories, which have to be included for coverage of the full spectrum of diseases, will not be endorsed frequently.

Previously, the common approach to investigate discriminatory power was examining the frequency distributions of responses, e.g. for ceiling or floor effects. A comprehensive, formal measure to express discriminatory power such as Shannon's indices seems clearly superior to such a 'face-value' method. Furthermore, when the number of categories is large (e.g. when comparing MAUIs as a whole), it becomes very difficult to make a sound comparison by just looking at the distributions.

We have demonstrated the use of the Shannon indices to compare the discriminatory power of different MAUIs, to show which instrument is more sensitive in differentiating between levels of health in the population at hand. But they may also be used to guide the development of new, or optimization of existing MAUIs, by helping determine how many levels are efficient for each dimension. This is a particularly relevant consideration for MAUIs, since adding extra levels in a descriptive system makes it increasingly complex, and the derivation of a robust set of preference weights more challenging.

Apart from MAUIs, the Shannon indices can also be used in a wide range of other classifications in the medical domain (e.g. the Karnofsky scale, the Spitzer QL index) and in the clinical domain (e.g. the APGAR score, the Child-Pugh classification).

A practical weakness of the Shannon approach is that when the sample size is exceeded by the total number of health states described by all permutations across all dimensions of a MAUI, informativity (for the instrument as a whole) has to be estimated. This implies that using the Shannon Evenness index by instrument is not very practical when a health classification system has a

large number of permutations as was the case in HUI3 (972,000 permutations). This however is not a disadvantage of the Shannon methodology per se, but also a matter of classification design (overload of dimensions with detailed response options producing an excessive amount of 'empty' permutations), or a practical problem (excessive data collection).

From a clinical or psychometric perspective it may seem tempting to extend any MAUI with extra levels or dimensions as it provides more clinically relevant detail generally and improves reliability. But Shannon's indices reveal that this may not always be a prudent approach. Increasing the number of levels per dimension (or permutations in the entire system) will probably result in higher H' values but J' values are likely to drop, as in fact our results for HUI3 indicate. This raises the question where the balance between H' and J' is optimal as more categories require more extensive subsequent studies to derive utility functions for the associated classification system.

How the Shannon indices will behave in a different population, such as patient populations, remains to be investigated. So far, Shannon's indices proved to be useful in showing weaknesses of level gradings used in EQ-5D, HUI2 and HUI3, and offers leads for improvement, establishing their practical psychometric value.

References

1. Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality-of-life. *Annals of Internal Medicine, 118*, 622–629.
2. Mishoe, S. C., & Maclean, J. R. (2001). Assessment of health-related quality of life. *Respiratory Care, 46*, 1236–1257.
3. Brazier, J., Deverill, M., Green, C., Harper, R., & Booth, A. (1999). A review of the use of health status measures in economic evaluation. *Health Technology Assessment, 3*, 1–164.
4. Coons, S. J., Rao, S., Keininger, D. L., & Hays, R. D. (2000). A comparative review of generic quality-of-life instruments. *Pharmacoeconomics, 17*, 13–35.
5. Rabin R., de Charro F. (2001) EQ-5D: a measure of health status from the EuroQol Group. *Annals of Medicine 33*, 337–343.
6. Feeny, D., Furlong, W., & Torrance, G. (1999). The health utilities index: An update. *Quality of Life Newsletter, 22*, 8–9.
7. Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The Health Utilities Index (HUI(R)): Concepts, measurement properties and applications. *Health Qual Life Outcomes, 1*, 54.
8. Brooks, R., Rabin, R. E., de Charro, F. Th. (2003). *The measurement and valuation of health status using EQ-5D: A European perspective*. Dordrecht: Kluwer academic publishers.
9. 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.
10. Kopec, J., & Willison, K. D. (2003). A comparative review of four preference-weighted measures of health-related quality of life. *Journal of Clinical Epidemiology, 56*, 317–325.
11. Brazier, J., & Deverill, M. (1999). A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Economics, 8*, 41–51.
12. Brazier, J., Jones, N., & Kind, P. (1993). Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Quality of Life Research, 2*, 169–80.
13. Sintonen, H. (2001). The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine, 33*, 328–336.
14. Murray, C. J., & Lopez, A. D. (2000). Progress and directions in refining the global burden of disease approach: a response to Williams. *Health Economics, 9*, 69–82.
15. Guyatt, H., Kirshner, B., & Jaeschke, R. (1992). Measuring health status: what are the necessary measurement properties? *Journal of Clinical Epidemiology, 45*, 1341–1345.
16. Streiner, D. L., & Norman, G. W. (1995). *Health measurement scales: A practical guide to their development and use*. Oxford: Oxford University Press.
17. Shannon, C. E (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379–423.
18. Macran, S., Weatherly, H., & Kind, P. (2003). Measuring population health—A comparison of three generic health status measures. *Medical Care, 41*, 218–231.
19. Myers, C., & Wilks, D. (1999). Comparison of Euroqol EQ-5D and SF-36 in patients with chronic fatigue syndrome. *Quality of Life Research, 8*, 9–16.
20. Conner-Spady, B., & Suarez-Almazor, M. E. (2003). Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Medical Care, 41*, 791–801.
21. Shaw, J. W., Johnson, J. A., & Coons, S. J. (2005). US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Medical Care, 43*, 203–220.
22. Johnson, J. A., Luo, N., Shaw, J. W., Kind, P., & Coons, S. J. (2005). Valuations of EQ-5D health states—Are the United States and United Kingdom different? *Medical Care, 43*, 221–228.
23. Furlong, W., Feeny, D., & Torrance, G. W. (2000). *Health Utilities Index (HUI ®) procudures manual: algorithm for determining HUI Mark 2 (HUI2)/Mark 3 (HUI3) health status classification levels, health states, single attribute level utility scores and overall health-related quality of life utility scores, from 15-item self-complete health status questionnaires*. Dundas: Ontario, Health Utilities Inc.
24. Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
25. Wiener, N. (1948). *Cybernetics: or, control, and communication in the animal and the machine*. Cambridge: MIT Press.
26. Krebs, C. J. (1989). *Ecological methodology*. New York: Harper & Row.
27. Miller, G. A, & Frick, F. C. (1949). Statistical behaviorists and sequences of responses. *Psychological Review, 56*, 311–24.

28. Reitsma, H. (1999). *Registers in cardiovascular epidemiology*. University of Amsterdam

29. Rao, G. S., Hamid, Z., & Rao, J. S. (1979). The information content of DNA and evolution. *Journal of Theoretical Biology, 81*, 803–807.

30. Yockey, H. P. (1992). *Information theory and molecular biology*. Cambridge: Cambridge University Press.

31. Basharin, G. P. (1959). On a statistical estimate for the entropy of a sequence of independent random variables. *Theory of Probability and its Applications, 4*, 333–336.

32. Lloyd, M., & Ghelardi, R. J. (1964). A table for calculating the 'equitability' component of species diversity. *Journal of Animal Ecology, 33*, 217–225.

33. Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology, 13*, 131–144.

34. Peet, R. K. (1974). The measurement of species diversity. *Annual Review of Ecology and Systematics, 5*, 285–307.

35. Mouillot, D., & Leprêtre, A. (1999). A comparison of species diversity estimators. *Research on Population Ecology, 41*, 203–215.