Research: Health Economics

Developing preference-based measures for diabetes: DHP-3D and DHP-5D

B. Mulhern[1,2] (iD), A. Labeit[2], D. Rowen[2], E. Knowles[2], K. Meadows[3], J. Elliott[4] (iD) and J. Brazier[2]

[1]University of Technology Sydney, Centre for Health Economics Research and Evaluation, Sydney, Australia, [2]School of Health and Related Research, University of Sheffield, [3]DHP Research and Consultancy and [4]Academic Unit of Diabetes, Endocrinology and Metabolism, University of Sheffield, Sheffield, UK

Abstract

Aims The aim of this study was to develop two diabetes-specific preference-based measures [the Diabetes Health Profile–3 Dimension (DHP-3D) and the Diabetes Health Profile–5 Dimension (DHP-5D)] for use in the calculation of Quality Adjusted Life Years, a key outcome in economic evaluation. These measures were based on the non-preference-based instrument the Diabetes Health Profile.

Methods For DHP-3D, psychometric and Rasch analyses were used to develop a health state classification system based on the Diabetes Health Profile–18 (DHP-18). The DHP-5D added two dimensions to the DHP-3D to extend the range of impacts measured. Each classification system was valued by 150 general public respondents in the United Kingdom using Time Trade Off (TTO). Multivariate regression was used to estimate utility value sets. The matched dimensions across each measure were compared using z-score tests.

Results The DHP-3D included three dimensions defined as mood, eating and social limitations, and the DHP-5D added dimensions defined as hypoglycaemic attacks and vitality. For both, the random effects generalized least squares regression model produced consistent value sets, with the DHP-3D and DHP-5D ranging from 0.983 (best state) to 0.717 (worst state), and 0.979 to 0.618 respectively. The addition of the two extra dimensions leads to significant differences for the more severe levels of each matched dimension.

Conclusions We have developed two diabetes-specific preference-based measures that, subject to psychometric assessment, can be used to provide condition-specific utility values to complement generic utilities from more widely validated measures such as the EuroQol-5 Dimension.

Diabet. Med. 000, 000–000 (2017)

Introduction

For the economic evaluation of health interventions, the Quality Adjusted Life Year (QALY), which combines length and quality of life into a single figure, is a commonly used outcome. The quality aspect (or utility value) is anchored on a scale of 1 (full health) to 0 (dead) and can be derived from a preference-based measure of health. Preference-based measures include two elements, a health state classification system and a utility value set. The utility value set is generally based on general population preferences for health states described by the measure and is elicited using a technique such as Time Trade Off (TTO) [1] to produce values to input into QALYs.

Generic preference-based measures are widely used in the estimation of QALYs because they are developed for use across conditions. The EuroQol-5 Dimension (EQ-5D) [2,3] is recommended by the United Kingdom (UK) National Institute for Health and Care Excellence (NICE) [4], although it is also noted that the EQ-5D may not be relevant in all circumstances. However, some generic measures may be insensitive in many patient groups because they do not cover important consequences of specific conditions [5,6]. In diabetes, there is some evidence for the psychometric validity of a range of generic measures, and some sensitivity to complications [7,8]. However, condition-specific measures may provide more sensitive assessments of a condition. In diabetes, one such measure is the Diabetes Health Profile (DHP-1/DHP-18) [9,10] which is used to provide a detailed assessment of emotional and behavioural health [11].

Condition-specific measures cannot be used to directly estimate QALYs as they are not preference based. It is not possible to derive value sets for these measures given that they describe many millions of health states that would be

Correspondence to: Brendan Mulhern.
E-mail: brendan.mulhern@chere.uts.edu.au

too complex to value. However, preference-based measures can be developed from existing measures to provide condition-specific utilities [12]. This is done by developing a shortened health state classification system from the longer measure, and valuing the classification system using a preference elicitation technique with a general population sample. The method has been applied widely in different conditions [13–18]. General population values are used in the development of preference-based measures, because in a publically funded healthcare system such as the UK, it is argued that it is the values of the tax payer funding the system that matters, and general population values that are recommended for use by NICE.

The aim of this study was to develop two diabetes-specific preference-based measures. This includes the following objectives:

• to develop a health state classification system from the DHP-18 (DHP-3D);

• to extend the scope of the DHP-3D and develop a second classification system that captures additional dimensions of importance (DHP-5D);

• to generate utility value sets for both measures using TTO.

The advantage of developing two preference-based measures is that utility values can be generated where only DHP-18 data is available (DHP-3D), and for other studies additional items can be included to estimate utilities reflecting a wider range of dimensions (DHP-5D).

Methods

Measures – DHP-18

The DHP-18 [8] was developed from the longer DHP-1 and measures the psychosocial functioning across three domains, psychological distress, barriers to activity and disinhibited eating. The DHP-18 has been used in clinical trials, health surveys and in the UK Department of Health long-term condition measurement study [19–23].

Data

The development of the DHP-3D was undertaken using data from a cross-sectional study of people with Type 2 diabetes recruited at a hospital in South Yorkshire, UK ($n = 237$;

response rate 80.6%) categorized into three treatment groups: insulin ($n = 122$); non-injectors/oral medication ($n = 37$); and Exenatide, an injectable medicine ($n = 78$). The sample is older than the general population (proportion aged 65+ years: 44.1% vs.19.3%), and contains a larger proportion of men (61.8% vs. 48.7%) and retired people (55.5% vs.13.5%).

Preference-based measure development process

To develop the preference-based measures, we followed a five-step process adapted from existing guidance [10].

Step 1 – Dimensionality assessment
DHP-3D: Exploratory factor analysis detects structure in the relationship between items and was used to examine DHP-18 dimensionality. Factor loadings indicate the level of correlation between the item and the factor. The number of dimensions was indicated by the number of factors with eigenvalues ≥ 1 (meaning that the factor accounts for more variability than a single item) [24]. Items loading < 0.35 on any factor were excluded.

DHP-5D: To develop the DHP-5D, a diabetes clinician and research nurse were consulted about other relevant dimensions. As part of this process, the authors and clinicians discussed the dimensions included in the DHP-1, the generic EQ-5D and Short Form-36 (SF-36), and also other potentially important dimensions.

Step 2 – Item selection
DHP-3D: To develop the dimension descriptions we used classical psychometric and Rasch analysis [25] to first, exclude and second, select item(s) to represent the underlying dimension. Classical psychometrics included floor (proportion in worst category) and ceiling (proportion in the best category) effects and missing data. Items with missing data < 5% and floor/ceiling effects < 80% were acceptable.

Rasch analysis is an item response theory technique that converts categorical item responses into a continuous latent (logit) scale. The probability of a response to each level of each item is used to assess the severity of the item against the underlying logit scale. To exclude items, we assessed item response ordering, differential item functioning across age and gender subgroups and fit to the dimension model. To select the best item, the two key criteria were the spread of responses at the average item difficulty (at logit 0) and item severity range coverage.

DHP-5D: To develop the extra dimensions, relevant items were selected from an appropriate measure.

Step 3 – Validation
The classification systems developed were validated by a diabetes clinician and research nurse who were shown the results of steps 1 and 2, and the original measures, and asked their opinions on the item selected to represent each dimension in comparison with the others available.

Step 4 – Valuation study
The valuation protocol used TTO with a sample of 300 UK general population respondents valuing nine health states each.

State selection and sample allocation: In the sections below, health states are described in terms of the levels of each dimension represented. For example, DHP-3D state 111 represents a state including the best level, level 1, of each dimension. This will range to 444 which is the worst possible DHP-3D state. For the DHP-5D, 33 unique states were selected using an orthogonal array, and allocated to four blocks of nine states (eight unique states plus the worst state). For the DHP-3D, the same states excluding the two additional dimensions were valued. This meant that 32 states were valued because state 111 was duplicated so appeared in two blocks. Each preference-based measure was valued by 150 respondents.

Interview procedure and the composite TTO task: Face-to-face interviews were carried out by a research agency. First, respondents completed demographic questions, and the classification system to be valued. They were informed that the states relate to diabetes and ranked those included in the block in order of preference.

The TTO task used the 'standard' method for states better than dead [2] and Lead Time–Time Trade Off [26] for states worse than dead. For states better than dead, respondents chose between 10 years in a health state (life A) and between 0 and 10 (*x*) years in full health (life B). The amount of time in full health was varied following a set iterative process until indifference was reached, with the value calculated as *x*/10. For states worse than dead (where 0 years in life B was preferable to 10 years in life A), a 'lead time' of 10 years in full health was added to each option. Respondents then chose between 10 years in full health followed by 10 years in the health state (life A), or between 0 and 20 (*x*) years in full health (life B).

Recruitment: Recruitment was carried out across the UK. First, respondents were recruited via an existing database of people interested in completing research studies and were provided with the study information and asked to participate. Second, interviewers approached new respondents in a neutral location such as a coffee shop or library, provided them with the study information and asked them to take part. The interviews took place either in the respondents' home or in a public place and participants received a £25 incentive. This process was approved by the School of Health and Related Research Ethics Committee at The University of Sheffield.

Exclusion criteria and data analysis: Respondents who valued every health state the same (but not at 1) were excluded. This is because it is expected that respondents should value states differently if they are willing to trade some time, but valuing every state at one indicates not giving up any time which could be a genuine preference.

Respondents who valued the worst state (444 or 44445) the highest were also excluded as this indicates lack of understanding.

Step 5 – Modelling
Individual and mean-level multivariate regression was used to estimate utility values for all states. The models estimated preference weights for each dimension severity level (with no problems the baseline). The classification systems are valued against 'full health', which is given a value of 1 with instrument-specific full health equivalent to the constant term.

Health state utilities were estimated using the following function:

$$U_{ij} = g(\beta' x_i) + \varepsilon_{it} \quad (1)$$

where U_{ij} gives the TTO disvalue which is calculated as 1 minus the TTO value, $i(1, 2…n)$ is the number of health states, $j(1,2,…m)$ is the number of respondents, g specifies the functional form and ε_{it} is the error term. X_i is a vector of binary dummy variables for each level of each dimension where the best level represents the baseline. The models included ordinary least squares regression, random effects generalized least squares (RE GLS) to account for repeated observations, and RE Tobit to account for the bounded nature of the data (where $-1 \leq U_{ij} \leq 1$). Ordinary Least Square models comprising one mean value per health state were also estimated.

Model performance criteria included the number of inconsistencies and significant coefficients within each dimension, root mean squared error and mean absolute error at the state level, and the number of states with absolute error > 5% and > 10%. Plots of actual and predicted values were examined.

To compare the matched dimensions across the DHP-3D and DHP-5D we compared the difference across each level coefficient using the standardized *z*-score test

$$z = (\beta 1 - \beta 2)/\sqrt{(se\beta 1^2 + se\beta 2^2)} \quad (2)$$

where $\beta 1$ and $\beta 2$ are the coefficients for each level of each dimension. A *z*-score of 1.96 or more indicates a significant difference at the 5% level.

Results

Step 1 – Dimensionality

DHP-3D
A three-dimension model was suggested (Table 1). Dimension 1 (Mood) includes six items from the psychological distress factor (explaining 34.9% of the variance). Dimension 2 (Eating) includes five items from the disinhibited eating factor plus one barriers to activity item (explaining 12.6%). Dimension 3 (Social limitations) includes five

Table 1 Dimension structure, psychometric and Rasch analysis of the DHP-18

Dimension	DHP-18 dimension	Loading	Psychometric analysis			Rasch analysis					
			% at ceiling	% at floor	% missing data	Disordered	Differential item functioning	Fit residual	P value	Spread at logit 0	Item range
Psychological distress											
Q6: Lose temper if people keep on at you	PD	0.68	53.2	8.0	2.1	Yes (Sometimes/usually)					
Q8: More arguments due to diabetes	PD	0.78	62.0	4.2	1.3		Gender	−0.42	0.22	0.37 to 0.84	−1.65 to 0.53
Q15: Get depressed	PD	0.59	38.0	7.6	0						
Q16: Lose temper or shout	PD	0.99	46.4	6.3	0	Yes (Sometimes/often)					
Q17: Get touchy or moody	PD	0.72	46.8	4.6	0			−0.45	0.72	0.14 to 0.85	−1.72 to 1.85
Q18: *Lose temper over small things*	PD	0.83	32.5	7.2	0			−0.20	0.68	0.04 to 0.70	−0.87 to 3.16
Eating											
Q1: Food control life	BA	0.60	24.9	7.2	2.1			0.66	0.41	0.20 to 0.85	−1.71 to 1.37
Q5: Eat to cheer self up	DE	0.68	31.6	6.3	1.7			−0.68	0.31	0.30 to 0.85	−1.71 to 0.85
Q7: Hard saying no to foods	DE	0.73	17.6	14.2	1.7	Yes (sometimes/usually)					
Q9: *Easy to stop eating*	DE	0.64	17.3	3.8	0.4		Gender	0.03	0.91	0.11 to 0.93	−2.64 to 2.04
Q10: Eat more when bored/fed up	DE	0.67	10.5	16.9	0.4						
Q12: Wish not so many nice things to eat	DE	0.49	29.1	22.8	0	Yes (a little/a lot)					
Social limitations											
Q2: Difficult staying out late	BA	0.63	52.3	5.5	2.1	Yes (usually/sometimes)					
Q3: *Days tied to meal times*	BA	0.73	34.6	3.8	2.5	Yes (usually/sometimes)		−0.76	0.11	0.11 to 0.88	−2.01 to 2.10
Q4: Avoid going out if sugars low	BA	0.71	49.8	11.8	2.5						
Q11: Worry about getting colds/flu	BA	0.39	48.9	7.2	0.4	Yes (a little/a lot)					
Q14: Get edgy	BA	0.53	55.3	6.8	2.1	Yes (a little/a lot)					
Excluded for non-loading											
Q13: Frightening going into busy shops	BA										

PD, psychological distress; BA, barriers to activity; DE, disinhibited eating. Items in italics were chosen for the descriptive system.

DHP-3D

Mood
You <u>never</u> find yourself losing your temper over small things
You <u>sometimes</u> find yourself losing your temper over small things
You <u>often</u> find yourself losing your temper over small things
You <u>very often</u> find yourself losing your temper over small things

Social limitations
Your days are <u>never</u> tied to meal times
Your days are <u>sometimes</u> tied to meal times
Your days are <u>usually</u> tied to meal times
Your days are <u>always</u> tied to meal times

Eating
When you start eating you find it <u>very easy</u> to stop
When you start eating you find it <u>quite easy</u> to stop
When you start eating you find it <u>not very easy</u> to stop
When you start eating you find it <u>not at all easy</u> to stop

DHP-5D

Mood
You <u>never</u> find yourself losing your temper over small things
You <u>sometimes</u> find yourself losing your temper over small things
You <u>often</u> find yourself losing your temper over small things
You <u>very often</u> find yourself losing your temper over small things

Social limitations
Your days are <u>never</u> tied to meal times
Your days are <u>sometimes</u> tied to meal times
Your days are <u>usually</u> tied to meal times
Your days are <u>always</u> tied to meal times

Eating
When you start eating you find it <u>very easy</u> to stop
When you start eating you find it <u>quite easy</u> to stop
When you start eating you find it <u>not very easy</u> to stop
When you start eating you find it <u>not at all easy</u> to stop

Hypoglycemic attacks
You <u>never</u> worry about doing too much and going hypo
You <u>sometimes</u> worry about doing too much and going hypo
You <u>usually</u> worry about doing too much and going hypo
You <u>always</u> worry about doing too much and going hypo

Vitality
You are tired <u>none of the time</u>
You are tired <u>a little of the time</u>
You are tired <u>some of the time</u>
You are tired <u>most of the time</u>
You are tired <u>all of the time</u>

FIGURE 1 DHP-3D and DHP-5D health state classification systems.

barriers to activity items (explaining 9.6%). One of the barriers to activity items (frightened to go into busy shops) cross loaded so it was excluded.

HCS2
Expert input based on the DHP-1 and SF-36 suggested the use of two extra dimensions that are key features of diabetes: 'Hypoglycaemic attacks' and 'Vitality'.

Step 2 – Dimension development

DHP-3D
The psychometric and Rasch analyses are shown in Table 1. Missing data rates are low (0–2.5%), and ceiling (17.3%–62.0%) and floor effects (3.8%–22.8%) were acceptable, and therefore do not clearly help to differentiate items.

Table 2 Valuation study demographics

Demographic	DHP-3D n (%)	DHP-5D n (%)	Significance	UK %
Completion	150	150		
Male	61 (40.7)	65 (43.3)	0.64	48.7
Age, years				
Mean (SD)	34.7 (14.2)	37.0 (14.8)	0.18	NA
Range	18–82	18–74		NA
Distribution			0.76	
18–40 years	100 (66.7)	94 (62.7)		41.6
41-65 years	45 (30.0)	51 (34.0)		39.1
Over 65 years	5 (3.3)	5 (3.3)		19.3
Ethnicity			0.37	
White British	116 (77.3)	125 (83.3)		85.7
White (other)	19 (12.7)	9 (6.0)		6.5
Mixed race	3 (2.0)	4 (2.7)		1.2
Asian	6 (4.0)	5 (3.3)		4.4
Black	4 (2.7)	3 (2.0)		2.0
Marital status			0.20	
Married/partner	75 (50.0)	86 (57.3)		NA
Single	64 (42.7)	59 (39.3)		NA
Separated/divorced/widowed	11 (7.3)	5 (3.3)		NA
Employment			0.45	
Employed	94 (62.7)	100 (66.7)		60.9
Retired	7 (4.7)	10 (6.6)		13.5
Student	44 (29.3)	33 (22.2)		7.3
Other	5 (3.3)	7 (4.7)		18.3
Education post minimum	140 (93.3)	136 (90.7)	0.40	NA
Has degree level education	92 (61.3)	97 (64.7)	0.55	NA
Householder status			0.25	
Own home	92 (61.3)	93 (62.0)		68.7
Rent	53 (35.3)	56 (37.3)		31.3
Experience of serious illness				
Yourself	27 (18.0)	30 (20.0)	0.91	NA
Family	113 (75.3)	110 (73.3)	0.54	NA
Caring for others	52 (34.7)	48 (32.0)	0.78	NA
Diabetes				
Yourself	2 (1.3)	8 (5.3)	0.05	NA
In family	68 (45.3)	57 (38.0)	0.43	NA
In caring for others	26 (17.3)	26 (17.3)	1.00	NA
Has long term health condition	48 (32.0)	55 (36.7)	0.40	NA
Time off work/other ill last 4 weeks	16 (10.7)	28 (18.7)	0.05	NA

NA, not available.
Values are given as number (%) unless indicated otherwise.

Psychological distress: Items 6 and 16 were disordered between the middle response categories (sometimes/usually and sometimes/often respectively). Item 15 displayed differential item functioning by gender. The remaining items (8, 17 and 18) had goodness of fit within the required parameters. Item 18 ('Do you find yourself losing your temper over small things?') covered the more severe end of the scale and was selected.

Eating: Items 7 and 12 displayed disordering between the middle response categories (sometimes/usually and a little/a lot respectively) and item 10 displayed differential item functioning by gender. The three remaining items (1, 5 and 9) displayed goodness of fit. Item 9 ('When you start eating how easy do you find it to stop?') had the largest spread at logit 0 and the largest severity range so was selected.

Social limitations: The middle response categories of items 2 and 4 (sometimes/usually) and items 11 and 14 (a little/a

lot) were disordered. Item 3 ('Does having diabetes mean your days are tied to meal times?') remained for selection and covered a large severity range so was selected.

DHP-5D
For the DHP-5D the same items as the DHP-3D were used for the matched dimensions along with the items for the extra two dimensions described below.

Hypoglycaemic attack: The DHP-1 includes four items representing this dimension: 'having nagging fear of hypos'; 'avoiding going too far in case of hypos'; 'worry about doing too much and going hypo'; and 'worry about going into a diabetic coma'. The item relating to 'worry about doing too much and going hypo' was selected as it describes the dimension in more general terms.

Vitality: The SF-36 vitality dimension includes four items measuring energy, tiredness, being worn out and being full of
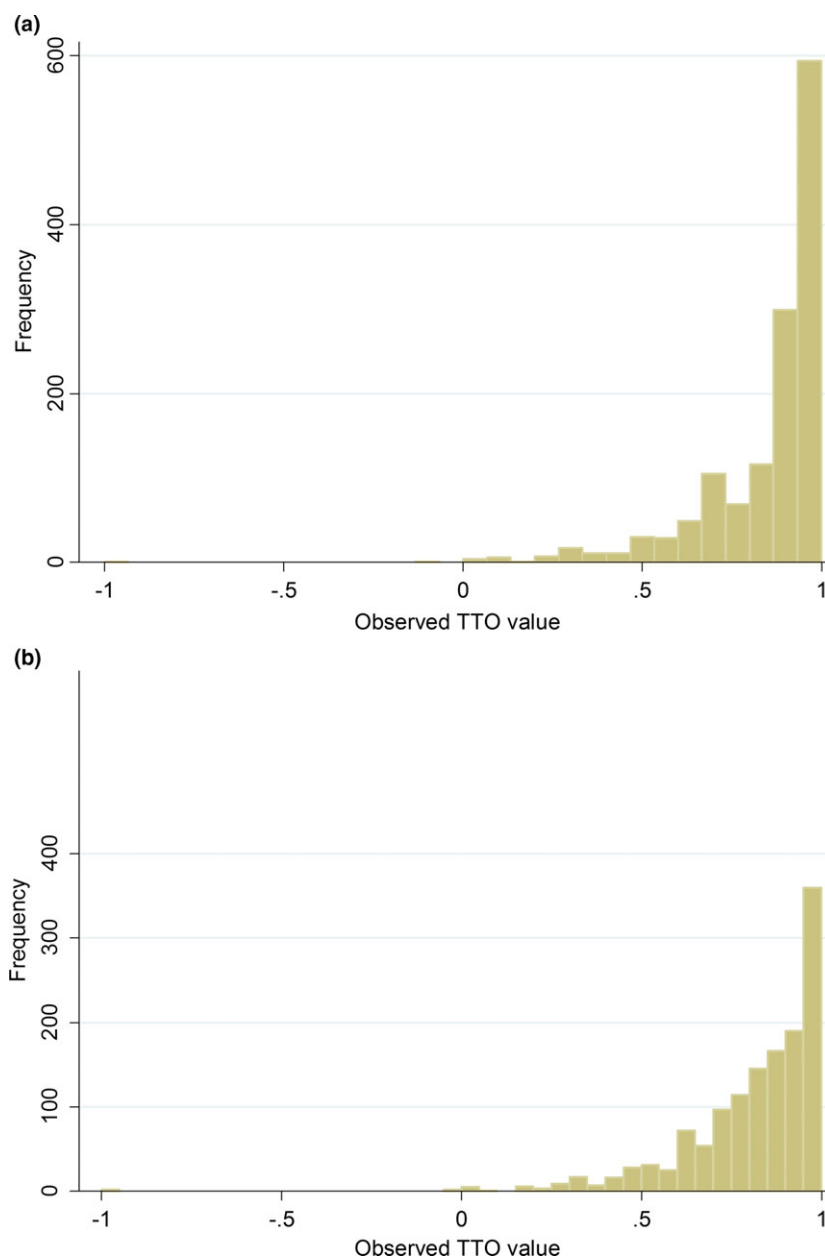
FIGURE 2 Histograms of observed DHP-3D (panel a) and DHP-5D (panel b) values.

life. The positively worded items were excluded as these may cause issues with the overall perception of the health states by respondents. Of the two negatively worded items, the more general item asking about frequency of feeling tired was selected.

Step 3 – Validation

The diabetes clinician and research nurse validated the items selected for both the DHP-3D and DHP-5D. The items selected were therefore rewritten to generate the health state classification systems for valuation (Fig. 1).

Step 4 – Valuation study

Sample

Overall, 300 interviews were fully completed (150, DHP-3D; 150, DHP-5D) with no significant demographic differences between the respondent samples (Table 2). The sample differed to the UK population, with a higher proportion of women, younger people and students, and fewer retired people. A substantial proportion of the sample had experience of diabetes in themselves, their family or in caring for others.

Table 3 Regression models estimating preference weights for the DHP-3D and DHP-5D

	DHP-3D [Coefficient (SE)]				
	(1) OLS	(2) RE-GLS	(3) RE-Tobit	(4) RE-Tobit (marginal effects)	(5) Mean model
Mood level 2	–0.030*** (0.008)	–0.026*** (0.009)	–0.038*** (0.010)	–0.025*** (0.007)	–0.029* (0.015)
Mood level 3	–0.098*** (0.016)	–0.085*** (0.010)	–0.100*** (0.011)	–0.072*** (0.009)	–0.098*** (0.015)
Mood level 4	–0.143*** (0.018)	–0.127*** (0.010)	–0.144*** (0.011)	–0.110*** (0.009)	–0.139*** (0.015)
Eating level 2	–0.004 (0.012)	0.000 (0.010)	–0.001 (0.011)	–0.001 (0.008)	–0.004 (0.015)
Eating level 3	–0.066*** (0.012)	–0.060*** (0.010)	–0.069*** (0.010)	–0.053*** (0.008)	–0.065*** (0.015)
Eating level 4	–0.058*** (0.010)	–0.074*** (0.010)	–0.083*** (0.010)	–0.065*** (0.008)	–0.053*** (0.015)
Social limitations level 2	0.011 (0.012)	–0.022** (0.010)	–0.026** (0.011)	–0.019** (0.008)	0.011 (0.015)
Social limitations level 3	–0.003 (0.020)	–0.044*** (0.012)	–0.051*** (0.013)	–0.039*** (0.010)	–0.003 (0.015)
Social limitations level 4	–0.044*** (0.010)	–0.065*** (0.009)	–0.071*** (0.010)	–0.055*** (0.008)	–0.039** (0.015)
Hypoglycaemic attacks level 2					
Hypoglycaemic attacks level 3					
Hypoglycaemic attacks level 4					
Vitality level 2					
Vitality level 3					
Vitality level 4					
Vitality level 5					
Constant	0.966*** (0.010)	0.983*** (0.015)	1.015*** (0.016)		0.964*** (0.017)
Observations	1278	1278	1278		1278
No. of respondents	142	142	142		32
R^2	0.165				0.870
No. of inconsistencies	2	0	0		2
No. of significant variables	6	8	8		6
Root mean square error	0.178	0.112	.		0.302
Mean absolute error	0.117	0.118	0.116		0.116
% states with AE>0.05	66.5	68.9	65.0		66.0
% states with AE>0.10	43.8	43.7	42.0		43.4
Ljung-Box test	1.617	6.343	7.274		1.715

GLS, Generalized Least Square OLS, ordinary least square; RE, random effects.
Significant at: *10%, **5%, ***1%.

Descriptive statistics and exclusions

Eight respondents were excluded from the DHP-3D (six valued all states the same, and two valued the worst state highest) and three (who valued all states the same) from the DHP-5D. DHP-3D mean values ranged between 0.965 (111) and 0.708 (444) and DHP-5D mean values were from 0.979 (11111) to 0.618 (44445). The mean value generally decreases as the severity of the state (indicated by the sum score) increases. The online supplement includes all descriptive statistics for all the observed TTO values for each instrument. Figure 2 displays the value distributions, with very few states valued negatively (or worse than dead).

Step 5 – Modelling

DHP-3D
Model 2 (RE GLS) was preferred because it includes a higher number of significant parameters and has good predictive ability, with the lowest root mean squared error and percentage of absolute errors (Table 3). The constant is equivalent to instrument-specific full health, so to calculate the utility value the decrement for each level is subtracted from this value. For example, state 333 would be calculated as [0.983 – 0.100 (L3 mood) – 0.069 (L3 eating) – 0.051 (L3 social) = 0.763]. The predicted values vary from 0.983 (111) to 0.717 (444). Mood has the largest overall decrement followed by eating and social limitations. There is no systematic difference in the observed and predicted mean values by state severity (Fig. 3).

DHP-5D
The RE GLS model (model 7) found that the eating and social limitations dimensions were disordered between levels 3/4 and 2/3 respectively (Table 3). These levels were combined to produce the ordered RE GLS model (model 8). The predicted values varied from 0.979 for state 11111 to 0.618 for the state

DHP-5D

	(6) OLS	(7) RE-GLS	(8) RE-GLS consistent model	(9) RE-Tobit	(10) RE-Tobit marginal effects	(11) Mean model	z-score 2 vs. 8
	−0.021** (0.008)	−0.016 (0.010)	−0.015 (0.010)	−0.019* (0.010)	−0.016* (0.009)	−0.021 (0.012)	−1.22
	−0.026** (0.013)	−0.027** (0.011)	−0.027** (0.011)	−0.032*** (0.012)	−0.027*** (0.010)	−0.026* (0.012)	−3.90**
	−0.060*** (0.016)	−0.051*** (0.012)	−0.051*** (0.011)	−0.054*** (0.012)	−0.047*** (0.010)	−0.060*** (0.012)	−5.11**
	−0.042*** (0.012)	−0.026** (0.012)	−0.026** (0.012)	−0.029** (0.012)	−0.024** (0.010)	−0.042*** (0.012)	1.66
	−0.059*** (0.012)	−0.045*** (0.011)	−0.043*** (0.010)	−0.050*** (0.011)	−0.043*** (0.010)	−0.059*** (0.012)	2.62**
	−0.043*** (0.010)	−0.041*** (0.011)	−0.043*** (0.010)	−0.045*** (0.011)	−0.039*** (0.009)	−0.043*** (0.012)	−2.19**
	−0.005 (0.011)	−0.014 (0.011)	−0.012 (0.010)	−0.018 (0.011)	−0.015 (0.010)	−0.005 (0.012)	−0.07
	−0.001 (0.016)	−0.009 (0.013)	−0.012 (0.010)	−0.014 (0.013)	−0.012 (0.011)	−0.001 (0.012)	−2.05**
	−0.026** (0.010)	−0.029*** (0.010)	−0.029*** (0.010)	−0.034*** (0.010)	−0.029*** (0.009)	−0.026** (0.012)	−2.68**
	−0.025** (0.011)	−0.019* (0.010)	−0.019* (0.010)	−0.022** (0.010)	−0.018** (0.008)	−0.025* (0.012)	
	−0.036*** (0.013)	−0.035*** (0.011)	−0.036*** (0.011)	−0.041*** (0.012)	−0.035*** (0.010)	−0.036*** (0.012)	
	−0.067*** (0.009)	−0.065*** (0.010)	−0.065*** (0.010)	−0.070*** (0.010)	−0.061*** (0.008)	−0.067*** (0.012)	
	−0.016* (0.008)	−0.028*** (0.010)	−0.028*** (0.010)	−0.033*** (0.011)	−0.026*** (0.008)	−0.016 (0.012)	
	−0.056*** (0.011)	−0.044*** (0.011)	−0.044*** (0.010)	−0.049*** (0.011)	−0.039*** (0.009)	−0.056*** (0.012)	
	−0.115*** (0.014)	−0.115*** (0.013)	−0.115*** (0.012)	−0.122*** (0.013)	−0.103*** (0.011)	−0.115*** (0.015)	
	−0.173*** (0.018)	−0.173*** (0.011)	−0.173*** (0.011)	−0.179*** (0.011)	−0.157*** (0.010)	−0.173*** (0.014)	
	0.988*** (0.008)	0.979*** (0.018)	0.979*** (0.018)	0.999*** (0.019)	0.121*** (0.003)	0.987*** (0.017)	
	1323	1323	1323	1323			
	147	147	1347	147		33	
	0.236					0.955	
	2	2	0	2		2	
	14	13	12	14		12	
	0.171	0.120	0.119			0.025	
	0.118	0.119	0.119	0.118		0.118	
	71.6	72.8	72.8	72.5		71.6	
	46.1	45.9	45.8	45.2		46.1	
	3.686	4.443	4.681	4.258		3.580	

44445. Vitality had the largest decrement followed by hypoglycaemic attacks. Again, there was no systematic difference in the observed and predicted values (Fig. 3).

The z-score comparison of the RE GLS models (model 2 for DHP-3D and model 8 for the DHP-5D) demonstrates that the addition of the two extra dimensions leads to significant differences for levels 3 and 4 of each matched dimension. For most levels (except eating levels 2 and 3), the addition of the extra DHP-5D dimensions reduces the magnitude of the coefficients.

Discussion

The measurement of condition-specific utilities can provide additional information for use in estimating QALYs. In this study, we developed two diabetes-specific preference-based measures. The DHP-3D can be used to estimate QALYs for the assessment of diabetes-specific interventions where the DHP-18 or DHP-1 is used. The DHP-5D can be used to estimate QALYs when the DHP-1 plus the extra SF-36 item are included.

The instruments developed in this study are not the first diabetes-specific preference-based measures, with the Diabetes Utility Index [27,28] derived from the Audit of Diabetes Dependent Quality of Life. The Diabetes Utility Index includes five dimensions (Physical Ability/Energy, Relationships, Mood/Feelings, Enjoyment of Diet, and Satisfaction with Management of Diabetes). There is limited overlap in the dimensionality of the Diabetes Utility Index and DHP-3D/DHP-5D (which is linked to the scope of the parent measures) and the valuation method (which was Standard Gamble with patients in the USA). This study used TTO with the UK general population for comparability with the valuation of EQ-5D, and therefore we have produced the first of diabetes-specific preference-based measures with UK value sets.
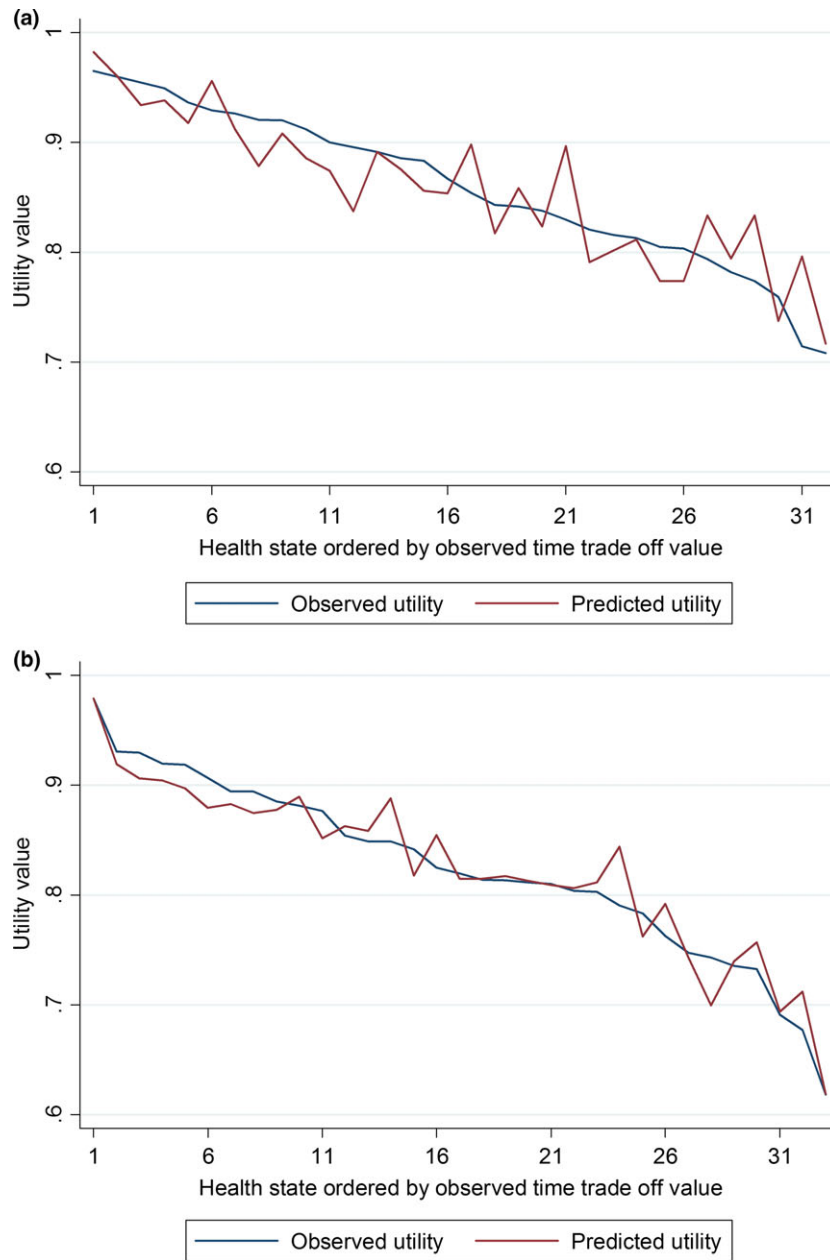
FIGURE 3 Observed vs predicted DHP-3D (panel a) and DHP-5D (panel b) values.

The DHP-3D and DHP-5D cover a small utility scale range in comparison with the EQ-5D [1]. This reflects the mildness of the descriptive system which measures the day-to-day impact of diabetes that may not be perceived as severe in comparison with losing years of life in the TTO exercise. The more severe impact of the longer-term complications of diabetes is not explicitly measured.

Developing two classifications mean that utility values can be generated where only DHP-18 data are available, and for studies collecting new data additional items can be used to generate utility values that are able to reflect a wider range of dimensions. The number of health states described by each measure is substantially different (64 vs. 1280). This impacts on the overall utility range and has implications for the sensitivity of the measures to change over time, with the DHP-5D potentially having increased sensitivity. However, the change in utility value between DHP-5D health states may be smaller than the DHP-3D given that 1280 states cover a utility range of 0.361 in comparison with 64 states covering a range of 0.220. It is also worth noting the potential loss of information in reducing the longer DHP-18 to a shorter classification system. In generating the short form, it is the aim of the analysis to retain as much information as possible to represent the dimensions included

in the longer measure. Therefore, factor, psychometric and Rasch analysis was used to understand the most representative dimension structure, and select the most valid item(s) to represent each domain.

The additional dimensions of the DHP-5D also change the magnitude of the coefficients of the DHP-3D dimensions, with significant differences found for the more severe levels of each. This demonstrates the importance of vitality and hypoglycaemic attacks, and the impact that these dimensions have on the overall importance of the matched dimensions. The difference is mainly driven by the inclusion of vitality, a more generic dimension which has the largest decrement.

Across the two measures there is also a difference in the percentage of the total states included in the valuation study, with (50% of DHP-3D states and 2.5% of DHP-5D). This was done given the widespread use of the DHP-18 in existing studies [19–23] and the practicalities of using the DHP-5D which may require the inclusion of other measures, or the addition of extra items. However, both models produce ordered utility scales, so this may not be a concern.

This study has limitations. First, the development dataset did not include sufficient sample to allow for validation of the classification systems as has been done for other condition-specific preference-based measures [14,16]. However, we did use expert input for validation as recently done by Mukuria et al. [15], and this provides a strong basis for the dimensions and items included. Second, the sample who valued the measures is also not directly representative of the UK general population with more young people taking part, and it may be possible that older people would value diabetes-specific health states differently. However, a many of the sample had experience of diabetes and this could help inform preferences. Third, it is possible that other dimensions that are important in diabetes were not included in the DHP-5D, for example weight gain. However, no data on useful pre-existing items to represent this dimension were found so weight gain was not included.

The aim of preference-based measures is to apply general population values to the self-reported health of people with diabetes to facilitate decision making. It is unclear whether diabetes patients have the same values for the health states as the general population, and this is an area for further work.

We have yet to establish the psychometric validity of the DHP-3D and DHP-5D in comparison with generic measures such as the EQ-5D. This is important because generic measures are widely used and accepted by reimbursement agencies such as NICE. Therefore, the validation of condition-specific preference-based measures in comparison with generic instruments is a key area for further work as this evidence will support the use of the DHP-3D and DHP-5D in the estimation of QALYs. It is worth noting that the DHP-18 and EQ-5D have a level of validity for use in the assessment of diabetes related health status [8,29,30].

In conclusion, we have developed two diabetes-specific preference-based measures that, subject to psychometric assessment, we suggest for use to provide condition-specific utility values to complement generic utilities from more widely validated measures.

Competing interests

KM was responsible for development of the DHP-1 and DHP-18.

References

1 Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997; **35**: 1095–1108.
2 Brooks R. EuroQol: the current state of play. *Health Policy* 1996; **37**: 53–72.
3 Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011; **20**: 1727–1736.
4 National Institute for Health and Care Excellence (NICE). *Methods for Health Technology Assessment 2013*. London: NICE, 2013.
5 Longworth L, Yang Y, Young T, Mulhern B, Hernandez-Alava M, Mukuria C et al. Use of generic and condition specific measures of health related quality of life in NICE decision making: systematic review, statistical modelling and survey. *Health Technol Assess* 2014; **18**: 9.
6 Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol Assess* 2014; **18**: 34.
7 Janssen M, Lubetkin EI, Sekhobo JP, Pickard AS. The use of the EQ-5D preference-based health status measure in adults with Type 2 diabetes mellitus. *Diabet Med* 2011; **28**: 393–413.
8 Richardson J, Khan M, Iezzi A, Maxwell A. Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI3, 15D, QWB, and AQoL-8D multiattribute utility instruments. *Med Decis Making* 2015; **35**: 276–291.

9 Meadows KA, Steen N, McColl E, Eccles M, Shiels C, Hewison J *et al*. The Diabetes Health Profile (DHP): a new instrument for assessing the psychosocial profile of insulin requiring patients – development and psychometric evaluation. *Qual Life Res* 1996; **5**: 242–254.

10 Meadows K, Abrams C, Sandaek A. Adaptation of the diabetes health profile (DHP-1) for use with patients with Type 2 diabetes mellitus: psychometric evaluation and cross-cultural comparison. *Diabet Med* 2000; **17**: 572–580.

11 Mulhern B, Meadows K. The construct validity and responsiveness of the EQ-5D, SF-6D and Diabetes Health Profile-18 in type 2 diabetes. *Health Qual Life Outcomes* 2014; **12**: 42.

12 Brazier JE, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y *et al*. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess* 2012; **16**: 32.

13 Rowen D, Brazier J, Young T, Gaugris S, Craig BM, King MT *et al*. Deriving a preference based measure for cancer using the EORTC QLQ-C30. *Value Health* 2011; **14**: 721–731.

14 King MT, Costa DSJ, Aaronson NK, Brazier J, Cella D, Fayers P *et al*. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. *Qual Life Res* 2016; **25**: 625–636.

15 Mukuria C, Rowen D, Brazier JE, Young T, Nafees B. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health* 2015; **18**: 846–855.

16 Rowen D, Mulhern B, Banerjee S, van Hout B, Young TA, Knapp M *et al*. Estimating preference based single index measures for dementia using DEMQOL and DEMQOL-Proxy. *Value Health* 2012; **15**: 346–356.

17 Mulhern B, Smith SC, Rowen D, Brazier JE, Knapp M, Lamping DL *et al*. Improving the measurement of QALYs in dementia: developing patient- and carer-reported health state classification systems using Rasch analysis. *Value Health* 2012; **15**: 323–333.

18 Mulhern B, Rowen D, Jacoby A, Marson T, Snape D, Hughes D *et al*. The development of a QALY measure for epilepsy: NEW-QOL-6D. *Epilepsy Behav* 2012; **24**: 36–43.

19 Erpelding ML, Boini S, Fagot-Campagna A, Mesbah M, Chwalow J, Penfornis A *et al*. Health related quality of life reference values (DHP) in people with diabetes living in France – Entred study, 2001–2003. *Bull Epidemiol Hebd (Paris)* 2009; **34**: 368–371.

20 Hippisley-Cox J, Yates J, Pringle M. Sex inequalities in access to care for patients with diabetes in primary care: questionnaire survey. *Br J Gen Pract* 2006; **56**: 342–348.

21 Farr A, Phillips CJ, Davies S, Morgan S. *Changes in health status of diabetic patients in Bridgend – Final report*. Swansea: University of Swansea, 2010.

22 Ruddock S, Fosbury J, Smith A, Meadows K, Crown A. Measuring psychological morbidity for diabetes commissioning: a cross-sectional survey of patients attending a secondary care diabetes clinic. *Practical Diabetes Int* 2010; **27**: 22–26.

23 Peters M, Crocker H, Jenkinson C, Doll H, Fitzpatrick R. The routine collection of patient-reported outcome measures (PROMs) for long-term conditions in primary care: a cohort survey. *BMJ Open* 2013; **4**: e003968.

24 Preacher KJ, MacCallum RC. Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics* 2013; **2**: 13–43.

25 Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press, 1980 (reprint).

26 Smith SC, Lamping DL, Banerjee S, Harwood R, Foley B, Smith P *et al*. Development of a new measure of health-related quality of life for people with dementia: DEMQOL. *Psychol Med* 2007; **37**: 737–746.

27 Janssen MF, Oppe M, Versteegh MM, Stolk EA. Introducing the composite Time Trade-Off: a test of feasibility and face validity. *Eur J Health Econ* 2013; **14**(Suppl. 1): 5–13.

28 Sundaram M, Smith MJ, Revicki DA, Elswick B, Miller LA. Rasch analysis informed the development of a classification system for a diabetes-specific preference-based measure of health. *J Clin Epidemiol* 2009; **62**: 845–856.

29 Sundaram M, Smith MJ, Revicki DA, Miller LA, Madhavan S, Hobbs G. Estimation of a valuation function for a diabetes mellitus-specific preference-based measure of health: the Diabetes Utility Index. *Pharmacoeconomics* 2010; **28**: 201–216.

30 Janssen M, Lubetkin EL, Sekhobo JP, Pickard AS. The use of the EQ-5D preference-based health status measure in adults with Type 2 diabetes mellitus. *Diabet Med* 2011; **28**: 393–413.

31 Kontodimopoulos N, Pappa E, Chadjiapostolou Z, Arvanitaki E, Papadopoulos AA, Niakas D. Comparing the sensitivity of EQ-5D, SF-6D and 15D utilities to the specific effect of diabetic complications. *Eur J Health Econ* 2012; **13**: 111–112.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Descriptive statistics of observed TTO for DHP-3D and DHP-5D (ordered by mean score)