Research: Health Economics

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

B. Mulhern¹,², A. Labeit², D. Rowen², E. Knowles², K. Meadows³, J. Elliott⁴ and J. Brazier²

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

Accepted 3 May 2017

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.

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
What’s new?

- We have developed the first diabetes-specific preference-based measures with United Kingdom value sets.
- The measures can be used to estimate Quality Adjusted Life Years for use in the economic evaluation of diabetes-specific interventions.

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 followed by 10 years in the 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 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:

\[ U_{ij} = g(\beta' x_i) + \epsilon_{it} \]  (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 \( \epsilon_{it} \) 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

\[ z = (\beta_1 - \beta_2)/\sqrt{(se_1^2 + se_2^2)} \]  (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>
<thead>
<tr>
<th>Dimension</th>
<th>DHP-18 dimension</th>
<th>Loading</th>
<th>% at ceiling</th>
<th>% at floor</th>
<th>% missing data</th>
<th>Psychometric analysis</th>
<th>Rasch analysis</th>
<th>Differential item functioning</th>
<th>Fit residual</th>
<th>Spread at logit 0</th>
<th>Item range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychological distress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological distress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6: Lose temper if people keep on at you</td>
<td>PD</td>
<td>0.68</td>
<td>53.2</td>
<td>8.0</td>
<td>2.1</td>
<td></td>
<td>Yes (Sometimes/ usually)</td>
<td>-0.42</td>
<td>0.22</td>
<td>0.37 to 0.84</td>
<td>-1.65 to 0.53</td>
</tr>
<tr>
<td>Q8: More arguments due to diabetes</td>
<td>PD</td>
<td>0.78</td>
<td>62.0</td>
<td>4.2</td>
<td>1.3</td>
<td></td>
<td>Yes (Sometimes/often)</td>
<td>-0.45</td>
<td>0.72</td>
<td>0.14 to 0.85</td>
<td>-1.72 to 1.85</td>
</tr>
<tr>
<td>Q16: Lose temper or shout</td>
<td>PD</td>
<td>0.99</td>
<td>46.4</td>
<td>6.3</td>
<td>0</td>
<td></td>
<td>Gender</td>
<td>-0.20</td>
<td>0.68</td>
<td>0.04 to 0.70</td>
<td>-0.87 to 3.16</td>
</tr>
<tr>
<td>Q17: Get touchy or moody</td>
<td>PD</td>
<td>0.72</td>
<td>46.8</td>
<td>4.6</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q18: Lose temper over small things</td>
<td>PD</td>
<td>0.83</td>
<td>32.5</td>
<td>7.2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1: Food control life</td>
<td>BA</td>
<td>0.60</td>
<td>24.9</td>
<td>7.2</td>
<td>2.1</td>
<td></td>
<td>Yes (sometimes/usually)</td>
<td>0.66</td>
<td>0.41</td>
<td>0.20 to 0.85</td>
<td>-1.71 to 1.37</td>
</tr>
<tr>
<td>Q5: Eat to cheer self up</td>
<td>DE</td>
<td>0.68</td>
<td>31.6</td>
<td>6.3</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7: Hard saying no to foods</td>
<td>DE</td>
<td>0.73</td>
<td>17.6</td>
<td>14.2</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9: Easy to stop eating</td>
<td>DE</td>
<td>0.64</td>
<td>17.3</td>
<td>3.8</td>
<td>0.4</td>
<td></td>
<td>Yes (sometimes/usually)</td>
<td>0.03</td>
<td>0.91</td>
<td>0.11 to 0.93</td>
<td>-2.64 to 2.04</td>
</tr>
<tr>
<td>Q10: Eat more when bored/fed up</td>
<td>DE</td>
<td>0.67</td>
<td>10.5</td>
<td>16.9</td>
<td>0.4</td>
<td></td>
<td>Gender</td>
<td>-0.68</td>
<td>0.31</td>
<td>0.30 to 0.85</td>
<td>-1.71 to 0.85</td>
</tr>
<tr>
<td>Q12: Wish not so many nice things to eat</td>
<td>DE</td>
<td>0.49</td>
<td>29.1</td>
<td>22.8</td>
<td>0</td>
<td></td>
<td>Yes (a little/a lot)</td>
<td>0.03</td>
<td>0.91</td>
<td>0.11 to 0.93</td>
<td>-2.64 to 2.04</td>
</tr>
<tr>
<td>Social limitations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2: Difficult staying out late</td>
<td>BA</td>
<td>0.63</td>
<td>52.3</td>
<td>5.5</td>
<td>2.1</td>
<td></td>
<td>Yes (usually/sometimes)</td>
<td>-0.76</td>
<td>0.11</td>
<td>0.11 to 0.88</td>
<td>-2.01 to 2.10</td>
</tr>
<tr>
<td>Q3: Days tied to meal times</td>
<td>BA</td>
<td>0.73</td>
<td>34.6</td>
<td>3.8</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4: Avoid going out if sugars low</td>
<td>BA</td>
<td>0.71</td>
<td>49.8</td>
<td>11.8</td>
<td>2.5</td>
<td></td>
<td>Yes (usually/sometimes)</td>
<td>-0.76</td>
<td>0.11</td>
<td>0.11 to 0.88</td>
<td>-2.01 to 2.10</td>
</tr>
<tr>
<td>Q11: Worry about getting colds/flu</td>
<td>BA</td>
<td>0.39</td>
<td>48.9</td>
<td>7.2</td>
<td>0.4</td>
<td></td>
<td>Yes (a little/a lot)</td>
<td>0.03</td>
<td>0.91</td>
<td>0.11 to 0.93</td>
<td>-2.64 to 2.04</td>
</tr>
<tr>
<td>Q14: Get edgy</td>
<td>BA</td>
<td>0.53</td>
<td>55.3</td>
<td>6.8</td>
<td>2.1</td>
<td></td>
<td>Yes (a little/a lot)</td>
<td>0.03</td>
<td>0.91</td>
<td>0.11 to 0.93</td>
<td>-2.64 to 2.04</td>
</tr>
<tr>
<td>Q13: Frightening going into busy shops</td>
<td>BA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

**Mood**
- You never find yourself losing your temper over small things
- You sometimes find yourself losing your temper over small things
- You often find yourself losing your temper over small things
- You very often find yourself losing your temper over small things

**Social limitations**
- Your days are never tied to meal times
- Your days are sometimes tied to meal times
- Your days are usually tied to meal times
- Your days are always tied to meal times

**Eating**
- When you start eating you find it very easy to stop
- When you start eating you find it quite easy to stop
- When you start eating you find it not very easy to stop
- When you start eating you find it not at all easy to stop

**DHP-5D**

**Mood**
- You never find yourself losing your temper over small things
- You sometimes find yourself losing your temper over small things
- You often find yourself losing your temper over small things
- You very often find yourself losing your temper over small things

**Social limitations**
- Your days are never tied to meal times
- Your days are sometimes tied to meal times
- Your days are usually tied to meal times
- Your days are always tied to meal times

**Eating**
- When you start eating you find it very easy to stop
- When you start eating you find it quite easy to stop
- When you start eating you find it not very easy to stop
- When you start eating you find it not at all easy to stop

**Hypoglycaemic attacks**
- You never worry about doing too much and going hypo
- You sometimes worry about doing too much and going hypo
- You usually worry about doing too much and going hypo
- You always worry about doing too much and going hypo

**Vitality**
- You are tired none of the time
- You are tired a little of the time
- You are tired some of the time
- You are tired most of the time
- You are tired all of the time

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

---

**FIGURE 2** Histograms of observed DHP-3D (panel a) and DHP-5D (panel b) values.
Table 3 Regression models estimating preference weights for the DHP-3D and DHP-5D

<table>
<thead>
<tr>
<th>(1) OLS</th>
<th>(2) RE-GLS</th>
<th>(3) RE-Tobit</th>
<th>(4) RE-Tobit (marginal effects)</th>
<th>(5) Mean model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mood level 2</td>
<td>-0.030*** (0.008)</td>
<td>-0.026*** (0.009)</td>
<td>-0.038*** (0.010)</td>
<td>-0.025*** (0.007)</td>
</tr>
<tr>
<td>Mood level 3</td>
<td>-0.098*** (0.016)</td>
<td>-0.085*** (0.010)</td>
<td>-0.100*** (0.011)</td>
<td>-0.072*** (0.009)</td>
</tr>
<tr>
<td>Mood level 4</td>
<td>-0.143*** (0.018)</td>
<td>-0.127*** (0.010)</td>
<td>-0.144*** (0.011)</td>
<td>-0.110*** (0.009)</td>
</tr>
<tr>
<td>Eating level 2</td>
<td>-0.004 (0.012)</td>
<td>0.000 (0.010)</td>
<td>-0.001 (0.011)</td>
<td>-0.001 (0.008)</td>
</tr>
<tr>
<td>Eating level 3</td>
<td>-0.066*** (0.012)</td>
<td>-0.060*** (0.010)</td>
<td>-0.069*** (0.008)</td>
<td>-0.053*** (0.008)</td>
</tr>
<tr>
<td>Eating level 4</td>
<td>-0.058*** (0.010)</td>
<td>-0.074*** (0.010)</td>
<td>-0.083*** (0.010)</td>
<td>-0.065*** (0.008)</td>
</tr>
<tr>
<td>Social limitations level 2</td>
<td>0.011 (0.012)</td>
<td>-0.022** (0.010)</td>
<td>-0.026** (0.011)</td>
<td>-0.019** (0.008)</td>
</tr>
<tr>
<td>Social limitations level 3</td>
<td>-0.003 (0.020)</td>
<td>-0.044*** (0.012)</td>
<td>-0.051*** (0.013)</td>
<td>-0.039** (0.010)</td>
</tr>
<tr>
<td>Social limitations level 4</td>
<td>-0.044*** (0.010)</td>
<td>-0.065*** (0.009)</td>
<td>-0.071*** (0.010)</td>
<td>-0.055*** (0.008)</td>
</tr>
<tr>
<td>Hypoglycaemic attacks level 2</td>
<td>-0.144*** (0.010)</td>
<td>-0.069*** (0.010)</td>
<td>-0.083*** (0.010)</td>
<td>-0.065*** (0.008)</td>
</tr>
<tr>
<td>Hypoglycaemic attacks level 3</td>
<td>-0.100*** (0.010)</td>
<td>-0.127*** (0.010)</td>
<td>-0.144*** (0.011)</td>
<td>-0.110*** (0.009)</td>
</tr>
<tr>
<td>Hypoglycaemic attacks level 4</td>
<td>-0.004 (0.012)</td>
<td>0.000 (0.010)</td>
<td>-0.001 (0.011)</td>
<td>-0.001 (0.008)</td>
</tr>
<tr>
<td>Vitality level 2</td>
<td>0.011 (0.012)</td>
<td>-0.022** (0.010)</td>
<td>-0.026** (0.011)</td>
<td>-0.019** (0.008)</td>
</tr>
<tr>
<td>Vitality level 3</td>
<td>-0.003 (0.020)</td>
<td>-0.044*** (0.012)</td>
<td>-0.051*** (0.013)</td>
<td>-0.039** (0.010)</td>
</tr>
<tr>
<td>Vitality level 4</td>
<td>-0.044*** (0.010)</td>
<td>-0.065*** (0.009)</td>
<td>-0.071*** (0.010)</td>
<td>-0.055*** (0.008)</td>
</tr>
<tr>
<td>Vitality level 5</td>
<td>0.011 (0.012)</td>
<td>-0.022** (0.010)</td>
<td>-0.026** (0.011)</td>
<td>-0.019** (0.008)</td>
</tr>
</tbody>
</table>

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 (44444). 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-3D

<table>
<thead>
<tr>
<th>OLS</th>
<th>RE-GLS consistent model</th>
<th>RE-Tobit marginal effects</th>
<th>Mean model 2 vs. 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.021** (0.008)</td>
<td>-0.016 (0.010)</td>
<td>-0.015 (0.010)</td>
<td>-0.019* (0.010)</td>
</tr>
<tr>
<td>-0.026** (0.013)</td>
<td>-0.027** (0.011)</td>
<td>-0.027** (0.011)</td>
<td>-0.032** (0.012)</td>
</tr>
<tr>
<td>-0.066*** (0.016)</td>
<td>-0.051*** (0.012)</td>
<td>-0.051*** (0.011)</td>
<td>-0.054*** (0.012)</td>
</tr>
<tr>
<td>-0.042** (0.012)</td>
<td>-0.026** (0.012)</td>
<td>-0.026** (0.012)</td>
<td>-0.029** (0.012)</td>
</tr>
<tr>
<td>-0.053*** (0.012)</td>
<td>-0.045*** (0.011)</td>
<td>-0.043** (0.010)</td>
<td>-0.050*** (0.011)</td>
</tr>
<tr>
<td>-0.043*** (0.010)</td>
<td>-0.041*** (0.011)</td>
<td>-0.043** (0.010)</td>
<td>-0.045*** (0.011)</td>
</tr>
<tr>
<td>-0.005 (0.011)</td>
<td>-0.014 (0.011)</td>
<td>-0.012 (0.010)</td>
<td>-0.018 (0.011)</td>
</tr>
<tr>
<td>-0.001 (0.016)</td>
<td>-0.009 (0.013)</td>
<td>-0.012 (0.010)</td>
<td>-0.014 (0.013)</td>
</tr>
<tr>
<td>-0.026** (0.010)</td>
<td>-0.029*** (0.010)</td>
<td>-0.029*** (0.010)</td>
<td>-0.034*** (0.010)</td>
</tr>
<tr>
<td>-0.025** (0.011)</td>
<td>-0.019* (0.010)</td>
<td>-0.019* (0.010)</td>
<td>-0.022** (0.010)</td>
</tr>
<tr>
<td>-0.036*** (0.013)</td>
<td>-0.035*** (0.011)</td>
<td>-0.036*** (0.011)</td>
<td>-0.041*** (0.012)</td>
</tr>
<tr>
<td>-0.067*** (0.009)</td>
<td>-0.065*** (0.010)</td>
<td>-0.065*** (0.010)</td>
<td>-0.070*** (0.010)</td>
</tr>
<tr>
<td>-0.016** (0.008)</td>
<td>-0.028*** (0.010)</td>
<td>-0.028*** (0.010)</td>
<td>-0.033*** (0.011)</td>
</tr>
<tr>
<td>-0.056*** (0.011)</td>
<td>-0.044*** (0.011)</td>
<td>-0.044*** (0.010)</td>
<td>-0.049*** (0.011)</td>
</tr>
<tr>
<td>-0.115** (0.014)</td>
<td>-0.115*** (0.013)</td>
<td>-0.115*** (0.012)</td>
<td>-0.122*** (0.013)</td>
</tr>
<tr>
<td>-0.173*** (0.018)</td>
<td>-0.173*** (0.011)</td>
<td>-0.173*** (0.011)</td>
<td>-0.179*** (0.011)</td>
</tr>
<tr>
<td>0.988*** (0.008)</td>
<td>0.979*** (0.018)</td>
<td>0.979*** (0.018)</td>
<td>0.999*** (0.019)</td>
</tr>
</tbody>
</table>

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.
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.

**FIGURE 3** Observed vs predicted DHP-3D (panel a) and DHP-5D (panel b) values.
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 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.

Funding sources
This study was funded by Merck & Co., Inc., Kenilworth, NJ, USA.

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

Acknowledgements
We would like to acknowledge Simon Heller (Academic Unit of Diabetes, Endocrinology and Metabolism, University of Sheffield) and Sue Hudson (Sheffield Teaching Hospitals) for their support in developing the DHP-3D and DHP-5D classification systems, and Charlotte Kosmas, Beenish Nafees and the interview team at ICON PLC for coordinating and carrying out the health state valuation interviews. We would also like to thank the respondents for taking part in the interview.

References


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)