eq5dmap: a command for mapping from 3-level to 5-level EQ-5D

Mónica Hernández-Alava
School of Health and Related Research
Health Economics and Decision Science
University of Sheffield
Sheffield, UK
monica.hernandez@sheffield.ac.uk

Stephen Pudney
School of Health and Related Research
Health Economics and Decision Science
University of Sheffield
Sheffield, UK
steve.pudney@sheffield.ac.uk

Abstract. This article describes a new Stata command, eq5dmap, for conditional prediction of the utility values of the newer 5-level version of EQ-5D from observed or specified values of 3-level EQ-5D conditional on age and gender. Predictions can be made either from the 5-item EQ-5D-3L health description, or from the (exact or approximate) utility score. The prediction process is based on a joint statistical model of the two variants of EQ-5D which have been fitted to alternative reference datasets. The underlying model is a system of ordinal regressions with a flexible residual distribution specified as Gaussian or as a copula-mixture. Use of the command is illustrated with an application which includes an investigation of the sensitivity of the mapping outcomes to the choice of reference dataset.

Keywords: st0001, EQ-5D, mapping, conditional prediction, copula, mixture model

1 Introduction

The quality-adjusted life year (QALY) is one of the most widely used health benefit measures in economic evaluations of interventions, services or programmes designed to improve health. The QALY allows health care decision makers to use a consistent approach across a broad range of disease areas, treatments and patients. It is the preferred outcome measure for the National Institute for Health and Care Excellence (NICE) in its appraisals of health interventions in England (NICE 2014). Preference-based measures such as the EQ-5D underpin the calculation of QALYs.

The EQ-5D instrument describes health states in terms of five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The original version, the EQ-5D-3L, measures each dimension on a three level scale (no problems, some or moderate problems, extreme problems). EQ-5D-3L can describe 243 different health states in this way. For example, the health state 11223 corresponds to no problems in the mobility and self-care dimensions, some problems in the usual activities and pain/discomfort dimensions, and extreme problems in the anxiety/depression dimension. Valuation studies in different countries assigned a utility score to each of the health states described by the instrument. Dolan (1997) carried out the first UK valuation study using general public preferences. Other countries have developed al-
ternative scoring systems but, in all countries the health state 11111 (full health) is assigned a utility score of 1 and death is assigned a value of 0. States with utility scores between 0 and 1 reflect some degree of impairment and states with negative valuations are considered worse than death.

A new version of the health description system, the EQ-5D-5L, has been developed to try to address concerns about the lack of sensitivity and floor/ceiling distortions of the EQ-5D-3L. The number of dimensions has remained unchanged but the new version extends the number of levels per dimension from three to five (no problems, slight problems, moderate problems, severe problems, extreme problems). To improve consistency across dimensions and aid understanding there have also been some wording changes. The number of discrete health states described by the new version is 3125. Utility value sets for EQ5D-5L have been released for England (Devlin et al. 2016), Japan (Ikeda et al. 2015), Canada (Xie et al. 2016), Uruguay (Augustovski et al. 2015), Netherlands (Versteegh et al. 2016) and Korea (Kim et al. 2016) and similar work is underway in other countries.

Many studies now include the five level version of EQ-5D instead of the original version. Since all studies, new and those previously completed, will form part of the available evidence in future economic evaluations, it is important to have a consistent way of translating health benefits measured using one of the two versions of EQ-5D into the other. Hernández-Alava and Pudney (2016) developed a flexible model which allows analysis of the joint responses to EQ-5D-3L and EQ-5D-5L. The underlying model is a system of ordinal regressions with a flexible copula mixture residual distribution. This model has been re-estimated using two different datasets and results reported elsewhere (Hernandez-Alava et al., 2017). In this article, we describe a new Stata command, eq5dmap, for conditional prediction of the EQ-5D-5L utility from observed EQ-5D-3L responses or specified utility values and age and gender. The command predictions are based on the models in Hernandez-Alava et al. (2017). Section 2 describes the two types of predictions that can be computed with the command. Section 3 explains briefly the underlying statistical model. The command syntax is fully described in Section 4 and Section 5 presents an illustrative example of the use of the command.

2 The mapping method

At present, the eq5dmap command offers only mapping from the 3-level version of EQ-5D to the 5-level version using the UK value sets. There are two reasons for this. First, EQ-5D-5L appears superior in terms of sensitivity and therefore gives a better basis for cost-effectiveness work. A second reason is that the most commonly-used 3L scoring system due to Dolan (1997) tends to generate a relatively high frequency of very low QALY outcomes, which is difficult to capture in 5L → 3L mapping (Hernández-Alava and Pudney, 2016). Future versions of eq5dmap will allow mapping in both directions and will include value sets for other countries.

1. See the EuroQol website http://www.euroqol.org/eq-5d-products/how-to-obtain-eq-5d.html for examples of the question wording used in EQ-5D-3L and EQ-5D-5L.
Let $Y_d \in \{1, 2, 3\}$ and $Y_{5d} \in \{1, 2, 3, 4, 5\}$ represent outcomes for the $d$th domain $(d = 1 \ldots 5)$ of the 3- and 5-level forms of EQ-5D. Define the vectors $Y_3 = (Y_{31} \ldots Y_{35})$ and $Y_5 = (Y_{51} \ldots Y_{55})$, and write the corresponding utility scoring scales $v_3(\cdot)$ and $v_5(\cdot)$. Our aim is to calculate the expectation of $v_5(Y_5)$ conditional on the values of a vector of covariates $X$, and also on available information about $Y_3$. Depending on the form of that information, two types of mapping can be done:

### 2.1 A specified value for $Y_3$

Let $S_5 = \{1 \ldots 5\}^5$ be the set of possible values that can be taken by the vector $Y_5$. If we know the conditioning value of $Y_3$, the expectation of $v_5$ can be computed as:

$$E(v_5|Y_3 = y_3, X) = \sum_{y_5 \in S_5} v_5(y_5) p(y_5|y_3, X)$$

(1)

where $p(y_5|y_3, X)$ is the form of conditional probability implied by the specified underlying statistical model for the joint distribution $Y_3, Y_5|X$.

### 2.2 A specified (approximate) value for $v_3(Y_3)$

In some cases, the user may know only the value of $v_3(Y_3)$, rather than $Y_3$ itself. Since the mapping $Y_3 \rightarrow v_3$ is not (quite) one-to-one, this case involves weaker conditioning information. Another possibility is that the user has only a predicted value $\hat{v}_3$, and the prediction may not correspond precisely to any valid 3L utility score. We handle both problems by distance-weighted averaging within a neighbourhood of the specified value, $\hat{v}_3$.

Let $C$ be a user-specified caliper. Define a set of vectors $\mathcal{S}(v) = \{Y_3 : |v_3(Y_3) - v| \leq C\}$ and a weight function of Epanechnikov form:

$$\omega(v_3 - v) = \begin{cases} 
1 - [(v_3 - v)/C]^2 & \text{for } |(v_3 - v)| < C \\
0 & \text{otherwise} 
\end{cases}$$

(2)

The estimate of the expected value of $v_5$ is:

$$E(v_5 | |v_3(Y_3) - \hat{v}_3| \leq C, X) = \frac{\sum_{y_5 \in \mathcal{S}(v_3)} \omega(v_3(y_5) - \hat{v}_3) \sum_{y_5} v_5(y_5) p(y_5|y_3, X)}{\sum_{y_5 \in \mathcal{S}(v_3)} \omega(v_3(y_5) - \hat{v}_3)}$$

(3)

where the summation over $y_5$ covers all 3125 possible outcome vectors for the EQ-5D-5L descriptive system.

### 3 The underlying statistical model

The predictive distribution $Pr(Y_5|Y_3, X)$ is derived from a model of the joint distribution of $Y_3, Y_5|X$. That model is a system of ten latent regressions, arranged in five
eq5dmap a mapping command for EQ-5D

groups, following the natural pairing of the dimensions in the two versions of EQ-5D, with domain $d$ containing the equations for $Y_{3d}$ and $Y_{5d}$:

\[
\begin{align*}
Y_{3d}^* &= X \beta_{3d} + U_{3d} \\
Y_{5d}^* &= X \beta_{5d} + U_{5d}
\end{align*}
\]

where $i$ indexes individual cases and we assume random sampling so that all sampled variables are independent across individuals. $X$ is a row vector of covariates and $\beta_{3d}, \beta_{5d}$ are column vectors of coefficients conformable with $X$. We assume that the covariate vector $X$ is the same for both the 3-level and 5-level version of the $r$th domain, but may differ between domains. $U_{3d}, U_{5d}$ are unobserved residuals which may be stochastically dependent and non-normal. The latent dependent variables $Y_{3d}^*, Y_{5d}^*$ are not observed directly but they have observable ordinal counterparts, $Y_{3d}, Y_{5d}$, generated by the following threshold-crossing conditions:

\[
Y_{kd} = q \text{  iff  } \Gamma_{kd} \leq Y_{kd}^* < \Gamma_{(q+1)d}; \quad q = 1...Q_k; \quad k = 3, 5 \quad (5)
\]

where $Q_k = 3$ or 5 is the number of categories of $Y_{kd}$ and the $\Gamma_{kd}$ are threshold parameters, with $\Gamma_{kd} = -\infty$ and $\Gamma_{(Q_k+1)d} = +\infty$.

To allow for background correlation between the five dimensions of EQ-5D, the residual $U_{kd}$ is decomposed into a single between-group factor $V_i$ which represents the individual’s general tendency to give more or less positive responses and a specific residual $\varepsilon_{kd}$ correlated within but not between dimensions:

\[
U_{kd} = \psi_{kd} V_i + \varepsilon_{kd}, \quad k = 3, 5; \quad d = 1...5 \quad (6)
\]

where the $\psi_{kd}$ are a set of ten parameters.

The model can be estimated under various alternative assumptions about the joint distribution of the residuals $\varepsilon_{3id}$ and $\varepsilon_{5id}$ within each dimension $d$ and the distributional form of the common factor $V_i$. The eq5dmap command offers two specifications: the Gaussian, where the pairs $\varepsilon_{3id}, \varepsilon_{5id}$ have bivariate normal distributions and $V_i \sim N(0, 1)$; and the copula specification where the distribution of each pair $\varepsilon_{3id}, \varepsilon_{5id}$ is specified in copula form with normal mixture marginals and $V_i$ as a normal mixture (see Hernández-Alava and Pudney (2016) for details).

These two variants of the mapping model have been estimated using two alternative datasets. The first comes from the National Data Bank for Rheumatic Diseases (NDB), which is a register of patients, mainly referred by US and Canadian rheumatologists (Wolfe and Michaud 2011). In 2011, there was a switch from the 3-level to the 5-level version of EQ-5D and both versions were collected in parallel during the January 2011 wave, which we used to estimate the reference model. The NDB dataset covers a Rheumatoid Arthritis (RA) specific sample. The second dataset comes from a data collection study coordinated and partly funded by the EuroQuol Group (EQG) between August 2009 and September 2010, in 6 countries: Denmark, England, Italy, the Netherlands, Poland and Scotland. It covered eight broad patient groups (cardiovascular disease, respiratory disease, depression, diabetes, liver disease, personality disorders,
arthritis, and stroke) and a student cohort (healthy population). This EQG dataset was intended to cover a wide range of responses across all the EQ-5D dimensions in a range of diseases (Janssen et al. 2013; van Hout et al. 2012). The EQG sample is younger than the NDB sample, with an average age of 51 versus 63, and it covers a wider age range. There is a big difference in gender composition: the EQG sample is 53% female, compared to 81% for NDB, in line with what is expected in an RA specific sample. Using the UK value sets, the EQG sample has lower health-related quality of life, with average UK utility values of 0.628 and 0.712 for 3- and 5-level EQ-5D respectively, versus 0.681 and 0.779 in the NDB dataset. Estimated reference models for these datasets are described in Hernández-Alava and Pudney (2016) and Hernandez-Alava et al. (2017).

4 Command syntax

```plaintext
eq5dmap outputvarname [ if ] [ in ] [ weight ], covariates(varlist) [
    model(modelname) items3(varlist) score3(varname)
    utility3(3Lutilityscalename) utility5(5Lutilityscalename) caliper(#) ]
```

Description

`eq5dmap` is a user-written program which allows outcomes measured using the older version of the instrument (EQ-5D-3L) to be converted into (expected) utility values measured using the newer 5-level version (EQ-5D-5L).

The predictions are constructed from an underlying statistical model as described in section 3. The model is not estimated by the `eq5dmap` command; instead, estimation results are selected from a collection of existing estimates derived from alternative model specifications and alternative reference datasets.

Output

`eq5dmap` returns the calculated conditional expectation of the 5-level EQ-5D utility score in the variable `outputvarname`. It also uses the `summarize` command to give a (weighted) summary of the predicted 5-level scores within the subset of observations defined by the `if` and `in` qualifiers, if they are specified.

Options

covariates(varlist) specifies the variables used as covariates. Mapping is age- and gender-specific, so there are two covariates. They must be specified as a `varlist` with the items ordered as: age (in years); gender (coded as female = 0, male = 1).

model(modelname) specifies the model to be used for the mapping. The available options for `modelname` are: NDBgauss, NDBcopula, EQGcopula and EQGgauss. If omitted, the default is EQGcopula.
If the predictor is a set of values for $Y_3$:

\textbf{items3(varlist)} specifies the variables which contain observed values for the five EQ-5D-3L domain items. They must be specified as a \textit{varlist} containing five variables ordered as: mobility; self-care; usual activities; pain; anxiety/depression. The variables should all be coded on a scale 1, 2, 3 where 1 = no problems, . . . 3 = extreme problems.

If the predictor is a value $\hat{\upsilon}_3$:

\textbf{score3(varname)} specifies a variable which contains the value of the 3L utility score.

\textbf{utility3(3Lutilityscalename)} specifies one of the alternative 3L utility scales offered. Currently, the only one offered is \textbf{UK} which specifies the utility scale described by Dolan (1997). If omitted, the default is \textbf{utility3(UK)}.

\textbf{utility5(5Lutilityscalename)} specifies one of the alternative 5L utility scales offered. Currently, the only one offered is \textbf{UK} which specifies the utility scale described by Devlin et al. (2016). If omitted, the default is \textbf{5Lutility(UK)}.

\textbf{caliper(#)} specifies the maximum distance to be allowed between the predictor value $\hat{\upsilon}_3$ and any valid 3L utility score. If omitted, the default is \textbf{caliper(0)}, which enforces exact matching of $\hat{\upsilon}_3$ to a point on the 3L utility scale. With the caliper set to 0, if there is a multiplicity of points on the 3L scale that match $\hat{\upsilon}_3$ exactly, then their average is returned as the value for outputvarname.

Either \textbf{items3} or \textbf{score3} must appear, but not both. If \textbf{score3} is used and it proves impossible to find a valid point on the 3L utility scale within a distance $C$ of $\hat{\upsilon}_3$, then \textbf{outputvarname} is returned with a missing value and a warning is written to the log file.

5 Examples

In sections 5.1 and 5.2, we give examples of the basic use of \textbf{eq5dmap} for $Y_3 \rightarrow \upsilon_5$ mapping and (exact) $\upsilon_3 \rightarrow \upsilon_5$ mapping. Section 5.3 gives recommendations on the choice of caliper for approximate $\upsilon_3 \rightarrow \upsilon_5$ mapping, and section 5.4 considers the potential sensitivity of results to the choice of reference dataset and model specification.

5.1 Basic use of \textbf{eq5dmap}: mapping from EQ-5D-3L items to a EQ-5D-5L utility score

\textbf{eq5dmap} is provided with a dataset that lists all possible EQ-5D-3L health states by gender and all ages from 16 to 100 (it thus contains $N = 243 \times 2 \times 85 = 41,310$ records. The dataset includes a set of 5 ordinal variables $Y_{3,1}$ to $Y_{3,5}$ corresponding to the 5 EQ-5D-3L dimensions\footnote{The dataset also contains the corresponding utility value, $u_3$, calculated using the UK value set, age and a dummy variable for gender. Two weight variables \textbf{fwEQG} and \textbf{fwNDB} are also included. For the purposes of this example, we have also included an additional variable $u_3hat$ containing an}. The data are summarized below.
// Load and examine the input dataset
use "eq5dmap_data.dta", clear
summarize

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y3_1</td>
<td>41,310</td>
<td>2</td>
<td>.8165065</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Y3_2</td>
<td>41,310</td>
<td>2</td>
<td>.8165065</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Y3_3</td>
<td>41,310</td>
<td>2</td>
<td>.8165065</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Y3_4</td>
<td>41,310</td>
<td>2</td>
<td>.8165065</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Y3_5</td>
<td>41,310</td>
<td>2</td>
<td>.8165065</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>male</td>
<td>41,310</td>
<td>.5</td>
<td>.5000061</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>age</td>
<td>41,310</td>
<td>58</td>
<td>24.53599</td>
<td>16</td>
<td>100</td>
</tr>
<tr>
<td>fwEQG</td>
<td>41,310</td>
<td>.0856693</td>
<td>.626692</td>
<td>0</td>
<td>37</td>
</tr>
<tr>
<td>fwNDB</td>
<td>41,310</td>
<td>.1259985</td>
<td>1.05797</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td>u3</td>
<td>41,310</td>
<td>.1367572</td>
<td>.3105279</td>
<td>-.594</td>
<td>1</td>
</tr>
<tr>
<td>u3hat</td>
<td>41,310</td>
<td>.1372946</td>
<td>.3011919</td>
<td>-.5935036</td>
<td>.781076</td>
</tr>
</tbody>
</table>

Our example uses the UK value sets (Devlin et al. 2016; Dolan 1997) for the 5- and 3-level utility scores respectively. We use the NDB reference dataset and the copula specification and predict the 5-level utility score from the vector of EQ-5D-3L descriptive items, that is, a $Y_3 \rightarrow v_5$ mapping. Note that it is not necessary to include the `utility5` option in the command, since the Devlin et al. (2016) UK score is the default (and currently only) choice.

```stata
.eq5dmap v5_y3, cov(age male) mo(NDBcopula)items3(Y3_1 Y3_2 Y3_3 Y3_4 Y3_5)
```

No 5L utility scale specified: default is UK scale
Summary of inputs to eq5dmap:
The 5-level utility scale is: UK
The age covariate is contained in input variable: age
The gender covariate is contained in input variable: male
Mapping from Y3 to v5
The 3-level descriptive items are contained in input variables: Y3_1 Y3_2 Y3_3 Y3_4 Y3_5
Unweighted mean of predicted 5L score within selected sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>v5_y3</td>
<td>41,310</td>
<td>.4881471</td>
<td>.2640421</td>
<td>-.2226453</td>
<td>.9635902</td>
</tr>
</tbody>
</table>

The striking feature of the predicted utility scores is that they do not cover the full range $[-0.281, 1.000]$ of the Devlin et al. (2016) scale. Instead, they vary between -0.223 and 0.964. This loss of dispersion is an inevitable feature of any minimum mean-squared error prediction based on the conditional expectation, since the purely random component of $v_5(Y_5)$ is inherently unpredictable. But note that the lower dispersion of predicted scores relative to directly-observed scores is not a problem if the scores are to be used in an economic evaluation based on aggregate net benefit, since the approximate EQ-5D-3L utility value such as those typically obtained from published papers or a previous mapping. These additional variables are used in later examples.
eq5dmap a mapping command for EQ-5D

The conditional mean of QALYs is not affected by the loss of dispersion \[ \text{mean} \] However, loss of dispersion does become a problem when confidence intervals are to be computed or the distribution, rather than mean or aggregate, of net benefit is required. See Hernández-Alava and Pudney (2016) for details of the full predictive distribution which can be used to handle such cases.

5.2 Basic use of eq5dmap: mapping from an EQ-5D-3L score to an EQ-5D-5L score

The second type of mapping generates the predicted 5-level utility score from a specified utility value (a \( v_3 \rightarrow v_5 \) mapping). In this example, we choose the precise utility score \( u_3 \) corresponding to the actual health state description in the dataset and an arbitrary small caliper of 0.001:

\[
\text{eq5dmap v5_u3, cov(age male) mo(NDBcopula) score3(u3) cal(0.001)}
\]

No 5L utility scale specified: default is UK scale
No 3L utility scale specified: default is UK scale
Summary of inputs to eq5dmap:
The 5-level utility scale is: UK
The age covariate is contained in input variable: age
The gender covariate is contained in input variable: male
Mapping from v3 to v5
The 3-level utility scale is: UK
The 3-level score is contained in input variable: u3
The caliper is: .001
Unweighted mean of predicted 5L score within selected sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>v5_u3</td>
<td>41,310</td>
<td>.4881471</td>
<td>.2623586</td>
<td>-.2226453</td>
<td>.9635902</td>
</tr>
</tbody>
</table>

The means of the two mapped variables, \( v_5 \cdot v_3 \) and \( v_5 \cdot u_3 \), are the same up to the seventh decimal point, but not their standard deviations. This is due to the weaker conditioning information contained in the utility score. For the majority of the UK utility values in EQ-5D-3L there is a one-to-one correspondence between the health state and its assigned value. For example the worst health state described by \( Y_3 = (3,3,3,3,3) \) is the only one that has a value of -0.594. However, for a small number of utility values there is not a one-to-one relationship because the same utility value corresponds to two different health states. For example, the two distinct health states described by \( Y_3 = (1,2,2,1,1) \) and \( Y_3 = (2,1,1,1,2) \) have the same utility value of 0.779. Thus, the calculation of the expected EQ-5D-5L needs to take into account that an EQ-5D-3L value of 0.779 can result from either of those two health states.

3. Provided calculation of QALYs involves only linear operations, as is typically the case.
5.3 Mapping from an approximate utility score: choice of caliper

We now repeat the mapping using the variable u3hat, which contains approximate utility values that do not match exactly any of the values found in the ‘official’ EQ-5D-3L utility set. Mismatch often arises in practice because utility scores have been predicted (e.g. from a previous mapping) rather than observed directly. The choice of caliper is then important, and we illustrate the effect of making alternative choices for the caliper. Again, we use the mixed copula model and the NDB reference dataset. The following code is used:

```plaintext
// Predicted UK utility scores:
// loop over four different calipers: 0.01, 0.03, 0.05, 0.1
foreach c of numlist 1 3 5 10 {
    local cal=`c´/100
    di "Caliper = `cal´, Specification = copula, Data = NDB"
    di "Predicted UK scores..."
    eq5dmap v5_u3_`c´ , covariates(age male) ///
        model(NDBcopula) score3(u3hat) caliper(`cal´)
    // Correlations with mapped actual score:
    corr v5_u3_`c´ v5_u3
}
```

The output from the first two passes of the loop is reproduced below and summarized in the first column of Table 1.

| Caliper = .01, Specification = copula, Data = NDB |
| Predicted UK scores... |
| No 5L utility scale specified: default is UK scale |
| No 3L utility scale specified: default is UK scale |
| Summary of inputs to eq5dmap: |
| The 5-level utility scale is: UK |
| The age covariate is contained in input variable: age |
| The gender covariate is contained in input variable: male |
| Mapping from v3 to v5 |
| The 3-level utility scale is: UK |
| The 3-level score is contained in input variable: u3hat |
| The caliper is: .01 |

Unweighted mean of predicted 5L score within selected sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>v5_u3_1</td>
<td>38,590</td>
<td>.5100907</td>
<td>.2319041</td>
<td>-.2226453</td>
<td>.9019883</td>
</tr>
</tbody>
</table>

Warning: It was not possible to find a valid point using the current caliper. Missing values generated.

(obs=38,590)

4. The variable u3hat was constructed as the prediction from a tobit model with lower and upper limits equal to the theoretical EQ-5D-3L bounds, (-0.594 and 1) conditional on the 3L descriptive items. This model has been used here for simplicity but has been shown not to be adequate to model EQ-5D-3L data (Hernandez-Alava et al. 2012).
Caliper = .03, Specification = copula, Data = NDB
Predicted UK scores...
No 5L utility scale specified: default is UK scale
No 3L utility scale specified: default is UK scale
Summary of inputs to eq5dmap:
The 5-level utility scale is: UK
The age covariate is contained in input variable: age
The gender covariate is contained in input variable: male
Mapping from v3 to v5
The 3-level utility scale is: UK
The 3-level score is contained in input variable: u3hat
The caliper is: .03
Unweighted mean of predicted 5L score within selected sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>v5_u3_3</td>
<td>41,310</td>
<td>0.4945065</td>
<td>0.2505553</td>
<td>-0.2226453</td>
<td>0.8491594</td>
</tr>
<tr>
<td></td>
<td>(obs=41,310)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v5_u3_3</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v5_u3</td>
<td>0.8938</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using the smallest caliper, 0.01, Stata gives a warning. For 7\% of the observations, matches could not be found within the caliper, generating missing mapped values. These observations tend to be at the extremes of the EQ-5D-3L distribution where the gaps between consecutive utility values are larger. Even this small number of missing values is enough to distort the mean since missingness is not symmetric across the upper and lower tails of the utility distribution. For that reason, it may be unwise to carry out precise mapping (i.e. a very small caliper) when utility values are approximate. In this case, use of a slightly larger caliper of 0.03 resolved the issue. In our artificial example, a larger caliper generates a result which is more highly correlated with the result of an exact mapping of actual scores: compare the rows of Table 1 for calipers 0.01, 0.03, 0.05 and 0.1. The mean of the predicted scores is fairly stable for different calipers and, as expected, the standard deviation decreases as the caliper increases. Note that, for any caliper, the mean of the mapped approximate utility is systematically above the mean of the mapping from the actual utility score. This happens to be a consequence of the arbitrary method we used to generate hypothetical utility values; it is not inherent in the mapping approach.

### 5.4 Sensitivity of mapping outcomes to the choice of reference dataset

A reference dataset is one that contains simultaneous observations on the two versions of EQ-5D. From that dataset, our estimate of the joint distribution of EQ-5D-3L and EQ-5D-5L responses is derived, which in turn is used to form a conditional predictor of EQ-5D-5L. Different reference datasets will yield different mapping results, and it is important to know how great those differences might be, to give some indication of
### Table 1: Summary statistics for NDB and EQG mapped values for EQ-5D-5L

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Mapping model based on...</th>
<th>Copula</th>
<th>Gauss</th>
<th>Copula</th>
<th>Gauss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NDB</td>
<td>EQG</td>
<td>NDB</td>
<td>EQG</td>
</tr>
<tr>
<td>Mapping actual UK utility score</td>
<td></td>
<td>NDB</td>
<td>0.488</td>
<td>0.416</td>
<td>0.473</td>
</tr>
<tr>
<td>Mean predicted EQ-5D-5L score</td>
<td></td>
<td>EQG</td>
<td>0.416</td>
<td>0.473</td>
<td>0.408</td>
</tr>
<tr>
<td>Minimum predicted EQ-5D-5L score</td>
<td></td>
<td>NDB</td>
<td>-0.223</td>
<td>-0.239</td>
<td>-0.216</td>
</tr>
<tr>
<td>s.d. predicted EQ-5D-5L score</td>
<td></td>
<td>EQG</td>
<td>0.262</td>
<td>0.281</td>
<td>0.258</td>
</tr>
<tr>
<td>Mapping predicted UK utility score: caliper = 0.01, n=38,590</td>
<td></td>
<td>NDB</td>
<td>0.510</td>
<td>0.441</td>
<td>0.495</td>
</tr>
<tr>
<td>Mean predicted EQ-5D-5L score</td>
<td></td>
<td>EQG</td>
<td>0.441</td>
<td>0.495</td>
<td>0.429</td>
</tr>
<tr>
<td>Minimum predicted EQ-5D-5L score</td>
<td></td>
<td>NDB</td>
<td>-0.223</td>
<td>-0.239</td>
<td>-0.216</td>
</tr>
<tr>
<td>s.d. predicted EQ-5D-5L score</td>
<td></td>
<td>EQG</td>
<td>0.232</td>
<td>0.252</td>
<td>0.230</td>
</tr>
<tr>
<td>Correlation with mapped actual score</td>
<td></td>
<td>NDB</td>
<td>0.848</td>
<td>0.874</td>
<td>0.860</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EQG</td>
<td>0.874</td>
<td>0.860</td>
<td>0.898</td>
</tr>
<tr>
<td>Mapping predicted UK utility score: caliper = 0.03, n=41,310</td>
<td></td>
<td>NDB</td>
<td>0.495</td>
<td>0.425</td>
<td>0.480</td>
</tr>
<tr>
<td>Mean predicted EQ-5D-5L score</td>
<td></td>
<td>EQG</td>
<td>0.425</td>
<td>0.480</td>
<td>0.415</td>
</tr>
<tr>
<td>Minimum predicted EQ-5D-5L score</td>
<td></td>
<td>NDB</td>
<td>-0.223</td>
<td>-0.239</td>
<td>-0.216</td>
</tr>
<tr>
<td>s.d. predicted EQ-5D-5L score</td>
<td></td>
<td>EQG</td>
<td>0.251</td>
<td>0.268</td>
<td>0.248</td>
</tr>
<tr>
<td>Correlation with mapped actual score</td>
<td></td>
<td>NDB</td>
<td>0.894</td>
<td>0.913</td>
<td>0.903</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EQG</td>
<td>0.913</td>
<td>0.903</td>
<td>0.930</td>
</tr>
<tr>
<td>Mapping predicted UK utility score: caliper = 0.10, n=41,310</td>
<td></td>
<td>NDB</td>
<td>0.494</td>
<td>0.424</td>
<td>0.480</td>
</tr>
<tr>
<td>Mean predicted EQ-5D-5L score</td>
<td></td>
<td>EQG</td>
<td>0.424</td>
<td>0.480</td>
<td>0.414</td>
</tr>
<tr>
<td>Minimum predicted EQ-5D-5L score</td>
<td></td>
<td>NDB</td>
<td>-0.223</td>
<td>-0.239</td>
<td>-0.216</td>
</tr>
<tr>
<td>s.d. predicted EQ-5D-5L score</td>
<td></td>
<td>EQG</td>
<td>0.249</td>
<td>0.267</td>
<td>0.246</td>
</tr>
<tr>
<td>Correlation with mapped actual score</td>
<td></td>
<td>NDB</td>
<td>0.900</td>
<td>0.917</td>
<td>0.909</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EQG</td>
<td>0.917</td>
<td>0.909</td>
<td>0.934</td>
</tr>
</tbody>
</table>

We again examine the two types of mapping implemented in `eq5dmap`, beginning with the \(Y_3 \rightarrow v_5\) mapping. The following code runs `eq5dmap`, using four loops to repeat the analysis over two values of age (40 and 70), both genders and the two choices for model specification and dataset. Correlations between the mapped results using the two different reference datasets are also computed.

```plaintext
// Loop over age/gender types
foreach a of numlist 40 70 {
    forvalues m=0/1 {
        // Loop over model specification and dataset
        foreach spec in gauss copula {

```
**eq5dmap a mapping command for EQ-5D**

```stata
test
foreach dat in NDB EQG {
    di "Age = `a´, Gender = `m´, Specification = `spec´, Data = `dat´"
    // eq5dmap produces means, etc. of predicted EQ-5D-3L utility scores across health states
    eq5dmap v5_`spec´_`dat´_`a´_`m´ if age==`a´ & male==`m´, ///
        covariates(age male) model(`dat´`spec´) ///
        items3(Y3_1 Y3_2 Y3_3 Y3_4 Y3_5)
}
}
// Compute correlations between results for different reference datasets
corr v5_gauss_NDB_`a´_`m´ v5_gauss_EQG_`a´_`m´ if age==`a´ & male==`m´
corr v5_copula_NDB_`a´_`m´ v5_copula_EQG_`a´_`m´ if age==`a´ & male==`m´
```

The output from the first two passes through the loops is reproduced below

```
Age = 40, Gender = 0, Specification = gauss, Data = NDB
No 5L utility scale specified: default is UK scale
Summary of inputs to eq5dmap:
The 5-level utility scale is: UK
The age covariate is contained in input variable: age
The gender covariate is contained in input variable: male
Mapping from Y3 to v5
The 3-level descriptive items are contained in input variables: Y3_1 Y3_2 Y3_3 Y3_4 Y3_5
> 3_4 Y3_5
Unweighted mean of predicted 5L score within selected sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>v5_gauss_N-0</td>
<td>243</td>
<td>.462048</td>
<td>.2630249</td>
<td>-.2160462</td>
<td>.9488086</td>
</tr>
</tbody>
</table>

Age = 40, Gender = 0, Specification = gauss, Data = EQG
No 5L utility scale specified: default is UK scale
Summary of inputs to eq5dmap:
The 5-level utility scale is: UK
The age covariate is contained in input variable: age
The gender covariate is contained in input variable: male
Mapping from Y3 to v5
The 3-level descriptive items are contained in input variables: Y3_1 Y3_2 Y3_3 Y3_4 Y3_5
> 3_4 Y3_5
Unweighted mean of predicted 5L score within selected sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>v5_gauss_E-0</td>
<td>243</td>
<td>.4147143</td>
<td>.2694663</td>
<td>-.2345741</td>
<td>.9768296</td>
</tr>
</tbody>
</table>
```

Table 2 summarizes the results in terms of the mean score, minimum score and standard deviation across EQ-5D-5L health states, together with the correlations between predictions produced by each model specification on the two reference datasets. The mean prediction is always larger using the NDB rather than EQG dataset and the mixed copula rather than Gaussian model. The differences in the mean predicted scores are slightly larger at age 70 than at 40. The correlations across the datasets are generally high but slightly lower for older individuals.

Figure 1 plots the empirical distributions (kernel densities) of the mapped EQ-5D-5L values for a 70 year old male. For a given choice of model (the mixed copula model is illustrated), there are noticeable differences between the results for the two reference
Table 2: Summary statistics for NDB and EQG mapped values for EQ-5D-5L

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Mapping model based on...</th>
<th>Age 40, female</th>
<th>Age 40, male</th>
<th>Age 70, female</th>
<th>Age 70, male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Copula Gauss</td>
<td>NDB</td>
<td>EqG</td>
<td>NDB</td>
<td>EqG</td>
</tr>
<tr>
<td>Mean predicted EQ-5D-5L score</td>
<td></td>
<td>0.472</td>
<td>0.422</td>
<td>0.462</td>
<td>0.415</td>
</tr>
<tr>
<td>Minimum predicted EQ-5D-5L score</td>
<td></td>
<td>-0.223</td>
<td>-0.237</td>
<td>-0.216</td>
<td>-0.235</td>
</tr>
<tr>
<td>s.d. predicted EQ-5D-5L score</td>
<td></td>
<td>0.268</td>
<td>0.285</td>
<td>0.263</td>
<td>0.269</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>0.970</td>
<td>0.979</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age 70, female</td>
<td>Age 70, male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean predicted EQ-5D-5L score</td>
<td></td>
<td>0.487</td>
<td>0.432</td>
<td>0.478</td>
<td>0.420</td>
</tr>
<tr>
<td>Minimum predicted EQ-5D-5L score</td>
<td></td>
<td>-0.215</td>
<td>-0.234</td>
<td>-0.206</td>
<td>-0.231</td>
</tr>
<tr>
<td>s.d. predicted EQ-5D-5L score</td>
<td></td>
<td>0.267</td>
<td>0.286</td>
<td>0.261</td>
<td>0.270</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>0.975</td>
<td>0.978</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

datasets. For a given reference dataset (the NDB is illustrated), the difference between the mixed copula and gaussian models is less pronounced, although there remain differences at the higher end of the utility values.

Next we repeat the exercise, for a $\nu_3 \rightarrow \nu_5$ mapping using the entire dataset. First we use $u_3$, the variable containing the actual UK utility scores corresponding to the health states described by the variables $Y_{3.1}$ to $Y_{3.5}$. The following code produces results which are summarized in the first panel of Table 1.

```plaintext
// Actual UK 3-level utility scores:
// loop over model specification and dataset
foreach spec in gauss copula {
    foreach dat in NDB EqG {
        di "Specification = ’spec’", Data = ’dat’"
        di "Actual UK scores..."
        eq5dmap v5_’spec’_’dat’_actual , covariates(age male) model(’dat’’spec’) ///
               score3(u3) caliper(0.001)
    }
}
```

As before, all mappings using the NDB dataset produce higher average mapped values than those using the EQG dataset (see Table 1). The smaller the caliper, the lower the correlations with the values mapped from the actual utility scores and also
Figure 1: Comparison of kernel densities of mapped values for EQ-5D-5L for a 70 year old male.

Figure 2: Kernel densities of mapped values for EQ-5D-5L using the copula model and the NDB reference dataset; actual vs approximate utility values for a choice of caliper choices.
5.5 Mapping using weights

Our dataset contains a row for every possible EQ-5D-3L health state by age and gender. The mean 3-level utility score in this artificial dataset is 0.137. Such a low level of utility is not typical of datasets encountered in practice and it is due to the large proportion of health states in the 3-level descriptive system with associated utility scores around and below 0 (equivalent to death). Even disease-specific samples in very severely-affected sub-populations have much larger average 3-level utility scores: for example, the mean 3-level utility values in the EQG and NDB samples are 0.628 and 0.681 respectively and the proportion of individuals with a negative utility value is 8% and 4%. This contrasts with the almost 35% health states with negative utility values in the whole descriptive system.

In this section we demonstrate the use of the command option \texttt{[weight]} to calculate the mean mapped 5-level utility for the populations in the EQG and the NDB samples. The variables \texttt{fwEQG} and \texttt{fwNDB} in the dataset contain frequency weights corresponding to the EQG and NDB datasets respectively. The following code continues with the choice of the NDB mixed copula mapping model to map the EQ-5D-3L descriptive items into a predicted 5-level utility score, but it uses the EQG frequency weights to calculate the weighted mean of the mapped utility score.

```
. eq5dmap v5_y3EQG [fw = fwEQG], cov(age male) mo(NDBcopula)items3(Y3_1 Y3_2 Y3_3 Y3_4 Y3_5)
No 5L utility scale specified: default is UK scale
Summary of inputs to eq5dmap:
The 5-level utility scale is: UK
The age covariate is contained in input variable: age
The gender covariate is contained in input variable: male
Mapping from Y3 to v5
The 3-level descriptive items are contained in input variables: Y3_1 Y3_2 Y3_3 Y3_4 Y3_5
Weighted mean of predicted 5L score within selected sample
Variable | Obs | Mean | Std. Dev. | Min | Max
--- | --- | --- | --- | --- | ---
v5_y3EQG | 3,539 | .7209188 | .2271723 | -.2155897 | .9622371
```

Now, we carry out the same mapping but using the NDB frequency weights instead.

```
. eq5dmap v5_y3NDB [fw = fwNDB], cov(age male) mo(NDBcopula)items3(Y3_1 Y3_2 Y3_3 Y3_4 Y3_5)
No 5L utility scale specified: default is UK scale
Summary of inputs to eq5dmap:
The 5-level utility scale is: UK
The age covariate is contained in input variable: age
The gender covariate is contained in input variable: male
Mapping from Y3 to v5
The 3-level descriptive items are contained in input variables: Y3_1 Y3_2 Y3_3 Y3_4 Y3_5
Weighted mean of predicted 5L score within selected sample
Variable | Obs | Mean | Std. Dev. | Min | Max
--- | --- | --- | --- | --- | ---
v5_y3NDB | 5,205 | .7769872 | .1649192 | -.2223183 | .9618608
```

The mean difference between 3-level and 5-level utility scores calculated for the EQG and NDB sample compositions are fairly small: 0.093 and 0.095 respectively. In com-
eq5dmap a mapping command for EQ-5D

Comparison, the average 3-level/5-level difference is 0.351 when calculated (unweighted) over all the 243 possible descriptive outcomes for EQ-5D-3L and all age-gender groups. This is consistent with the findings in (Hernández-Alava and Pudney 2016) and (Hernandez-Alava et al. 2017) of minor differences between 3-level and 5-level EQ-5D at the top of their range but much larger differences at the bottom.

We now investigate the effect of using alternative weights on the correlations between the mappings based on different reference datasets. The following code calculates the correlations between the \( \nu_3 \rightarrow \nu_5 \) mapped values in Section 5.4 using the mixed copula model.

\[
\begin{align*}
\text{.} & \text{ // calculate correlations - unweighted} \\
\text{.} & \text{ correlate v5_copula_NDB_actual v5_copula_EQG_actual} \\
& \text{(obs=41,310)} \\
& \text{v5_copula___} \quad 1.0000 \\
& \text{v5_copula___} \quad 0.9534 \quad 1.0000
\end{align*}
\]

\[
\begin{align*}
\text{.} & \text{ // calculate correlations - weighted} \\
\text{.} & \text{ correlate v5_copula_NDB_actual v5_copula_EQG_actual \ [fw = fwNDB]} \\
& \text{(obs=5,205)} \\
& \text{v5_copula___} \quad 1.0000 \\
& \text{v5_copula___} \quad 0.9964 \quad 1.0000
\end{align*}
\]

\[
\begin{align*}
\text{.} & \text{ correlate v5_copula_NDB_actual v5_copula_EQG_actual \ [fw = fwEQG]} \\
& \text{(obs=3,539)} \\
& \text{v5_copula___} \quad 1.0000 \\
& \text{v5_copula___} \quad 0.9955 \quad 1.0000
\end{align*}
\]

The unweighted correlation between the mapped values using different reference datasets is high, 0.9534, when calculated unweighted across all possible EQ-5D-3L outcomes, and it increases to a value close to 1 when the sample is weighted to either the NDB or the EQG composition. This serves to demonstrate that inconsistencies between 3-level and 5-level utility scores are moderate in samples with a realistic composition, and that their main feature is difference in the mean rather than lack of correlation.

6 Conclusion

In this article we presented a new Stata command, eq5dmap, which calculates the prediction of the newer 5-level version of EQ-5D utility scores from observed or specified values of 3-level EQ-5D (individual items or utility score), age and gender. The predictive distribution was derived from a joint model of the two versions of EQ-5D developed in Hernández-Alava and Pudney (2016) and applied to two different reference datasets.
The use of the command is illustrated through several examples. The first two examples demonstrate the basic use of the command. Further examples illustrate the consequences of a) using different calipers when mapping using approximate utility scores and b) using different reference datasets. The final example demonstrates the use of weights.

7 Acknowledgements

This work was supported by the National Institute for Health and Care Excellence through its Decision Support Unit and the Medical Research Council (grant MR/L022575/1). Pudney acknowledges further ESRC funding through the Centre for Micro-Social Change (grant RES-518-28-5001). The views expressed in this article, and any errors or omissions, are those of the authors only.

8 References


eq5dmap a mapping command for EQ-5D


About the authors

Mónica Hernández-Alava is an applied microeconometrician in the Health Economics and Decision Science section in ScHARR, University of Sheffield, UK.

Steve Pudney is Professor of Health Econometrics at ScHARR, University of Sheffield, UK.