VALUE IN HEALTH 20 (2017) 1157–1165

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Discordance in Utility Measurement in Persons with Neurological Conditions: A Comparison of the SF-6D and the HUI3

Hannah Abel, MSc[1,*], *George Kephart, PhD*[1], *Tanya Packer, PhD*[2], *Grace Warner, PhD*[2]

[1]*Department of Community Health and Epidemiology, Dalhousie University, Halifax, Nova Scotia, Canada;* [2]*School of Occupational Therapy, Dalhousie University, Halifax, Nova Scotia, Canada*

ABSTRACT

Objectives: To examine the extent of disagreement in estimated utility between the six-dimensional health state short form (SF-6D) and the Health Utilities Index—Mark 3 (HUI3) in Canadians with neurological conditions and how discordance varied by participant and neurological condition attributes. **Methods:** The study analyzed cross-sectional survey data from the Living with and Managing the Impact of a Neurological Condition Study. Self-reported data were collected on the burden and impact of neurological conditions on participants' everyday lives. Disagreement was examined by comparing utility distributions, paired *t* tests of the means, Spearman ρ correlations, intraclass correlations, and Bland-Altman plots. Associations between participant and neurological condition attributes and utility differences were assessed using multiple regression models. **Results:** Disagreement between the SF-6D and the HUI3 was substantial, with a mean utility difference of 0.15 (95% confidence interval 0.13–0.17). An intraclass correlation coefficient of 0.41 suggests only marginal agreement. The Bland-Altman plot and regression analysis showed systematic variation in utility difference associated with level of utility. Depending on the level of utility, utility differences between the SF-6D and the HUI3 shift in magnitude and direction. The pattern of disagreement did not vary substantially by participant or neurological condition characteristics. **Conclusions:** The SF-6D and the HUI3 provide inconsistent evaluations of utility in persons with neurological conditions. The magnitude and direction of differences in estimated utility are strongly associated with level of utility. Depending on the health status of the sample, the SF-6D and the HUI3 could provide widely contradictory utility estimates. A concern is that utility scores, and hence potential evaluations and health care decisions, may vary simply according to the choice of instrument.

Keywords: discordance, health-related quality of life, Health Utilities Index, neurological conditions, Neuro-QoL, preference-based quality of life, SF-6D, utility measures.

Introduction

Quality of life is frequently compromised for people living with neurological conditions that are diverse in their symptoms, trajectories, and sequelae. Collectively, neurological conditions are a significant cause of disease burden and are characterized by diverse functional limitations that pose daily challenges to individuals and their families [1–5]. Very few neurological conditions are curable. As a result, the treatment emphasis is often on managing the impact of symptoms and the maintenance of health-related quality of life (HRQOL) [4–6].

HRQOL is a multidimensional construct that represents an individual's subjective perception of improvements or detriments in quality of life and well-being that are central to health such as the physical, mental, and psychosocial-emotional aspects of health [7]. Assessments of HRQOL contribute to a more comprehensive understanding of disease effects and impacts on quality of life than do measures of clinical change alone [4,8–11]. There are various condition-specific and generic instruments to evaluate HRQOL. Although condition-specific instruments can target domains that are relevant to an individual condition, generic instruments facilitate comparative assessments across conditions and in persons with multimorbidity [12].

Multi-attribute, preference-based HRQOL instruments (routinely referred to as "utility" instruments), such as the Health Utilities Index—Mark 3 (HUI3) [13], the six-dimensional health state short form (SF-6D) [14,15], and the EuroQol five-dimensional questionnaire (EQ-5D) [16], are a class of generic instruments that evaluate HRQOL in terms of utility [7]. These instruments are standardized outcome measures designed to describe the extent of disability and impact of ill health by attribute and also assess the strength of people's *preferences* for health states as a single

Conflicts of interest: The authors have no conflicts of interest to declare.

* *Address correspondence to:* Hannah Abel, Department of Community Health and Epidemiology, Dalhousie University, 5790 University Avenue, Halifax, Nova Scotia B3H 1V7, Canada.

E-mail: Hannah.Abel@dal.ca.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2017.04.008

utility-based index score, where 1 signifies perfect health and 0 represents health states considered equivalent to death. It is possible to obtain scores less than 0, which represent health states considered worse than dead [7,17]. Preference-based instruments consist of a descriptive health classification system, comprising a number of multilevel HRQOL domains, and a scoring algorithm that assigns utility to health states [7]. The scoring algorithm incorporates "preferences" that reflect a valuing of different health states. Preferences can be obtained either directly from the participant, an individual experienced in the health state, or indirectly from a representative community sample [12]. Instruments using indirect preferences therefore reflect "societal" rather than individual preferences for health states [8,12,18].

Even though utility instruments are purported to measure the same concept (HRQOL), they vary in their descriptive health classification systems, valuation methods, and scoring algorithms. As a result, different utility instruments often produce dissimilar utility scores and distributions and vary in their sensitivity and responsiveness [8,19–21]. The accumulating empirical evidence of discrepancies between utility instruments is of significant concern and has important implications for interpreting HRQOL status and challenges the validity required for decision making. The concern is that utility scores, and hence potential evaluations and health care decisions, may vary simply according to the choice of instrument [17,22,23].

The agreement and discrepancies between utility instruments have been assessed in head-to-head comparisons in various populations and conditions, including specific neurological conditions; it has not been evaluated in a sample of people with diverse neurological conditions [4,19,24,25]. The aim of this study was to examine the extent of disagreement in utility scores between the SF-6D and the HUI3 in Canadian adults with neurological conditions and determine whether utility differences vary by participant and neurological condition attributes.

Methods

Data

The data were drawn from the Canadian LINC (Living with and Managing the Impact of a Neurological Condition) Study. The purpose of the study was to describe the health burden and impact on health, quality of life, resource utilization, and participation in everyday life of people with neurological conditions [26]. The LINC Study collected data in a cross-sectional survey of 948 Canadians, 17 years of age or older, with at least one identified neurological condition, between September 14, 2011, and July 1, 2012. Subjects were recruited primarily through the Neurological Health Charities of Canada and their member organizations. The LINC Study survey instrument was constructed using a series of validated scales and standardized instruments as well as self-reported data on diagnoses and health status (diagnosis, time since diagnosis, self-rated health, and the degree of neurological impairment and symptoms) and sociodemographic variables (age, sex, employment status, and education). Question formats matched those used in Statistics Canada health surveys. The LINC Study survey was primarily administered online using Opinio (ObjectPlanet, Inc. Øvre Slottsgate 5, 0157 Oslo, Norway), a Web-based survey application. The remaining respondents completed the survey on paper (8%) or by telephone (3%). Data were cleaned and coded using Excel and SAS (SAS Institute, Inc., Cary, NC). Coding, including handling of missing data, was as per protocols for each instrument.

This analysis used data from survey respondents (N = 948) who had utility scores for both the SF-6D and the HUI3 (N = 776). Ethical approval was received from Health Canada and the Public Health Agency of Canada Research Ethics Board as well as the appropriate ethics review boards at Dalhousie University, Queens University, the University of Manitoba, and the University of Prince Edward Island.

Instruments

Three HRQOL instruments were used in the analysis: the HUI3, the SF-6D, and the Quality of Life in Neurological Disorders (Neuro-QoL). The HUI3 and the SF-6D are generic, indirect, preference-based multi-attribute health state classification and utility scoring systems. The Neuro-QoL is a condition-specific psychometric measure for use in neurological diagnoses.

The HUI3 health classification system includes eight HRQOL domains with a focus on bodily functions: vision, hearing, speech, dexterity, ambulation, cognition, emotion, and pain. Each domain has five to six levels of function. The HUI3 can describe 972,000 unique health states. The HUI3 scoring function was developed by rating single-deficit states using the visual analogue scale and standard gamble techniques in a multi-attribute multiplicative model and incorporated preferences of a community sample (n = 504) in Ontario, Canada. HUI3 utility scores can range from −0.36 to 1.00. Minimally important clinical difference in HUI3 utility scores range between 0.02 and 0.05 [13,27]. The HUI3 is widely used in general and clinical applications including various neurological conditions [25,28–30]. It has been included in major population health surveys in Canada since 1990, including more recently (2010) the Survey of Living with a Neurological Condition in Canada [31,32].

The SF-6D is derived from the widely used short form 36 health survey using an algorithm developed by Brazier et al. [14,15,33,34]. Although less frequently used than other utility measures such as the HUI3 and the EQ-5D, it provides a means of measuring utility when data from the short form 36 health survey are available. The SF-6D describes a broader definition of health within its six domains: physical functioning, role limitations (physical and emotional), bodily pain, vitality, social functioning, and mental health. Each domain has one to three items, with four to six levels of function to describe up to 18,000 unique health states. SF-6D health state valuation used standard gamble valuations that included complex, multideficit states from a sample of 836 individuals in the United Kingdom. A linear additive model was used to aggregate domain scores. The SF-6D utility scores range from 0.30 to 1.00 [14,15,33]. Minimally important clinical difference in SF-6D scores range from 0.011 to 0.097 in different clinical populations [35–37]. The SF-6D has been administered in various patient populations with neurological conditions [6,24,25].

The LINC Study evaluated neurological condition–specific attributes using the Neuro-QoL instrument. The Neuro-QoL has a series of separate psychometric scales that assess domains relevant to neurological conditions (see Table 2 for the 13 included scales). Each scale uses eight to nine items to describe a domain. Neuro-QoL scale T scores (mean 50, SD 10), derived from the scoring tables in the Neuro-QoL user manual, were used as neurological condition attribute covariates in the analysis. A 0.5 SD change in Neuro-QoL T score is considered a clinically relevant difference [10,11,38].

Analysis

Discrepancy or difference in measured utility between the SF-6D and the HUI3 was the primary outcome variable for this study. Difference in utility (dUTY) scores was measured for each person, i, as follows:

$$dUTY_i = SF6D_i - HUI3_i$$

Variation in dUTY was examined with respect to a number of patient- and condition-specific attributes. Patient attributes included sex, age group (<26, 27–55, 56–65, and >65 years), educational attainment (less than high school diploma, high school diploma, some postsecondary, postsecondary degree), years since diagnosis (0–5, 6–10, and >10), and change in health status since previous year.

Descriptive statistics (means, SDs, and percentiles) and graphs were used to describe sample characteristics and the distributions of the SF-6D and the HUI3. Differences in mean utility scores were assessed using paired *t* tests and associations between each of the SF-6D and the HUI3 domains were evaluated using a Spearman correlation matrix, in which correlations greater than 0.50 represented a strong relationship (results not

Table 1 – Characteristics of the study population (n = 776).

Characteristic/condition	n (%)	SF-6D (95% CI)	HUI3 (95% CI)	Mean difference*
Sample characteristics				
Whole sample utility score by instrument	776 (100)	0.62 (0.62–0.63)	0.47 (0.45–0.49)	0.15 (0.13–0.17)
Sex				
Male	244 (35)	0.62 (0.61–0.64)	0.43 (0.39–0.47)	0.19 (0.16–0.22)
Female	444 (65)	0.63 (0.62–0.64)	0.49 (0.46–0.52)	0.13 (0.11–0.16)
Age† (y)				
<26	44 (6)	0.66 (0.63–0.70)	0.49 (0.39–0.59)	0.18 (0.09–0.26)
27–55	368 (53)	0.61 (0.60–0.62)	0.47 (0.44–0.50)	0.14 (0.12–0.17)
56–65	163 (24)	0.62 (0.60–0.64)	0.44 (0.39–0.49)	0.18 (0.14–0.22)
>65	115 (17)	0.66 (0.64–0.68)	0.51 (0.45–0.56)	0.15 (0.11–0.19)
Education				
No high school	45 (7)	0.60 (0.56– 0.64)	0.37 (0.27–0.46)	0.24 (0.16–0.31)
High school	78 (12)	0.61 (0.59–0.63)	0.41 (0.34–0.47)	0.20 (0.15–0.26)
Some postsecondary	53 (8)	0.62 (9.58–0.65)	0.41 (0.34–0.49)	0.21 (0.14–0.27)
Postsecondary degree	491 (73)	0.63 (0.62–0.64)	0.50 (0.48–0.53)	0.13 (0.11–0.15)
Self-rated health				
Excellent	44 (6)	0.73 (0.68–0.77)	0.60 (0.50–0.69)	0.13 (0.05–0.21)*
Very good	171 (22)	0.68 (0.66–0.69)	0.63 (0.58–0.67)	0.05 (0.01–0.09)*
Good	304 (39)	0.63 (0.62–0.64)	0.51 (0.48–0.54)	0.12 (0.09–0.14)
Fair	194 (25)	0.58 (0.57–0.59)	0.35 (0.31–0.39)	0.23 (0.20–0.26)
Poor	57 (7)	0.50 (0.48–0.53)	0.13 (0.06–0.19)	0.38 (0.32–0.43)
Time since diagnosis (y)				
0–5	146 (19)	0.63 (0.47–0.57)	0.52 (0.47–0.57)	0.12 (0.08–0.16)
6–10	128 (16)	0.60 (0.58–0.62)	0.43 (0.38–0.49)	0.16 (0.12–0.21)
>10	502 (65)	0.63 (0.62–0.64)	0.47 (0.44–0.50)	0.16 (0.14–0.18)
Health in past year				
Health same or better	510 (66)	0.65 (0.64–0.66)	0.53 (0.50–0.55)	0.12 (0.09–0.14)
Health declined	265 (34)	0.57 (0.52–0.58)	0.36 (0.33– 0.40)	0.21 (0.18–0.24)
Selected major neurological conditions within the sample‡				
ALS	26	0.58 (0.55–0.61)	0.26 (0.17–0.36)	0.32 (0.23–0.40)
Alzheimer disease	11	0.61 (0.54–0.68)	0.31 (0.11–0.51)	0.30 (0.12–0.48)*
Epilepsy	117	0.65 (0.62–0.67)	0.52 (0.46–0.59)	0.13 (0.08–0.17)
Brain injury	108	0.61 (0.59–0.63)	0.38 (0.32–0.44)	0.23 (0.18–0.28)
Brain/spinal cord tumor	24	0.60 (0.54–0.65)	0.42 (0.29–0.55)	0.18 (0.07–0.28)*
Cerebral palsy	14	0.58 (0.52–0.63)	0.44 (0.29–0.60)	0.13 (0.01–0.26)*
Dystonia	43	0.65 (0.61–0.69)	0.56 (0.47–0.64)	0.09 (0.03–0.16)*
Multiple sclerosis	183	0.61 (0.59–0.62)	0.44 (0.40–0.49)	0.16 (0.13–0.20)
Muscular dystrophy	65	0.61 (0.59–0.63)	0.38 (0.31–0.45)	0.23 (0.17–0.29)
Parkinson disease	147	0.64 (0.62–0.66)	0.54 (0.49–0.57)	0.10 (0.06–0.14)
Spina bifida and hydrocephalus	62	0.62 (0.59–0.65)	0.47 (0.40–0.54)	0.15 (0.08–0.23)
Spinal cord injury	47	0.59 (0.56–0.62)	0.37 (0.29–0.45)	0.22 (0.15–0.29)
Stroke	38	0.59 (0.56–0.63)	0.34 (0.24–0.43)	0.26 (0.18–0.34)

ALS, Amyotrophic lateral sclerosis; HUI3, Health Utilities Index—Mark 3; SF-6D, six-dimensional health state short form.

* Difference was compared using two-sided paired *t* tests. Mean differences between the SF-6D and the HUI3 are all significant at a *P* value <0.001 except self-rated health: Excellent *P* value 0.003; self-rated health: Very good *P* value 0.007; cerebral palsy *P* value 0.0390; dystonia *P* value 0.0080; Alzheimer disease *P* value 0.0036; and brain/spinal cord tumor *P* value 0.0019.

† A focus of the LINC Study was the impact of neurological conditions on everyday life; hence, age grouping reflects developmental stages related to education, work, and retirement. People younger than 27 y are often still in school or beginning their careers; people aged 27–55 y are within the typical working years. The age group of 56–65 y finds people with mixed working/retired profiles, particularly true for people with chronic conditions. Most people older than 65 y are retired.

‡ Frequency and utility of selected neurological conditions were included in the sample. The sample also included Huntington's (n = 7) and Tourette's (n = 8).

shown) [12,39]. Ceiling and floor effects were considered present if more than 10% of scores fell within the upper limit (between 0.90 and 1.00) or the lower limit (between 0.30 and 0.40 in the SF-6D and between −0.28 and −0.18 in the HUI3) of an instrument's distribution [40].

Agreement between SF-6D and HUI3 utilities was assessed using a two-way mixed model intraclass correlation (ICC) in which the SF-6D and HUI3 utility scores were treated as fixed effects and the interactions between the participant and the instrument were treated as random effects [41]. 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 less than 0.40 was considered poor agreement [42].

Difference in utility (dUTY) and its relationship with HRQOL status (the average of the SF-6D and the HUI3, or mean utility) were examined using a Bland-Altman plot, with dUTY plotted on the y-axis and the mean of the two utilities on the x-axis. The variation in paired observations was estimated as the 95% limits of agreement (mean ± 1.96 [SD]) [43].

The effect of participant and neurological condition attributes (Neuro-QoL domains) on dUTY was explored using ordinary least squares multiple regression to assess whether the magnitude and pattern of utility discrepancy varied by patient and condition attributes. A base regression model describing the relationship between dUTY and mean utility observed in the Bland-Altman plot was first estimated. Because the relationship was nonlinear, three linear splines with knots at mean utility of 0.30 and at 0.77 were used. Sensitivity analysis, using five splines, was conducted but did not alter the results. Participant and neurological (Neuro-QoL) attribute covariates were then added to the base model, one at a time, to assess their association with dUTY. Main effects were used to assess mean differences and interactions with the spline coefficients to assess differences in slope. Regression coefficients were assessed for statistical significance (t tests; P value ≤ 0.05) and for the magnitude of change in utility discrepancy from the baseline model. The joint statistical significance of groups of variables, or interactions, was assessed using F tests for the reduction in sum-of-squares error. Minimal clinically important differences, or the smallest change in score that patients would perceive as beneficial or detrimental, vary between utility instruments and clinical populations [27,35,44]. On the basis of reports of clinically meaningful change in the SF-6D and the HUI3, the frequently cited minimal clinically important dUTY of 0.03 was adopted as a threshold of clinically meaningful discrepancy between the SF-6D and the HUI3 [27,35]. All analyses were conducted using STATA 12 (StataCorp, College Station, Texas).

Results

The 776 individuals included in the analyses had a mean age of 50.8 years (95% confidence interval [CI] 49.7–51.9), a mean of 17.3 years (range <1 year to 71 years) since diagnosis (95% CI 16.2–18.5), and were predominantly female (64.5%). Participants reported, on average, 1.5 ± 0.83 neurological conditions. Despite 67.5% of the sample self-rating their health as good, very good, or excellent, the SF-6D and the HUI3 summary utility scores were very low (Table 1).

A high level of disagreement was observed between the SF-6D and the HUI3 utility scores (Fig. 1 and Table 1). The HUI3 had a mean utility of 0.47 (95% CI 0.45–0.49) with a range of −0.28 to 1.00, whereas the SF-6D had a higher mean utility of 0.62 (95% CI 0.62–0.63) and a narrower range of 0.30 to 1.00. The HUI3 showed a mild ceiling effect, with 10.2% of participants between a utility of 0.9 and 1.00 compared with only 1.8% in the SF-6D. Neither instrument showed a floor effect as defined in our methods.

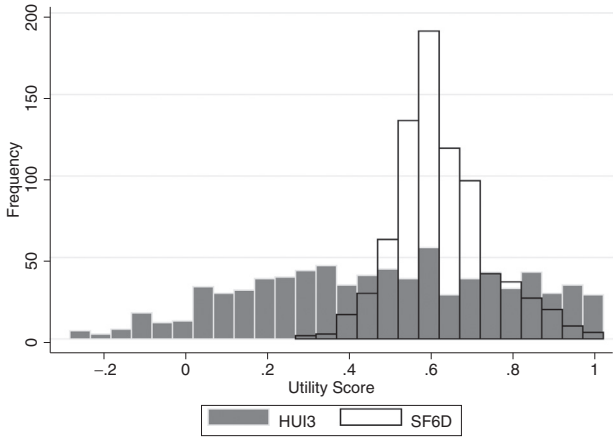

Fig. 1 – Distribution of the SF-6D and the HUI3. HUI3, Health Utilities Index—Mark 3; SF-6D, six-dimensional health state short form.

There was, however, a clustering of SF-6D scores (68.3%) between utilities of 0.50 and 0.69. The mean dUTY was 0.15 (95% CI 0.13–0.17) and 92% of respondents had a dUTY greater than the predefined disagreement threshold of 0.03. Significant utility discrepancies across all participant attributes were confirmed by paired t tests (Table 1). Despite a strong linear association between the SF-6D and the HUI3 utility scores (Pearson $r = 0.62$, $P < 0.0001$, and Spearman $\rho = 0.61$, $P < 0.0001$), the ICC demonstrated marginal agreement, with a coefficient of 0.41 (95% CI 0.35–0.46) suggesting that only 41% of the variation in utility was due to between-subject differences.

Differences between the two measures depend greatly on the level of HRQOL (mean utility), a relationship that is not evident by examining mean dUTY alone. In the Bland-Altman plot (Fig. 2), the data points deviate widely from the agreement line in a reverse checkmark pattern. The result of this pattern is that

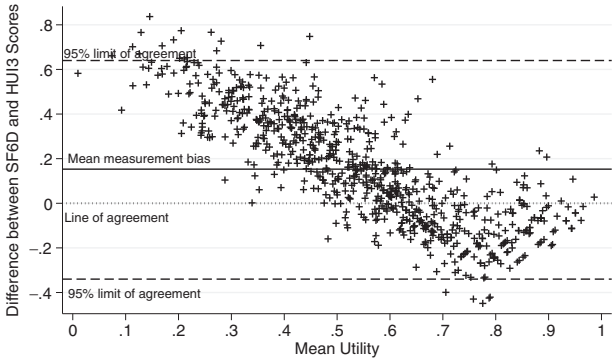

Fig. 2 – Bland-Altman plot of the SF-6D and the HUI3. The solid line at y = 0.15 indicates the mean measurement bias, or mean difference between the SF-6D and the HUI3, and indicates that overall the SF-6D provides a utility score that is 0.15 higher than that of the HUI3. The dashed lines at y = −0.34 and y = 0.65 represent the 95% limits of agreement (mean utility difference ± 1.96) and represent the expected limits of agreement for 95% of the observations. The 95% limits of agreement are wide (−0.34 to 0.65) with potential variation in utility (0.99) almost equivalent to the theoretical 0.00–1.00 utility score range. HUI3, Health Utilities Index—Mark 3; SF-6D, six-dimensional health state short form.

depending on the value of mean utility, the relationship between dUTY and utility shifts in both magnitude *and* direction. The highest values of dUTY are at the lower end of the utility distribution (lower HRQOL status, left side of Fig. 2) with the SF-6D providing a much higher utility score than the HUI3, with differences between the two instruments as high as 0.60. There is a null range between mean utility of approximately 0.56 and 0.70 where dUTY between the SF-6D and the HUI3 is negligible (dUTY $\leq$ 0.03). For utility scores more than 0.66, the HUI3 provides higher utility scores than the SF-6D (up to dUTY of 0.18). A similar pattern has been observed between the SF-6D and the EQ-5D [45].

Regression analyses assessed whether the pattern of association between dUTY and HRQOL status (mean utility) varied by patient and condition attributes (Table 2). Mean utility, modeled with three splines, explained 73% of the variation in dUTY, emphasizing the strong effect utility has on dUTY. Although most of the neurological condition attributes (Neuro-QoL covariates) showed statistically significant effects on dUTY (Table 2), the magnitude of dUTY was modest in comparison with the large

Table 2 – Multiple regression modeling associations between participant characteristics and Neuro-QoL domains and utility difference.

Participant characteristics	n (%)	Model 1*			Model 2†		
		β Coefficient‡	P value	F test	β Coefficient‡	P value	F test
Mean difference in utility	766 (100)	0.151	0.000	<0.000	0.155	0.000	<0.000
Sex							
Male	244 (35)	−0.016	0.122	0.122	−0.011	0.266	0.265
Female	444 (65)	Reference category			Reference category		
Age (y)							
<26	44 (6)	0.062	0.003	<0.001	0.062	0.003	<0.001
27–55	368 (53)	Reference category			Reference category		
56–65	163 (24)	0.021	0.081		0.021	0.093	
>65	115 (17)	0.047	0.001		0.044	0.002	
Education							
No high school diploma	45 (7)	0.009	0.667	0.483	0.002	0.908	0.512
High school diploma	78 (12)	0.022	0.157		0.023	0.153	
Some postsecondary	53 (8)	0.015	0.422		0.013	0.500	
Postsecondary degree	491 (73)	Reference category			Reference category		
Time since diagnosis (y)							
0–5	146 (19)	−0.012	0.318	0.224	−0.021	0.087	0.058
6–10	128 (16)	−0.021	0.112		−0.027	0.042	
>10	502 (65)	Reference category			Reference category		
Health in past year							
Health same or better	510 (66)	Reference category		<0.001	Reference category		<0.001
Health declined	265 (34)	−0.037	0.000		−0.040	0.000	

Neuro-QoL variables	n	Model 1* β Coefficient‡,§	Model 1* P value	Model 2† β Coefficient‡,§	Model 2† P value
Upper extremity‖	763	−0.018	0.000	−0.015	0.007
Lower extremity‖	762	−0.017	0.001	−0.013	0.019
Cognition—general‖	753	0.008	0.145	0.008	0.174
Cognition—executive function‖	747	0.007	0.229	0.008	0.154
Fatigue¶	763	−0.075	0.000	−0.073	0.000
Sleep disturbance¶	760	−0.065	0.000	−0.067	0.000
Depression¶	746	−0.068	0.000	−0.064	0.000
Anxiety¶	749	−0.060	0.000	−0.057	0.000
Emotional behavioral dyscontrol¶	748	−0.030	0.000	−0.026	0.000
Positive effect and well-being‖	746	0.055	0.000	0.053	0.000
Stigma¶	742	−0.043	0.000	−0.041	0.000
Ability to participate in social roles‖	719	0.074	0.000	0.070	0.000
Satisfaction with social roles‖	719	0.075	0.000	0.072	0.000

Neuro-QoL, Quality of Life in Neurological Disorders.

* Model 1 : Difference in utility (dUTY)$=\alpha+f(\beta_{\text{mean utility}})+\beta_j x_j$, where f is the spline function of three linear splines of mean utility and $\beta_j x_j$ are covariates of interest.

† Model 2 : Difference in utility (dUTY)$=\alpha+f(\beta_{\text{mean utility}})+\beta_2$ sex$+\beta_2$ age$+\beta_j x_j$.

‡ Significance based on a minimally important difference (effect size) of >0.03 and statistical significance P value <0.05. Coefficient represents utility difference associated with the category of interest vs. the reference category.

§ Regression coefficient represents change in utility difference for 1 SD change in the associated Neuro-QoL variable.

‖ Higher Neuro-QoL score relates to *better* self-reported health.

¶ Higher Neuro-QoL score relates to *worse* self-reported health.

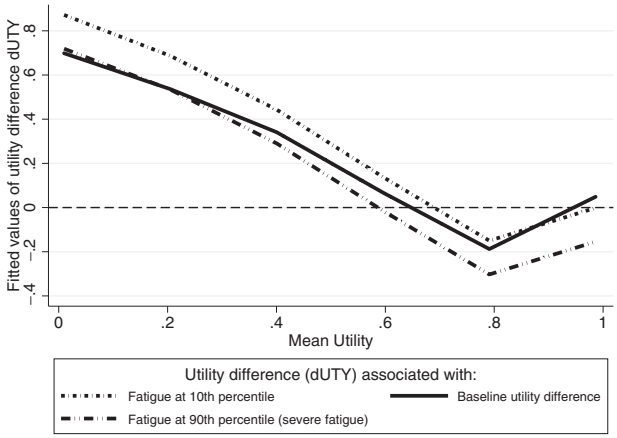

Fig. 3 – Comparison of dUTY by mild and severe fatigue. Fatigue showed the highest estimate of dUTY compared with all other covariates with a mean estimated dUTY of −0.75. This plot shows that although there are differences in the magnitude of dUTY between mild (10th percentile—upper line) and severe fatigue (90% percentile—lower line) and the baseline dUTY, the difference is modest (mean difference −0.075) compared with the substantial dUTY associated with health status (mean utility) throughout the full range of utility. At its highest, the fitted values of dUTY are estimated at 0.60 (SF-6D higher than HUI3) between mean utility values of 0.00 and 0.2 (left side of the figure). In the opposite direction, dUTY is approximately −0.18 (SF-6D lower than HUI3) in higher levels of utility (right side of the figure). dUTY, utility difference; HUI3, Health Utilities Index—Mark 3; SF-6D, six-dimensional health state short form.

effect of mean utility. For example, dUTY for the Neuro-QoL attribute fatigue is plotted against baseline dUTY in Figure 3. Although the magnitude of dUTY varies between individuals with severe fatigue and mild fatigue across the distribution of utility, the pattern of dUTY is similar to the baseline dUTY and is marginal in comparison.

Each model was rerun with interaction terms between the explanatory covariates and the splines of mean utility (results not shown). None was found to be significant, indicating that the explanatory covariates were associated only with modest variations in the magnitude of dUTY, and not with a change in the pattern of dUTY observed in the Bland-Altman plot. This remained true after adjusting for age and sex.

Discussion

Utility instruments, such as the SF-6D and the HUI3, are designed to measure the same construct (HRQOL) and are promoted as instruments that in theory permit broad comparisons across populations and interventions. The pairwise agreement of the SF-6D and the HUI3 utilities was compared in a sample of individuals with various neurological conditions and were found to provide widely divergent utility estimates that vary substantially and systematically across the entire utility distribution. The magnitude and direction of discrepancies between the SF-6D and the HUI3 utility scores were most dependent on the level of HRQOL (mean utility). At the group or population level, this means that the direction and magnitude of utility discrepancies will depend on the health and HRQOL status and the utility distribution of the study sample. This finding suggests that important health care decisions and prioritizations could be affected by the choice of instrument, particularly if used in cost-utility applications, and that either instrument could favor or disadvantage individuals depending on whether the individual has high or low health status and HRQOL.

This is the first comparison of the SF-6D and the HUI3 in a sample of individuals with various neurological conditions. The results are consistent with evidence from other studies demonstrating that there are differences in predicted utility scores in pairwise comparisons between the SF-6D and the HUI3 in the same sample [23,46–48]. Utility discrepancies in this study are the rule rather than the exception, with 92% of respondents exhibiting clinically important differences between their SF-6D and HUI3 utilities higher than the disagreement threshold of 0.03. It is important to acknowledge that minimal clinically important differences vary by population, diagnosis, and context and may be higher than our chosen disagreement threshold of 0.03 [27,35,36]. Nonetheless, the mean dUTY of 0.15 found in this study is 5 times more than the studies' threshold of disagreement and is well above the upper limits of clinically important difference for either the SF-6D or the HUI3. The clinically important differences in utility found between the SF-6D and the HUI3 further cast doubt on whether utility estimates between instruments are comparable or universally valid.

The factor that contributed most to dUTY in this study was that the SF-6D cannot provide utility scores less than 0.30. Although both instruments have a theoretical range of 0.00 to 1.00, in reality the lower limit of each instrument is very different. This affected the magnitude and pattern of discrepancies between the SF-6D and the HUI3 utilities. For example, for the 243 (31.3%) individuals who had HUI3 utility scores less than the lower limit of the SF-6D (0.30), the mean dUTY was 0.44 (95% CI 0.42–0.45). For the remaining 533 individuals with an HUI3 score greater than 0.30, the mean dUTY was 0.02 (95% CI 0.006–0.036). This finding is consistent with Richardson et al. [17] who estimated that the difference in the SF-6D and the HUI3 scales contributes to 47% of the variation between utility scores.

According to findings in this study, particular caution should be exercised when interpreting and comparing results of utilities in samples with low health status and HRQOL. The mean dUTY of 0.15 in this study was greater than previous comparison studies of the SF-6D and the HUI3 in general and clinical populations [4,24,25,49–52], but consistent with findings by Pickard et al [19] that greater differences occurred in the presence of lower utility scores. Pickard et al. observed a dUTY of 0.36 associated with a baseline mean SF-6D score of 0.55 and an HUI3 score of 0.19 in persons recovering from stroke. Therefore, in clinical samples with very low health status and subsequently very low utility, the difference in results between the two may be so substantial that utility estimates cannot be meaningfully interpreted.

Population surveys, however, are likely to sample across a broader spectrum of health and result in higher mean utility scores. For example, the Survey on Living with a Neurological Condition in Canada conducted in 2010 found a mean HUI3 utility of 0.71 across 18 selected neurological conditions [32]. dUTY associated with HUI3 scores in the range of 0.65 to 0.75 were negligible (dUTY = 0.05; 95% CI 0.04–0.07) in the present study, suggesting that the SF-6D and the HUI3 scores would be relatively more comparable in population samples with modest utility levels than in clinical samples with low health status and low mean utility. Nevertheless, within-sample subpopulations should also be considered. For example, within the Survey on Living with a Neurological Condition in Canada sample there is great variation in HUI3 utility scores between diagnoses, with a low in Alzheimer disease (0.21) and a high in Tourette's (0.71) and

migraine (0.79). Variation in HRQOL status and utility scores is also observed between diagnoses in the LINC Study sample. Correspondingly, greater utility disparities are seen in the LINC Study sample diagnoses with lower HUI3 utility scores (see Results section for paired *t* test for diagnoses; Table 1). A similar pattern has been observed between the EQ-5D and the SF-6D in people with differing severities of chronic heart failure [53]. These disparities could have implications in interpreting and comparing utility scores for individuals with higher or lower health status within the same sample.

Sociodemographic characteristics influence utility scores [45]. Nevertheless, similar to the findings of Wee et al. [54], no significant associations between sociodemographic characteristics and utility discrepancies were observed in this study [54]. Although overshadowed by the effect of HRQOL status (mean utility), clinical characteristics (Neuro-QoL covariates) played a small but statistically significant role in contributing to the magnitude of discrepancy between the SF-6D and the HUI3 utility scores. This is consistent with studies that find variation in utility scores in populations with comorbid and diverse clinical symptoms, ultimately impacting HRQOL evaluations and utility agreement [44,55].

HRQOL assessment is constrained by the content of an instrument's descriptive health classification system [17,23]. Differences between these descriptive systems are a cause of variation and discrepancy between utility scores and contribute to differences in sensitivity and discrimination, measurement error, and between-instrument agreement [17,19,24,25,45,50,56]. Richardson et al. [17] estimated that differences in descriptive systems account for 51% of differences in utilities observed between the SF-6D and the HUI3. Furthermore, omissions of important health attributes or attribute levels in the descriptive system can lead to variation in the range and severity of potential health states captured and ceiling and floor effects [23]. The health status in neurological populations is well below that of the general population [31,32]; therefore, the HUI3 and its apparent ability/benefit over the SF-6D to characterize HRQOL in lower HRQOL states may appear preferential. This "benefit," however, may not universally apply to all neurological conditions or outcomes of interest, especially those that impact participation in social roles. The HUI3's descriptive system is conceptually limited to "within the skin" health domains and focuses on an individual's body functions, whereas the SF-6D also includes psychosocial health domains of role limitations and social interaction that are domains considered "outside the skin" [17,21,50]. These descriptive system differences are reflected in research that finds that the HUI3 is more sensitive to outcomes of physical and cognitive functioning, whereas the SF-6D is more sensitive to outcomes that impact social functioning [19,24,25,56]. The present study found evidence of discrepancies in utility estimates that were aligned with differences in descriptive system domains. Utility discrepancies did not vary by the loosely shared "within the skin" domains of the SF-6D and the HUI3 of ambulation, dexterity, and cognition, but did vary and by "outside the skin" domains such as fatigue, "ability to participate in activities," and "satisfaction in social roles," attributes reflected in the SF-6D's, but not the HUI3's, health classification system.

This study highlights that both the HRQOL status (mean utility) of a sample and clinical outcomes of interest (condition characteristics and condition severity) should be considered when selecting an instrument or when comparing utility scores. This study compared discrepancies associated with both sociodemographic and clinical characteristics and also adjusted for mean utility in comparing agreement between the SF-6D and the HUI3, which has not been considered in previous studies. A limitation of this study is that the LINC Study sample was not based on a probability sample of the population and therefore was not representative of the population of persons with neurological conditions. Nevertheless, given that discrepancies between utilities were most strongly associated with HRQOL status, and this relationship was consistent across both participant- and condition-specific attributes, we expect our findings to be robust and applicable in other study samples. Further comparative studies in neurological population and clinical samples with mild, moderate, and severe impairment are needed to corroborate the strong association between utility discrepancies and HRQOL status and its implications on perceiving important clinical differences in neurological conditions. Moreover, recent studies are exposing discrepancies and variation in utilities in populations with chronic, multifaceted conditions and subsamples with differing disease severity [45,53,55]. The potential impact these discrepancies may have on interpreting utilities and the physical, cognitive, and psychosocial-emotional aspects of HRQOL warrants additional investigation. The cross-sectional design of the LINC Study precluded evaluating whether discrepancies between the SF-6D and the HUI3 impact the instruments' responsiveness to changes in health status when monitoring disease progression over time and is thus an area for further research [40]. In addition, expanding analyses to include other utility instruments (generic and condition-specific) would further understanding of measurement discrepancies between instruments in persons with neurological conditions.

Conclusions

Differences in the way preference-based HRQOL instruments define, describe, and value health affect the interpretability, comparability, meaningfulness, and agreement across utilities measured [17,23,57]. Utility estimates are among the most important but imprecise data inputs in cost-utility models [7]. Nevertheless, beyond their role in economic valuation, utility instruments are frequently administered as generic measures of HRQOL in which the quality of utility estimates is still critical to the accurate health care assessment, monitoring, and decision-making processes [7,17]. This study adds to the growing body of evidence regarding the methodological differences between utility instruments. Discrepancy between the SF-6D and the HUI3 utility scores in persons with neurological conditions was substantial and varied systematically with the level of HRQOL. Between-instrument utility comparisons should be avoided, especially at the extremities of utility distributions. Depending on the HRQOL status of the sample, the SF-6D and the HUI3 could provide widely contradictory utility predictions confirming that SF-6D and HUI3 are not interchangeable and both instruments could favor or disadvantage individuals depending on their utility.

Acknowledgments

The survey data used in this study were part of the National Population Health Study of Neurological Conditions. We acknowledge the membership of the Neurological Health Charities Canada and the Public Health Agency of Canada for their contribution to the success of this initiative. The opinions expressed in this publication are those of the authors/researchers and do not necessarily reflect the official views of the Public Health Agency of Canada. We thank the many participants in the LINC Study as well as the large research team that made the study possible.

Source of financial support: Funding for the survey data on which this study was based was provided by the Public Health Agency of Canada.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2017.04.008 or, in a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Gaskin J, Gomes J, Darshan S, et al. Burden of neurological conditions in Canada 2016. Available from: . http://dx.doi.org/10.1016/j.neuro.2016.05.001. [Accessed June 1, 2016].
[2] Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2224–60.
[3] Carod-Artal FJ. Determining quality of life in stroke survivors. Expert Rev Pharmacoecon Outcomes Res 2012;2:199–211.
[4] Heffernan C, Jenkinson C. Measuring outcomes for neurological disorders: a review of disease-specific health status instruments for three degenerative neurological conditions. Chronic Illn 2005;2:131–42.
[5] Karlsen H, Tandenberg E, Arsland D, et al. Health related quality of life in Parkinson's disease: a prospective longitudinal study. J Neurol Neurosurg Psychiatry 2000;69:584–9.
[6] Vossius C, Nilsen OB, Larsen JP. Health state values during the first year of drug treatment in early-stage Parkinson's disease: a prospective, population-based, cohort study. Drugs Aging 2009;26:973–80.
[7] Wolowacz SE, Briggs A, Belozeroff V, et al. Estimating health-state utility for economic models in clinical studies: an ISPOR Good Research Practices Task Force report. Value Health 2016;19:704–19.
[8] Kuspinar A, Mayo NE. Do generic utility measures capture what is important to the quality of life of people with multiple sclerosis? Health Qual Life Outcomes 2013;11:71–81.
[9] Mayo NE, Moriello C, Asano M, et al. The extent to which common health-related quality of life indices capture constructs beyond symptoms and function. Qual Life Res 2011;20:621–7.
[10] Cella D, Nowinski C, Peterman A, et al. The neurology quality-of-life measurement initiative. Arch Phys Med Rehabil 2011;92:S28–36.
[11] Cella D, Lai JS, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. Neurology 2012;78:1860–7.
[12] Guyatt H, Feeny H, Patrick L. Measuring health-related quality of life. Ann Intern Med 1993;118:622–9.
[13] Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. Med Care 2002;40:113–28.
[14] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271–92.
[15] Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004;42:851–9.
[16] Brooks R, Rabin R, de Charro F. The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2003.
[17] Richardson J, Iezzi A, Khan MA. Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and "micro-utility" effects. Qual Life Res 2015;24:2045–53.
[18] Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. Med Care 2003;41:791–801.
[19] Pickard AS, Johnson JA, Feeny DH. Responsiveness of generic health-related quality of life measures in stroke. Qual Life Res 2005;14:207–19.
[20] Marra A, Marion A, Guh P, et al. Not all "quality-adjusted life years" are equal. J Clin Epidemiol 2007;60:616–24.
[21] O'Brien BJ, Spath M, Blackhouse G, et al. A view from the bridge: agreement between the SF-6D utility algorithm and the health utilities index. Health Econ 2003;12:975–81.
[22] Fryback G, Palta M, Cherepanov D, et al. Comparison of 5 health-related quality-of-life indexes using item response theory analysis. Med Decis Making 2010;30:5–15.
[23] Whitehurst DG, Bryan S. Another study showing that two preference-based measures of health-related quality of life (EQ-5D and SF-6D) are not interchangeable. But why should we expect them to be? Value Health 2011;14:531–8.
[24] Langfitt T, Vickrey G, McDermott P, et al. Validity and responsiveness of generic preference-based HRQOL instruments in chronic epilepsy. Qual Life Res 2006;15:899–914.
[25] Fisk D, Brown MG, Sketris IS, et al. A comparison of health utility measures for the evaluation of multiple sclerosis treatments. J Neurol Neurosurg Psychiatry 2005;76:58–63.
[26] Versnel J, Packer T, Weeks LE, et al. The everyday experience of living with and managing a neurological condition (the LINC Study): study design. BMC Neurol 2013;13:30.
[27] Horsman J, Furlong W, Feeny D, et al. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual Life Outcomes 2003;1:54.
[28] Pohar SL, Jones CA, Warren S, et al. Health status and health care utilization of multiple sclerosis in Canada. Can J Neurol Sci 2007;34:167–74.
[29] Pohar SL, Allyson Jones C. The burden of Parkinson disease (PD) and concomitant comorbidities. Arch Gerontol Geriatr 2009;49:317–21.
[30] Edwards D, Koehoorn M, Boyd A, et al. Is health-related quality of life improving after stroke? A comparison of health utilities indices among Canadians with stroke between 1996 and 2005. Stroke 2010;41:996–1000.
[31] Statistics Canada. Canadian Community Health Survey (CCHS)—annual component. Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3226. [Accessed April 14, 2015].
[32] Public Health Agency of Canada. Survey of living with a neurological condition in Canada (SLNCC). Available from: http://www.phac-aspc.gc.ca/publicat/cd-mc/mc-ec/section-1-eng.php. [Accessed February 28, 2014].
[33] Kharroubi SA, Brazier JE, Roberts J, et al. Modelling SF-6D health state preference data using a nonparametric Bayesian method. J Health Econ 2007;26:597–612.
[34] McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. J Health Econ 2006;25:418–31.
[35] Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health Qual Life Outcomes 2003;1:4.
[36] Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 2005;14:1523–32.
[37] Kim SK, Kim SH, Jo MW, Lee SI. Estimation of minimally important differences in the EQ-5D and SF-6D indices and their utility in stroke. Health Qual Life Outcomes 2015;13:32.
[38] National Institute of Neurological Disorders and Stroke. User manual for the quality of life in neurological disorders (Neuro-QOL) measures, version 1.0. Available from: http://www.neuroqol.org/Resources/Resources%20documents/Neuro-QOL-UserManual-2013.pdf. [Accessed April 1, 2013].
[39] Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 1997;50:869–79.
[40] Seymour J, McNamee P, Scott A, et al. Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. Health Econ 2010;19:683–96.
[41] Shrout E, Fleiss L. Interclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.
[42] Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med 1990;20:337–40.
[43] Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60.
[44] Jaeschke R, Guyatt G. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–15.
[45] Kontodimopoulos N, Pappa E, Papadopoulos AA, et al. Comparing SF-6D and EQ-5D utilities across groups differing in health status. Qual Life Res 2009;18:87–97.
[46] McDonough CM, Grove MR, Tosteson TD, et al. Comparison of EQ-5D, HUI, and SF-36-derived societal health state values among Spine Patient Outcomes Research Trial (SPORT) participants. Qual Life Res 2005;14:1321–32.
[47] Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Qual Life Res 2008;17:485–95.
[48] Kaplan RM, Tally S, Hays RD, et al. Five preference-based indexes in cataract and heart failure patients were not equally responsive to change. J Clin Epidemiol 2011;64:497–506.
[49] Fryback DG, Dunham NC, Palta M, et al. US norms for six generic health-related quality-of-life indexes from the National Health Measurement study. Med Care 2007;45:1162–70.
[50] Davison SN, Jhangri GS, Feeny DH. Comparing the Health Utilities Index Mark 3 (HUI3) with the Short Form-36 preference-based SF-6D in chronic kidney disease. Value Health 2009;12:340–5.
[51] Marra CA, Woolcott JC, Kopec JA, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQOL and the HAQ) in rheumatoid arthritis. Soc Sci Med 2005;60:1571–82.

[52] Feeny D, Wu L, Eng K. Comparing short form 6D, standard gamble, and Health Utilities Index Mark 2 and Mark 3 utility scores: results from total hip arthroplasty patients. Qual Life Res 2004;13:1659–70.
[53] Kontodimopoulos N, Argiriou M, Theakos N, et al. The impact of disease severity on EQ-5D and SF-6D utility discrepancies in chronic heart failure. Eur J Health Econ 2011;12:383–91.
[54] Wee H-L, Machin D, Loke W-C, et al. Assessing differences in utility scores: a comparison of four widely used preference-based instruments. Value Health 2007;10:256–65.
[55] Kontodimopoulos N, Pappa E, Chadjiapostolou Z, et al. Comparing the sensitivity of EQ-5D, SF-6D and 15D utilities to the specific effect of diabetic complications. Eur J Health Econ 2012;13:111–20.
[56] Hatoum HT, Brazier JE, Akhras KS. Comparison of the HUI3 with the SF-36 preference based SF-6D in a clinical trial setting. Value Health 2004;7:602–9.
[57] Lenert L, Kaplan RM. Validity and interpretation of preference-based measures of health-related quality of life. Med Care 2000;38(Suppl): II138–50.